Functional Mapping of Inner Speech Areas: A Preliminary Study with Portuguese Speakers

Carlos Ferreira[1,2,10], Bruno Direito[1], Alexandre Sayal[1,2],
Marco Simões[2,3,4], Inês Cadório[5,6], Paula Martins[5,7,8],
Marisa Lousada[5,6], Daniela Figueiredo[5,6], Miguel Castelo-Branco[2,3],
and António Teixeira[8,9(✉)]

[1] Institute of Nuclear Sciences Applied to Health,
University of Coimbra, Coimbra, Portugal
cdferreira@ua.pt, c_dferreira@yahoo.com
[2] CIBIT Coimbra Institute for Biomedical Imaging and Translational Research,
ICNAS, University of Coimbra, Coimbra, Portugal
[3] Faculty of Medicine, University of Coimbra, Coimbra, Portugal
[4] Center for Informatics and Systems, University of Coimbra, Coimbra, Portugal
[5] School of Health Sciences, University of Aveiro, Aveiro, Portugal
[6] Center for Health Technology and Services Research,
University of Aveiro, Aveiro, Portugal
[7] Institute of Biomedicine, University of Aveiro, Aveiro, Portugal
[8] Institute of Electronics and Telematics Engineering
of Aveiro (IEETA), Aveiro, Portugal
[9] Department of Electronics, Telecommunications and Informatics,
University of Aveiro, Aveiro, Portugal
ajst@ua.pt
[10] Perspectum Diagnostics, Oxford, UK

Abstract. Inner speech can be defined as the act of talking silently with ourselves. Several studies aimed to understand how this process is related to speech organization and language. Despite the advances, some results are still contradictory. Importantly, language dependency is scarcely studied. For this first study of inner speech for Portuguese native speakers using fMRI, we selected a confrontation naming task, consisting of 40 black and white line drawings. Five healthy participants were instructed to name in inner and in overt speech the visually presented image. fMRI data analysis considering the proposed inner speech paradigm identified several brain areas such as the left inferior frontal gyrus, including Broca's area, supplementary motor area, precentral gyrus and left middle temporal gyrus including Wernicke's area. Our results also show more pronounced bilateral activations during the overt speech task when compared to inner speech, suggesting that inner and overt speech activate similar areas but stronger activation can be found in the later. However, this difference stems in particular from significant activation differences in the right pre-central gyrus and middle temporal gyrus.

1   Introduction

Inner Speech is defined as the act of talking to ourselves silently [6,13,15]. Several studies imply inner speech in memory tasks, reading, comprehension, consciousness, inner thought (self-reflection tasks) [6,14] and prospective thought [16]. According to the literature, two levels of inner speech can be defined: one (more abstract) designated of "language of mind" where the syntax is not fully structured and semantics is more personal and subjective; the other level is more concrete and phonological and phonetic components can be present [6].

Aside these two features intrinsic to inner speech, there is still lack of comprehension on how inner speech is related with speech organization and language. To that end, recent work has been developed to better understand the relation between inner and overt speech, and their correspondence to language pathways.

Despite all previous efforts, there is still lack of consensus regarding the relation between inner and overt speech [6,18]. Some of the factors that contribute to this are: the paradigm variability to explore inner and overt speech, some of the studies did not compare inner and overt speech and others did not monitor participants performance [6,18].

To assess their neural underpinnings features, different methods such as Positron Emission Tomography (PET), electroencephalography (EEG), Transcranial Magnetic Stimulation TMS) and Functional Magnetic Resonance Imaging (fMRI) can be used. The recent advances in the field of Magnetic Resonance Imaging (MRI), combining optimized spatial and improved temporal resolution, multivariate supervised learning methods (allowing assessments in real time), established the use of this technique as one of the most important in the understanding of brain mechanisms. The fact that it does not use ionizing radiation (as in PET imaging) also represents a significant advantage of fMRI to assess brain function. fMRI uses the contrast between oxygenated and deoxygenated blood, the blood-oxygenation-level-dependent (BOLD) effect, which is based on the coupling between the hemodynamic response and neuronal activity. Currently, fMRI using BOLD effect, is one of the preferred methods to map neuronal activity [26]. High field MRI scanners are being used to increase signal-to-noise ratio, ultimately improving the ability to map brain function based on the BOLD signal [10,26].

