





Cortico-Cerebellar Monitoring of Speech Sequence Production

Snežana Todorović^{1,2} , Jean-Luc Anton^{3,4}, Julien Sein^{3,4}, Bruno Nazarian^{3,4},
Valérie Chanoine², Birgit Rauchbauer^{1,2}, Sonja A. Kotz^{5,6}, and Elin Runnqvist^{1,2} 

¹Laboratoire Parole et Langage, CNRS–Aix-Marseille Université, Aix-en-Provence, France

²Institute of Language, Communication and the Brain, Aix-en-Provence, France

³Centre IRM, Marseille, France

⁴INT, CNRS–Aix-Marseille Université, Marseille, France

⁵Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands

⁶Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Keywords: cerebellum, fMRI, internal modeling, monitoring, speech production

ABSTRACT

In a functional magnetic resonance imaging study, we examined speech error monitoring in a cortico-cerebellar network for two contrasts: (a) correct trials with high versus low articulatory error probability and (b) overtly committed errors versus correct trials. Engagement of the cognitive cerebellar region Crus I in both contrasts suggests that this region is involved in overarching performance monitoring. The activation of cerebellar motor regions (superior medial cerebellum, lobules VI and VIII) indicates the additional presence of a sensorimotor driven implementation of control. The combined pattern of pre-supplementary motor area (active across contrasts) and anterior cingulate cortex (only active in the contrast involving overt errors) activations suggests sensorimotor driven feedback monitoring in the medial frontal cortex, making use of proprioception and auditory feedback through overt errors. Differential temporal and parietal cortex activation across contrasts indicates involvement beyond sensorimotor driven feedback in line with speech production models that link these regions to auditory target processing and internal modeling-like mechanisms. These results highlight the presence of multiple, possibly hierarchically interdependent, mechanisms that support the optimizing of speech production.

INTRODUCTION

Speakers monitor their utterances for errors in speech production and rely on a broad cortico-cerebellar network to do so. Specific cognitive mechanisms such as internal modeling, conflict monitoring, and feedback control are all linked to this network (e.g., Gauvin et al., 2016; Hansen et al., 2019a; Okada et al., 2018; Runnqvist, 2023; Runnqvist et al., 2016, 2021; Volfart et al., 2022). Aside from detecting overt articulation errors (external monitoring), monitoring in speech planning (internal monitoring) can also take place and result in the reduction of contextually inappropriate speech errors; speakers sometimes also swiftly repair errors at a speed that would not be possible if they simply relied on sensory feedback (e.g., Hartsuiker & Kolk, 2001; Levelt, 1983; Motley et al., 1981, 1982; Severens et al., 2012). An inherent assumption of existing speech monitoring models is that both internal and external monitoring rely on the same mechanism(s) applied in different processing stages, before and after articulation (e.g., Levelt et al., 1999; Nozari et al., 2011; Pickering & Garrod, 2013). However,

Citation: Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S. A., & Runnqvist, E. (2024). Cortico-cerebellar monitoring of speech sequence production. *Neurobiology of Language*, 5(3), 701–721. https://doi.org/10.1162/nol_a_00113

DOI:
https://doi.org/10.1162/nol_a_00113

Received: 17 April 2022
Accepted: 27 June 2023

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:
Elin Runnqvist
elin.runnqvist@univ-amu.fr

Handling Editor:
Julie Fiez

Copyright: © 2023
Massachusetts Institute of Technology
Published under a Creative Commons
Attribution 4.0 International
(CC BY 4.0) license

External monitoring:

Error detection processes taking place after speech becomes overt.

Internal monitoring:

Error detection processes taking place during speech planning.

several studies have shown that inner speech and articulated speech differ in error patterns (e.g., Oppenheim & Dell, 2008) and velocity of repair after error detection (Nooteboom & Quené, 2017) and that they display partially different neural correlates of error monitoring (e.g., Hansen et al., 2019a; Okada et al., 2018; Runnqvist et al., 2021). Specifically, previous research has shown dissociations between the regions and mechanisms involved when monitoring is indexed by high error probability contexts in which speakers ultimately produce a correct utterance (i.e., tapping into internal monitoring), as opposed to a context in which overt errors are committed (i.e., also tapping into external monitoring). While both contexts engage the cerebellar Crus I, external monitoring also engages a broader set of medial frontal and temporo-parietal regions and the superior medial cerebellum (SMC; e.g., Gauvin et al., 2016; Hansen et al., 2019a; Loh et al., 2020; Okada et al., 2018; Runnqvist et al., 2021). Consequently, the current study aimed at a better understanding of the differential contributions of different parts of this cortico-cerebellar monitoring network. We wanted to:

- 1) assess whether the relevant variable for dissociable brain regions and mechanisms in monitoring is indeed internal versus external monitoring, or rather how distant a linguistic representation is from articulation, or a combination of both;
- 2) refine our understanding of the contributions of distinct cerebellar regions to speech production monitoring, ultimately informing theories about functional cerebellar subdivisions; and
- 3) gain further insight into the specific functional roles of the medial frontal and parieto-temporal cortices in speech monitoring that are currently ambiguous and subject to much debate.

Which Variables Are Relevant in Triggering Differentiation of Brain Regions in Monitoring?

Language production is a complex motor skill that involves several processing levels, ranging from semantic retrieval to articulatory programming and execution. Speech errors can originate at any level of processing, resulting in diverse speech errors (e.g., saying “cat” instead of “dog” is a semantic speech error, while saying “gat” instead of “cat” is a phonological or articulatory-phonetic speech error). Some of the processes in speech production only occur during speech planning (semantic retrieval, lexical access) and are thus more cognitive in nature, while others are concomitant to articulation and are thus more motor related. Prior literature assessed internal monitoring at a typically cognitive stage of speech/language processing (pragmatic or lexical; e.g., Hansen et al., 2019b; Runnqvist et al., 2021; Severens et al., 2012; Teghipco et al., 2023). In turn, external monitoring has mostly been assessed by pooling together multiple overt speech errors, including an undeniable yet unknown ratio of motor source (e.g., Gauvin et al., 2016; Runnqvist et al., 2021). Hence, there is a potential confound between internal and external monitoring as opposed to monitoring of cognitive versus motor properties of speech. This confound can be countered in the case of external monitoring by separating overt errors based on their functional sources (e.g., semantic vs. articulatory errors). However, doing so is difficult as overall error rates are often low, even in error-eliciting language production tasks (but see Ganushchak & Schiller, 2008a, 2008b, and McCall et al., 2023).

Here we adopted an alternative approach to address this potential confound by manipulating error probability (high vs. low), allowing to assess different degrees of internal monitoring (high vs. low), of a motor-related variable, namely, syllabification. Concretely, participants produced pseudowords made up of sound combinations that were either novel (phonotactically illegal)

Phonotactically illegal sequences:
Combinations of speech sounds that
do not exist in the repertoire of a
given language.

or part of their native French language repertoire (phonotactically legal). As in previous studies (e.g., Runnqvist et al., 2021; Severens et al., 2012), the contrast between the more error-prone (illegal) condition and the less error-prone (legal) condition was taken as an index of internal monitoring but at a more motor-related level compared to prior studies. That is, given the illegal condition entails a higher risk of making errors than the legal condition, more monitoring is required to avoid speech errors. Hence, a difference between the two conditions on correct trials necessarily comprises a difference in internal monitoring. Furthermore, as in previous studies (e.g., Gauvin et al., 2016; Riès et al., 2011; Runnqvist et al., 2021), a contrast of correct trials and errors was taken to index external monitoring. As speakers detect their overt errors quite accurately, committing an error involves error detection (e.g., Gauvin et al., 2016; Postma & Noordanus, 1996).