Recent studies tried to use fMRI to understand and identify brain areas involved in inner speech. Areas like the left inferior frontal gyrus (IFG) (including Broca's area), Wernicke's area, right temporal cortex, supplementary motor area (SMA), insula, right superior parietal lobule (SPL) and right superior cerebellar cortex were found to be involved in inner speech [6,9,12,15]. Geva [6] mentions that structural connectivity patterns near the supramarginal gyrus (SMG) (implicated in the dorsal pathway of language) are predictive of internal speech.

In a critical review, it is mentioned that the planning without speech production and articulation is supported by connections between the prefrontal cortex and the left IFG (Broca's area) [9]. Also is stated the existence of projections between areas related with speech production and auditory cortex as a relevant process in verbal self-monitorization of internal speech [9]. It is also mentioned that inner speech nature is supported by connections between frontal and temporal regions, as to inform the areas related with language perception of the self-generated nature of the verbal output. To map the areas related with inner speech, several paradigms are being used. One example [19], analyzes the relation between frontal and temporal activity instructing the participants to say the same word (word repetition task) in different time points - at each second, each 4 s (condition fast vs. slow), and at each second, each 2 s, each 4 s (conditions fast vs. medium vs. slow). The moment where the participants need to perform the task was indicated by a visual cue [19]. Other example used a task to map the inner speech during a working memory task where the authors exploit a storage condition and a manipulation condition with sub-vocal reproduction of letters [12]. This paradigm allowed to identify active brain areas related with working memory during an inner speech task.

Paradigms that include letter or object naming, animal name generation, verb generation, reading, rhyme judgement, counting or semantic fluency tasks are also used to assess inner speech brain related areas [6,18].

In the present study, we focus in the optimization of a paradigm that can be easily used to study inner and overt speech and possible relation between the areas recruited by both processes. We will use a confrontation naming task to evaluate the variability/differences between both speech mechanisms and try to map areas that could be related only with pure inner speech. We also want to assess the feasibility of assessing inner speech related areas when performing a language task in the context of the European Portuguese language.

Paper Structure: The paper is structured as follows: a brief introductory section presenting related work; Sect. 2 details methods of fMRI data acquisition (including the stimulation protocol, MR parameters) and the tools used for image processing and analyses; Sect. 3 provides the most relevant results obtained so far; in Sect. 4 we discuss the results comparing our findings with published literature; finally, the conclusions that can be drawn are presented.

2   Methods

The study consisted in the recording and analysis of fMRI data while native speakers of Portuguese performed inner and overt speech tasks in response to visual stimuli.

Participants: Five healthy native Portuguese speakers volunteers (mean age: 22.2 years old; 3 males) were enrolled in this study. All participants had normal or corrected to normal vision, and no history of neurological disorders.

The Edinburgh handedness test was applied to the participants to ensure they were all right handed (mean 92% right) and they all declare Portuguese as their native language. The study was approved by the Ethics Commission of the Faculty of Medicine of the University of Coimbra and was conducted in accordance with the declaration of Helsinki. All subjects provided written informed consent to participate in the study.

Data Collection: The data was collected using a Siemens Magnetom Trio 3 T scanner (Erlangen, Germany) with a 12-channel head coil. Anatomical images were acquired using a sagittal T1 3D MPRAGE sequence with the following parameters: TR = 2530 ms; TE = 3.42 ms; TI = 1100 ms; flip angle = 7°; 176 slices; matrix size 256×256; voxel size $1 \times 1 \times 1$ mm. After the anatomical scan, functional maps were obtained using axial gradient echo-planar imaging BOLD sequences parallel to the bi-commissural plane with the following parameters: TR = 3000 ms; TE = 30 ms; 40 slices; matrix size 70×70; voxel size $3 \times 3 \times 3$ mm. Visual stimuli were presented on a NordicNeuroLab (Bergen, Norway) LCD monitor, with a resolution of 1920×1080 pixels, refresh rate 60 Hz.

Stimulation Protocol: The experimental protocol consisted in a picture naming task - inner and overt speech - of 40 black and white line drawings selected from Snodgrass & Vanderwart corpus [20] (Fig. 1). The selection of black and white line drawings was preferred over colored pictures because of their simplicity. Additionally, ambiguous pictures that could retrieve more than one target word (e.g. bottle with water) were excluded from this task. The inner and overt speech runs consisted in a block design experiment with nine rest blocks of 15 s and 8 task blocks of 30 s where each image was presented during 3 s, 10 images per block and two repetitions per image in the run. Each run had a total duration of 125 volumes (Fig. 1). In the baseline condition, the participants were instructed to focus on the fixation cross presented. During the task condition, each participant was instructed to name the object silently in the inner speech run and overtly in the overt speech run.

Data Analysis: Preprocessing and analysis were conducted using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, Netherlands). First, individual functional data were analyzed in order to assess data quality (e.g. head motion) and participants' engagement and ability to perform the task proposed. All participants successfully performed the task and were included in the analysis. Preprocessing of single-subject fMRI data included slice-time correction, realignment to the first image to compensate for head motion and temporal high-pass filtering to remove low-frequency drifts. The anatomical images were co-registered to the functional volumes and all images were normalized to Talairach coordinate space [24].

After preprocessing, in the first-level analysis of the functional data, general linear model (GLM) analysis was used for each run. Predictors were modeled as a boxcar function with the length of each condition, convolved with the canonical hemodynamic response function (HRF). Six motion parameters (three
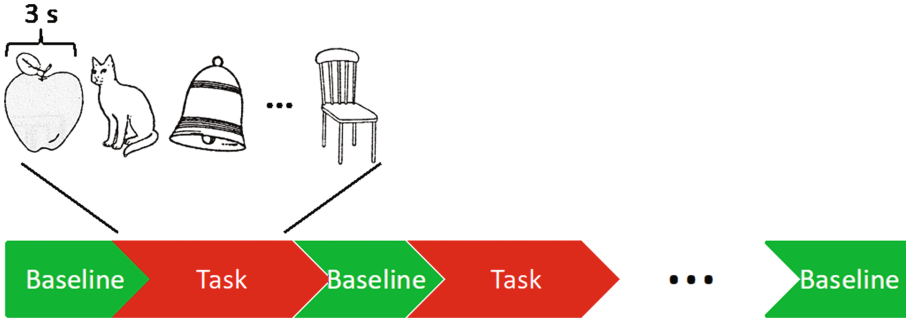
Fig. 1. Stimulation paradigm. Baseline - consisted in a fixation cross and the participants were instructed to focus on it. Task - picture naming task - consisted in a sequence of images visually presented for the participants to name silently (in the inner speech run) or overtly (in the overt speech run). Each image was presented at the screen during 3 s in a total of 10 images per block.