Note that our internal monitoring contrast also likely involved some degree of external monitoring, and similarly the external monitoring contrast involved some degree of internal monitoring. Importantly, the ratio of each monitoring type should differ as in one case speakers manage to avoid errors despite a high error probability (presumably relying more on internal monitoring), while in the other case they do not (presumably relying more on external monitoring). Dissociations between the contrasts of internal and external monitoring would support the hypothesis that they are indeed relevant variables for triggering differential engagement of the monitoring network. Commonalities would instead support the hypothesis that the more relevant dimension is the cognitive to motor continuum.

What Are the Distinct Cerebellar Contributions to Speech Production Monitoring?

The cerebellum is considered central to the internal modeling of self-produced actions (e.g., Blakemore et al., 2001; Knolle et al., 2013). Parallel to programming actions, we also predict the sensory consequences these actions will result in through an efference copy. In turn, the efference copy inhibits the expected neural response in the relevant part of somatosensory cortex (reafference cancellation), which then leads to an error signal should expected and actual sensory consequences mismatch (e.g., McCloskey, 1981; Wolpert et al., 1995). Importantly, such efference copying can take place internally, before an action becomes overt (e.g., Blakemore & Sirigu, 2003; Desmurget & Grafton, 2000), and is found for motor and mental actions, including language (e.g., Argyropoulos, 2016; Ito, 2008; Lesage et al., 2017; Moberget et al., 2014).

Some authors have proposed that this internal modeling is being carried out across the whole cerebellum as a unique computation (e.g., Medina & Mauk, 2000; Moberget & Ivry, 2016). This hypothesis was generated indirectly based on the structural and synaptic properties of the cerebellum. In fact, the same type of basic neural circuitry is repeated throughout all cerebellar subdivisions, constituting the fundamental functional module of the cerebellum (e.g., Purves et al., 2001). In turn, this suggests that a single, characteristic computation may be common to all the functions cerebellar circuitry is involved in. A strong candidate for such a unique computation is performance monitoring through internal modeling of self-generated actions (e.g., Ito, 2008; Peterburs & Desmond, 2016). This computation could then be applied to different types of functions/representations in different cerebellar subcompartments. In fact, just as language can be divided into more cognitive and more motor-related parts, the cerebellum has been characterized by cognitive and motor subcompartments, respectively (e.g., Buckner, 2013; Fiez, 1996; Middleton & Strick, 1994; Stoodley & Schmahmann, 2009; Strick et al., 2009). Such functional topography largely corresponds to the so-called double motor representation (lobules VI and VIII) and triple cognitive representation (Crus I, Crus II and