translational and three rotational) and predictors based on spikes (outliers in the BOLD time course) were also included into the GLM as covariates. At the group level, to map the most important brain regions involved in inner and overt speech, we used the contrast "task" > "baseline". First we applied a 3D spatial smoothing with a Gaussian filter of 6 mm. Taking into account the feasibility nature of our study, we performed a fixed-effects (FFX) analysis. To address the multiple comparison problem, we applied False Discovery Rate (FDR) correction (considering a false discovery rate of 0.01). We also aimed at comparing inner and overt speech mechanisms. To this end, we selected a set of regions of interest (ROI) involved in the speech/word formation network (based on a literature review [5,6,12,15,19,23]. Each individual ROI was selected based on the corresponding anatomic landmarks and on the highest t-statistic voxel of the inner speech run statistical map (contrast "confrontation naming task" > "baseline"). Each ROI was defined as a volume with a maximum of 1000 voxels around the peak value (using BrainVoyager QX interface tool to define ROIs). We then computed and compared ROI-GLM t-statistic per ROI between inner and overt speech. We performed a two-sided Wilcoxon rank sum test (Matlab 2017a) to test the statistical significance of the difference between the results obtained considering the inner and overt speech in the naming task.

3   Results

3.1   Whole Brain Analysis - Brain Map of the Naming Task

The FFX-GLM statistical map regarding the inner speech naming task (FFX, q(FDR)< 0.01), considering the contrast of interest "picture naming task" > "baseline" (Fig. 2a), revealed significant activations in the IFG and Middle Frontal Gyrus (MFG) (including Broca's area), preCentral Gyrus (pGC), SMA,

Middle Temporal Gyrus (MTG) (including Wernicke's area), Intraparietal Sulcus (IPS), Occipital areas and Fusiform Gyrus (FG).

Figure 2b presents the FFX-GLM statistical map from the overt speech naming task (FFX, q(FDR) < 0.01), considering the contrast of interest "picture naming task" > "baseline" in which it is possible to identify several brain regions such as the IFG (including Broca's area), pCG, SMA, MTG (including Wernicke's area), Occipital areas and FG.
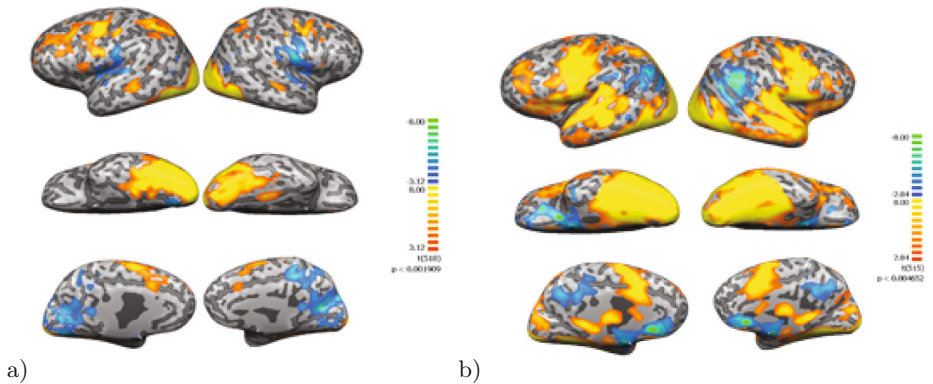

a)   b)

Fig. 2. (a) FFX-GLM group activation map for the inner speech runs (q(FDR) < 0.01), showing areas with higher activation during the task relative to the baseline. (b) FFX-GLM group activation map for the overt speech runs (q(FDR) < 0.01), showing areas with higher activation during the task relative to the baseline. The regions in blue, in both, show the expected deactivation particular in the default mode network. (Color figure online)

3.2 Comparing Inner and Overt Speech - ROI-Based Analysis and the Speech Brain Network

One of the aims of the study was to compare inner and overt speech activation patterns. To this end, considering a literature review on speech-related brain networks, we identified a total of 16 ROIs (summarized in Table 1).

In order to functionally define each ROI, we identified the relevant anatomical landmarks and selected a ROI around the highest t-statistic voxel considering the whole brain inner speech statistical map. Table 1 presents the coordinates of the center of gravity of each ROI (in Talairach coordinates) and the total number of voxels. The beta weights of the contrast "picture naming task" > "baseline" for each region and condition (ROI-GLM) were extracted per participant and run (these weights reflect the BOLD signal variation during the task condition relative to the baseline).

To evaluate the statistical significance of the difference between inner and overt speech naming tasks, we performed a two-sided Wilcoxon rank sum test on the beta values for each ROI. The results are presented in Table 1.

Table 1. Inner vs Overt speech: Talairach coordinates and number of voxels per ROI. Wilcoxon statistical test results to evaluate main differences between inner and overt speech.

Region	Num. voxels	ROI			inner (median)	overt (median)	U	p-value
		x	y	z				
Left middle temporal gyrus	738	−47	−46	9	−0.0244	0.2949	3	0.0556
Left frontal lobe (Broca's area)	1000	−48	146	11	0.0293	0.0908	12	> 0.9999
Left putamen	826	−21	−5	14	0.1152	0.2666	4	0.0952
Right putamen	412	22	0	10	0.1875	0.3125	10	0.6905
Supplementary motor area	962	−3	−5	57	0.2725	0.3740	6	0.2222
Right middle temporal gyrus	414	47	−31	1	0.1699	0.6025	1	**0.0159**
Left M1 - Precentral gyrus	970	−43	−5	41	0.1699	0.5049	5	0.1508
Left intraparietal sulcus	979	−46	−41	41	−0.0137	0.0615	7	0.3095
Left inferior parietal lobule	989	−32	−59	45	0.2920	0.1875	15	0.6905
Right inferior parietal lobule	951	32	−58	44	0.2822	−0.1025	17	0.4206
Right M1 - precentral gyrus	973	48	−4	45	0.3379	0.9150	2	**0.0317**
Left FG	1000	−42	−50	−14	0.3125	0.4873	7	0.3095
Right FG	1000	41	−60	−20	0.2783	1.5811	3	0.0556
Left inferior occipital gyrus	940	−26	−88	−7	1.4863	1.4590	10	0.6905
Right inferior occipital gyrus	941	27	−88	−8	1.6309	1.6074	9	0.5476
Right inferior frontal gyrus	525	48	36	5	0.0156	0.3828	6	0.2222

Our results show that overt speech elicits a stronger activation pattern. Statistical significant differences were found in the right MTG and the right pCG.

Additionally, we computed the subtraction between the overt and inner speech activation maps (Fig. 3). The results suggest that the overt speech activation pattern in most brain structures is higher than the activation pattern presented by the inner speech task, complying with the ROI-GLM results.
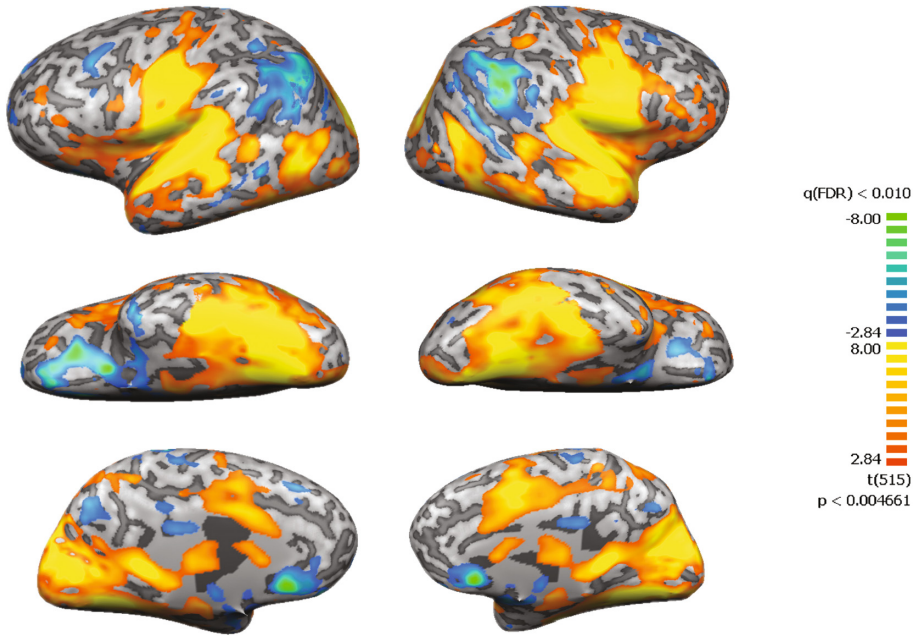
Fig. 3. Subtraction FFX-GLM group activation map between the overt and inner speech runs (q(FDR) < 0.01), showing areas with higher activation during the overt relative to the inner.