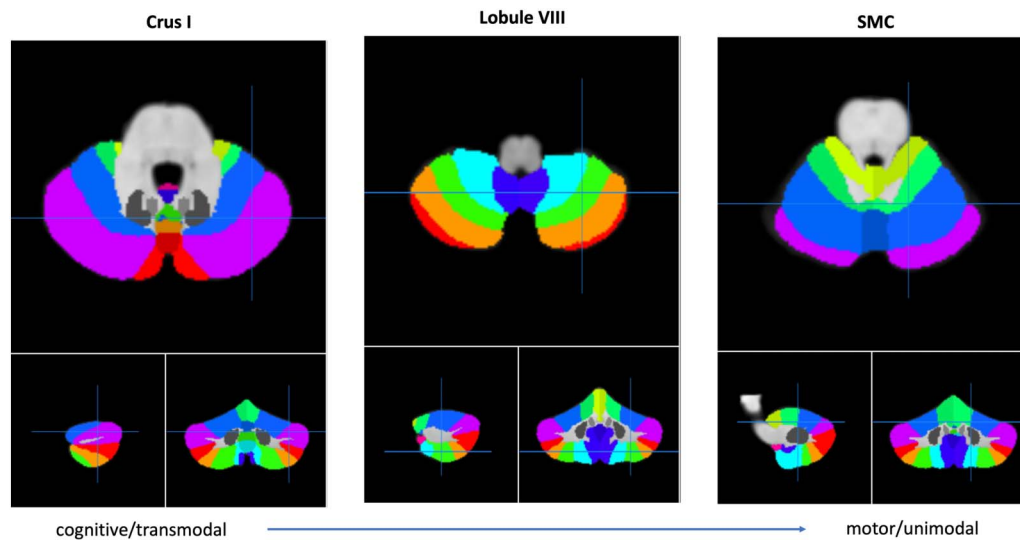


Figure 1. Illustrations of the cerebellum obtained from the SUI online atlas tool (Diedrichsen et al., 2009). The crosshair in each of the three panels is located at different functional subcompartments (from left to right): the cognitive or transmodal region Crus I, the motor region lobule VIII, and the motor region SMC (superior medial cerebellum).

lobule IX; e.g., Guell et al., 2018). A different take on linking macrostructure to function are functional gradients (e.g., D’Mello et al., 2020; Guell et al., 2018). By calculating the similarity of intracerebellar resting state functional connectivity patterns, Guell et al. (2018) established a functional gradient, extending bilaterally from lobules IV/V/VI and lobule VIII to posterior aspects of Crus I and II as well as medial regions of lobule IX. The first extreme of the gradient was related to motor function (i.e., modality specific) and the other extreme to functions involving several modalities (i.e., transmodal; see Figure 1). An interesting question explores the relationship between cerebellar anatomy and motor versus non-motor related speech and monitoring functions. Further, a previous study reported that the cerebellar motor region SMC (or lobule IV) was only involved in external monitoring while the cognitive or transmodal region Crus I was involved in both internal and external monitoring.

Two hypotheses can be generated based on these findings. (1) Different compartments, motor and non-motor, might be involved in monitoring motor and non-motor aspects of speech, respectively. If this were the case, manipulating an articulatory motor variable such as phonotactic probability should only activate cerebellar motor compartments, such as SMC and lobule VIII, but not Crus I. (2) Cognitive or transmodal parts of the cerebellum such as Crus I engage in the monitoring of all types of speech related actions, whether motor or cognitive in nature, reflecting that any part of speech production triggers a hierarchically superior type of control. Consequently, Crus I should be implicated regardless of the monitoring type and level of linguistic representation. Here we aimed at contrasting these two hypotheses.

What Are the Functional Roles of the Medial Frontal and Parieto-Temporal Cortices in Speech Monitoring?

Several studies have observed activation of the medial frontal cortex in external monitoring as indexed by overt speech errors contrasted with correct speech production. Notably, both the pre-supplementary motor area (pre-SMA) and the anterior cingulate cortex (ACC) show differential activation in such circumstances (e.g., Gauvin et al., 2016; Runnqvist et al., 2021). One way to account for this pattern is the concept of conflict-based monitoring (e.g., Gauvin &

Hartsuiker, 2020; Nozari et al., 2011). In this framework, conflict itself triggers monitoring in the medial frontal cortex, that, in turn, relays the information to the prefrontal cortex to regulate behavior. This mechanism (e.g., de Zubicaray et al., 1998; Gauvin et al., 2016; Nozari et al., 2011; Riès et al., 2011) would detect conflict at different levels of speech processing (more cognitive and more motor related) and hence in both external and internal error monitoring. However, activation of the pre-SMA in speech production has also been linked to motor-related difficulties, suggesting its role in early stages of motor sequence learning related to the planning and execution of complex speech sequences (Alario et al., 2006). Linking this evidence to that of the pre-SMA in overt error processing, a plausible alternative to conflict monitoring is that this region is involved in monitoring articulatory aspects of production (thus sensorimotor driven). When manipulating an articulatory, motor proximate, variable, this account predicts the engagement of the pre-SMA both for internal monitoring (in situations of high articulatory error probability) and external monitoring (for overt articulatory errors). Regarding the role of the ACC in speech monitoring, an alternative to conflict monitoring could be monitoring of feedback provided via the overt errors (Loh et al., 2020; Runnqvist et al., 2021), where only external monitoring should differentially engage this region (sensorimotor driven).