4 Discussion

In this study we sought to assess brain activity patterns when performing two speech tasks - one related with inner speech and other related with overt speech. One of the new findings that has not been reported in other studies, was IPS activity. This finding can be explained by the involvement of IPS in tasks related with working memory, attention and attentional control by left fronto-parietal network that can be flexibly allocated to language processing as a function of task demands [2,7,11].

Another interesting finding is the activation of the FG, specially the visual word form area (VWFA) during both tasks. Usually related with the processing of visually presented letter strings, words, pseudowords but also to nonwords stimuli [3–5,23,25], the VWFA was active during the performance of speech tasks with image presentation (non-verbal material) in both tasks. This can be supported by Cohen [3] that mention the relation between the visual system and left lateralized regions engaged in language processing and by Stevens [22] that mention a functional connectivity between visual word form area and core regions of language processing. Bouhali [1] recently showed functional and anatomical connections between visual word form area and most perisylvian language-related areas, including Broca's area.

The major task related difference in the statistical analysis indicates more activation in right precentral gyrus during overt speech task. This is in concordance with some results published in the literature that assume the need of a strong motor response to produce the overt speech controlling all the elements involved in speech production while, the inner speech, as is not so dependent of activating articulatory elements, should have lower pCG activation [8,17,18,21]. Another source of difference is right middle temporal gyrus (rMTG) that shows stronger activation in overt speech when comparing to inner speech. This finding remains controversial in the literature where some authors mentioned that in the inner speech conditions they found high activations in other MTG subregions [18].

These intriguing results can be explained by some limitations in our study. First the small sample size of this exploratory study is more prone to be influenced by single individual results, in particular in an FFX analysis. There is also the possibility that distinct subregions in MTG modulate differentially.

Nevertheless, this first approach in Portuguese speaking participants allow us to map the mechanisms involved in inner speech even without the use of verbal material (e.g. words and sentences). This proof of concept/pilot study may pave way to further explore the mechanisms involved in inner speech when using verbal stimuli.

5   Conclusions

In this work we were able to map the inner speech related areas that are in accordance with the literature and explicit a wider bilateral brain activation during overt speech when compared with inner speech, although these differences dominate mainly in two regions (a part of MTG and M1). Future research should focus on expanding the understanding of the neural correlates of inner and overt speech. In this sense, we believe that using a parametric difficulty level paradigm design (e.g. from vowel to sentence) may represent an important tool to evaluate major differences between the several areas engaged in the inner speech performance when difficulty of the task is increasing.

References

1. Bouhali, F., de Schotten, M.T., Pinel, P., Poupon, C., Mangin, J.F., Dehaene, S., Cohen, L.: Anatomical connections of the visual word form area. J. Neurosci. **34**(46), 15402–15414 (2014)
2. Bray, S., Almas, R., Arnold, A.E., Iaria, G., MacQueen, G.: Intraparietal sulcus activity and functional connectivity supporting spatial working memory manipulation. Cereb. Cortex **25**(5), 1252–1264 (2013)
3. Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., Michel, F.: The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain **123**(2), 291–307 (2000)

4. Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S.: Language-specific tuning of visual cortex? functional properties of the visual word form area. Brain **125**(5), 1054–1069 (2002)

5. Dehaene, S., Le Clec'H, G., Poline, J.B., Le Bihan, D., Cohen, L.: The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport **13**(3), 321–325 (2002)

6. Geva, S., Jones, P.S., Crinion, J.T., Price, C.J., Baron, J.C., Warburton, E.A.: The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. Brain **134**(10), 3071–3082 (2011)

7. Grefkes, C., Fink, G.R.: The functional organization of the intraparietal sulcus in humans and monkeys. J. Anat. **207**(1), 3–17 (2005)

8. Huang, J., Carr, T.H., Cao, Y.: Comparing cortical activations for silent and overt speech using event-related fMRI. Hum. Brain Mapp. **15**(1), 39–53 (2002)

9. Jones, S.R., Fernyhough, C.: Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. Clin. Psychol. Rev. **27**(2), 140–154 (2007)

10. Logothetis, N.K.: What we can do and what we cannot do with fMRI. Nature **453**(7197), 869 (2008)

11. Majerus, S.: Language repetition and short-term memory: an integrative framework. Front. Hum. Neurosci. **7**, 357 (2013)

12. Marvel, C.L., Desmond, J.E.: From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. Brain Lang. **120**(1), 42–51 (2012)

13. Morin, A.: Inner speech. In: Hirstein, W. (ed.) Encyclopedia of Human Behavior, pp. 436–443. Elsevier, London (2012)

14. Morin, A., Hamper, B.: Self-reflection and the inner voice: activation of the left inferior frontal gyrus during perceptual and conceptual self-referential thinking. Open Neuroimaging J. **6**, 78–89 (2012)

15. Morin, A., Michaud, J.: Self-awareness and the left inferior frontal gyrus: inner speech use during self-related processing. Brain Res. Bull. **74**(6), 387–396 (2007)

16. Morin, A., Uttl, B., Hamper, B.: Self-reported frequency, content, and functions of inner speech. Procedia - Soc. Behav. Sci. **30**, 1714–1718 (2011)

17. Palmer, E.D., Rosen, H.J., Ojemann, J.G., Buckner, R.L., Kelley, W.M., Petersen, S.E.: An event-related fMRI study of overt and covert word stem completion. Neuroimage **14**(1), 182–193 (2001)

18. Perrone-Bertolotti, M., Rapin, L., Lachaux, J.P., Baciu, M., Loevenbruck, H.: What is that little voice inside my head? inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behav. Brain Res. **261**, 220–239 (2014)

19. Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M., McGuire, P.K.: Modulation of activity in temporal cortex during generation of inner speech. Hum. Brain Mapp. **16**(4), 219–227 (2002)

20. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exper. Psychol. Hum. Learn. Mem. **6**(2), 174 (1980)

21. Stephan, F., Saalbach, H., Rossi, S.: How the brain plans inner and overt speech production: a combined EEG and fNIRS study. In: 23rd Annual Meeting of the Organization for Human Brain Mapping (OHBM), Vancouver, Canada (2017)

22. Stevens, W.D., Kravitz, D.J., Peng, C.S., Tessler, M.H., Martin, A.: Privileged functional connectivity between the visual word form area and the language system. J. Neurosci. **37**(21), 5288–5297 (2017)

23. Tagamets, M.A., Novick, J.M., Chalmers, M.L., Friedman, R.B.: A parametric approach to orthographic processing in the brain: an fMRI study. J. Cogn. Neurosci. **12**(2), 281–297 (2000)
24. Talairach, J., Tournoux, P.: Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York (1988)
25. Vigneau, M., Jobard, G., Mazoyer, B., Tzourio-Mazoyer, N.: Word and non-word reading: what role for the visual word form area? Neuroimage **27**(3), 694–705 (2005)
26. Willinek, W.A., Schild, H.H.: Clinical advantages of 3.0 T MRI over 1.5 T. Eur. J. Radiol. **65**(1), 2–14 (2008)