Finally, there is an ongoing debate about the involvement of temporal and temporo-parietal regions in speech monitoring (e.g., Meekings & Scott, 2021). Traditionally, the detection of speech production errors was considered to exclusively relate to sensory feedback (e.g., Hartsuiker & Kolk, 2001; Levelt et al., 1999; Roelofs, 2020). Either phonologically encoded inner speech or articulated and audible speech would feed into the same speech comprehension loop, leading to the conceptual level of processing, where the appropriateness of an utterance would be monitored. That is, speakers would detect errors by listening to and understanding their own inner and overt speech, similarly to how they listen to and understand someone else's speech. Because of the central role of speech comprehension in this theory, its hypothesized neural basis is the posterior superior temporal gyrus (pSTG). More recent models of speech production also integrate the temporal (pSTG) and temporoparietal (Sylvian fissure at the temporoparietal boundary, or SPT) in relation to auditory targets and error maps. In these models, mechanisms akin to internal modeling have been implemented as an error signal generated by a mismatch between a (somato-)sensory target and a motor command (e.g., Guenther et al., 2006; Hickok, 2012, 2014; Pickering & Garrod, 2013; Tourville & Guenther, 2011). Furthermore, in these models the cerebellum is hypothesized to contribute to sensory predictions that form the auditory and somatosensory targets, though their role is limited to speech motor function. Hence, these models would also predict pSTG or SPT activation during internal and external monitoring, arguably in combination with cerebellar activation. Indeed, there is converging evidence that the pSTG engages in external monitoring as indexed by articulated, though not necessarily audible speech, overt speech errors, and halted speech (e.g., Hansen et al., 2019a; Okada et al., 2018; Runnqvist et al., 2021). Note that some studies manipulating auditory feedback (e.g., Hirano et al., 1997) or looking into auditory hallucinations (e.g., Shergill et al., 2000) also link pSTG to internal monitoring (e.g., Indefrey & Levelt, 2004). Nonetheless, these studies varied in their use of auditory variables. It is unclear though whether the pSTG activation is actually reflecting internal monitoring, especially as studies explicitly controlling for auditory variables do not seem to observe pSTG in conditions taxing internal monitoring (e.g., Hansen et al., 2019a; Runnqvist et al., 2021). Here we wanted to shed further light on the functional role of temporal and parietal regions by assessing their potential involvement in internal and external monitoring.

Table 1. MNI coordinates and references of the ROI classified by anatomical regions.

Broad anatomical region	Anatomical subregion	MNI coordinates
Cerebellum	R Crus I	(38, -64, -30) Stoodley and Schmahmann, 2009*
	R SMC	(16, -59, -23) Golfinopoulos et al., 2010*
	R posterior lobule VIII	(28, -56, -54) Rauschecker et al., 2008
Medial frontal cortex	L ACC	(-6, 20, 34) Gauvin et al., 2016
	L Pre-SMA	(-6, 8, 49) Gauvin et al., 2016
Temporo-parietal cortex	L pSTG	(-65, -33, 14) Golfinopoulos et al., 2010*
	L SPT	(-54, -30, 14) Okada and Hickok, 2006

Note. R = right, L = left. Asterisks indicate meta-analysis or model-based coordinates. SMC = superior medial cerebellum; ACC = anterior cingulate cortex; pre-SMA = pre-supplementary motor area; pSTG = posterior superior temporal gyrus; SPT = sylvian fissure at the parieto-temporal boundary; MNI = Montreal Neurological Institute.

Cerebellar and Cortical ROI Definition in the Current Study

To zoom in on the monitoring network of interest, we used a region of interest (ROI) approach, in which seven ROIs within the cerebellum, medial frontal, temporal, and temporo-parietal cortices were examined in two contrasts. Six of these regions were taken from Runnqvist et al. (2021) and one from Rauschecker et al. (2008; see Table 1 in the next section). As speech is assumed to be left-lateralized in cortex and right-lateralized in the cerebellum, we limited the cortical ROIs to the left and the cerebellar ROIs to the right hemisphere. In the cerebellum, we chose three right hemispheric regions: Crus I on the transmodal side of Guell et al. (2018) functional gradient, the phylogenetically older SMC, and the phylogenetically younger posterior lobule VIII on the motor side of the gradient. In addition, we chose two regions in the medial frontal cortex (left ACC and pre-SMA) and two regions in temporal and temporo-parietal cortex (left pSTG and SPT). This selection sheds light on potential dissociations across internal and external monitoring for the articulatory variable of interest. Complementing the ROI analyses, we also examined the differential activations related to internal and external monitoring contrasts in broader anatomical ROIs (cerebellum, medial frontal cortex, and temporo-parietal cortex) in a whole brain analysis.

MATERIALS AND METHODS

Participants

The study received ethical approval (filed under Id 2017-A03614-49 from the regional ethical committee, Comité de protection des personnes sud Méditerranée I). Twenty-six (17 females) right-handed native speakers of French participated in exchange for monetary compensation. Two participants (1 female, 1 male) were excluded from further analyses due to excessive head movements during data acquisition. The average age of the remaining 24 participants was 26.2 (*SD* 3.5). No participant reported any history of language, neurological, or hearing disorders.

Materials

Target stimuli were 36 CCVCC syllables (C = consonant, V = vowel) composed of French phonemes. In one half of the stimuli, these phonemes were combined in a phonotactically legal

way in French, and their production therefore required less monitoring, even if the syllable was meaningless. The other half of the stimuli were syllables consisting of novel combinations of phonemes (phonotactically illegal), requiring increased monitoring as their production had to be learned (e.g., Segawa et al., 2015). Two-consonant clusters were selected from the French language database *lexique.org* (New et al., 2001) and filtered for their frequency of occurrence at the beginning or at the end of the syllable (New & Spinelli, 2013). The clusters with high frequency (141.6 ± 251.6) were used to form low monitoring load stimuli, and those with a frequency close to zero (0.3 ± 0.8) to form the high monitoring load ones. The resulting syllables were checked for orthographic neighbors using WordGen software (Duyck et al., 2004). Only one stimulus had one orthographic neighbor (spald), the rest of stimuli had none.

All stimuli were recorded by a Serbian speaker, as Serbian allows pronunciation of all used combinations. Stimuli were also visually presented according to French orthographic rules. As this experiment was part of a larger study that took place over two sessions, any given participant was only presented with half of the stimuli in a counterbalanced manner. Each stimulus was presented 25 times over five experimental runs. The order of the stimuli was pseudorandomized in eight different lists so that the same category (high or low monitoring load) was not presented more than twice in a row and that each stimulus was presented N times before any other stimulus was presented $N + 1$ times. Thus, each participant was presented with 225 (25×9) high monitoring load and 225 low monitoring load syllables.

Procedure

Syllables remained on the screen for 798 ms and were synchronously presented auditorily for 660 ms. They were preceded by a fixation cross for 532 ms and followed by a blank screen during which the participants were instructed to pronounce the presented syllable. The time for pronunciation was 1,530 ms. The jittered interstimulus-interval was 930 to 1,464 ms. In sum, on average a single trial lasted 4.059 seconds, and the experimental runs lasted 6 minutes each. Figure 2 shows an exemplary trial. Stimulus presentation was controlled by a custom-made software compiled using the LabVIEW development environment (National Instruments) and OptoActive audio system (OptoAcoustics). This software also allowed the recording of vocal productions: Trials were recorded and labeled independently using a multifunction NI PCIe-6353 DAQ (National Instruments) and the FOMRI-III optical noise canceling microphone (OptoAcoustics). Each trial was saved in an independent 16-bit WAV audio file.

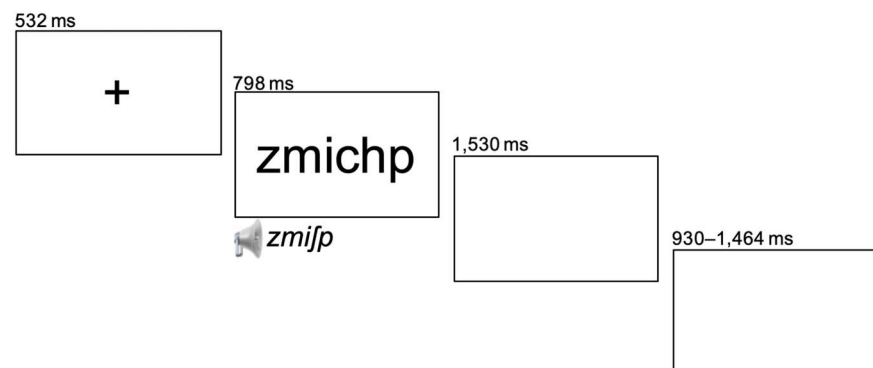


Figure 2. Depiction of an experimental trial.

MRI Data Acquisition

Data were collected on a 3-Tesla Siemens Prisma scanner at the Marseille MRI Center using a 64-channel head coil. Functional images covering the whole brain, including the cerebellum, were acquired during task performance, using a multiband blood oxygen level dependent (BOLD)-sensitive gradient echo planar imaging (EPI) sequence (Moeller et al., 2010) (repetition time [TR] = 1,224 ms, echo time [TE] = 30 ms, flip angle = 66°, 54 slices with a thickness of 2.5 mm, field of view [FOV] = 210 × 210 mm², matrix = 84 × 84, slice thickness = 2.5 mm, multiband factor = 3). Whole-brain anatomical magnetic resonance imaging (MRI) data were acquired using high-resolution structural T1-weighted image (MPRAGE sequence, voxel size = 1 × 1 × 1 mm³, data matrix 256 × 256, TR/TI (inversion time)/TE = 2,300/900/2.98 ms, flip angle = 9°). Prior to functional imaging and to correct functional image for susceptibility induced distortions, a pair of spin-echo EPI sequences with the same spatial parameters as the BOLD images was acquired twice with opposite phase encode directions along the anterior-posterior axis with the following parameters: TR/TE = 7,060/59 ms, voxel size = 2.5 × 2.5 × 2.5 mm³, slices = 54, FOV = 210 × 210 mm². Participants' movements during data acquisition were controlled using Framewise Integrated Real Time MRI Monitoring (Dosenbach et al., 2017).

Behavioral Data Processing and Analyses

All syllable productions were transcribed and labeled as incorrect if they contained insertions, omissions, hesitations, or self-repairs, if they were impartial or missing, or if a syllable was pronounced as two or more syllables. They were labeled as correct in all other cases.

Image Processing and Analyses

Results derive from preprocessing, using fMRIPrep Version 20.2.0 (Esteban et al., 2019; Esteban, Markiewicz, Goncalves, et al., 2020), which is based on Nipype Version 1.5.1 (Esteban, Markiewicz, Burns, et al., 2020; Gorgolewski et al., 2011). A singularity image for fMRIPrep was used and it was run on a cluster of the Mesocentre platform.

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4Bias-FieldCorrection (Tustison et al., 2010), distributed with ANTs Version 2.3.3 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and gray matter (GM) was performed on the brain-extracted T1w using FAST (FSL 5.0.9; Zhang et al., 2001). Volume-based spatial normalization to one standard space was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009), TemplateFlow ID: MNI152Nlin2009cAsym (Esteban et al., 2022).

Functional data preprocessing

For each of the five BOLD runs found per participant (across all sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging one single-band reference (SBRefs). A B0-non-uniformity map (or *fieldmap*) was estimated based on two (or more) EPI references with opposing

phase-encoding directions, with 3dQwarp (Cox & Hyde, 1997; AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using FLIRT (FSL Version 5.0.9; Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost-function. Co-registration was configured with 9 *df* to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using MCFLIRT (FSL Version 5.0.9; Jenkinson et al., 2002). The BOLD time series were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time series are referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time series were resampled into standard space, generating a *preprocessed BOLD run in MNI152Nlin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS (i.e., spatial standard deviation of successive difference images, e.g., Smyser et al., 2011), and three region-wise global signals. FD was computed using two formulations following Power et al. (2014; absolute sum of relative motions) and Jenkinson et al. (2002; relative root mean square displacement between affines). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor; Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time series (using a discrete cosine filter with 128 s cut-off) for anatomical CompCor (aCompCor). For aCompCor, two probabilistic masks (CSF, WM) were generated in anatomical space. The implementation differs from that of Behzadi et al. (2007). Instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks were subtracted from a mask of pixels that likely contain a volume fraction of GM. This mask was obtained by thresholding the corresponding partial volume map at 0.05, and it ensured components were not extracted from voxels containing a minimal fraction of GM. Finally, these masks were resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). For each CompCor decomposition, the 12 components with the largest singular values were retained. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confounded time series derived from head-motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). All resamplings were performed with a *single interpolation step* by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms`, configured with Lanczos's interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964).

The preprocessed data were imported to and analyzed using the Statistical Parametric Mapping software (SPM12; <https://www.Fil.ion.ucl.ac.uk/spm/software/spm12/>) in MATLAB R2018b (Mathworks Inc., Natick, MA). Smoothing was performed with an isotropic Gaussian kernel (full-width at half-maximum = 5 mm). Regressors of no interest included 24 head movement regressors, the mean signal of the CSF and WM mask, and 24 aCompCor regressors related to CSF and WM.

For the univariate whole brain analysis, we created a general linear model (GLM) for each participant. The GLM included, for each of the five runs, the following regressors of interest: AnyLoad_Incorrect (corresponding to all overtly committed errors), HighLoad_Correct (corresponding to the correctly produced phonotactically illegal pseudowords), LowLoad_Correct (corresponding to the correctly produced phonotactically legal pseudowords). Regressors of interest were convolved with the canonical hemodynamic response function. Functional data were filtered with a 128 s high-pass filter. For each participant, we estimated the two contrasts: AnyLoad_Incorrect (corresponding to all overtly committed errors) versus HighLoad_Correct and LowLoadCorrect (corresponding to all correctly produced pseudo words) [2 -1 -1] and HighLoad_Correct (corresponding to the correctly produced phonotactically illegal pseudowords) versus LowLoad_Correct (corresponding to the correctly produced phonotactically legal pseudowords) [0 1 -1]. These two contrasts were then explored at the group level with a random effect analysis. All statistical comparisons were performed with a voxelwise threshold of $p < 0.001$ and a cluster extent threshold of 25 voxels.

For the univariate analysis of ROIs, seven anatomical ROIs were created based on the previous literature (Table 1). Spheres with a Montreal Neurological Institute (MNI) coordinates center and a 10 mm radius were created using the MarsBar SPM toolbox (Brett et al., 2002) and for each participant a mask was created by keeping only voxels within each sphere whose GM probability content (as determined by fMRIPrep segmentation) was higher than 0.2. For a given ROI mask and on the basis of unsmoothed functional images, we extracted the mean of the beta value differences between the contrasted experimental conditions as specified above. For each ROI, we performed one-tailed one-sample t tests comparing the distribution of the beta value differences to the null hypothesis (no difference). Bonferroni correction was used to correct for multiple comparisons.

RESULTS

Out of the 10,800 syllable productions of all participants, 3,548 were marked incorrect (32.8%, mean standard error [MSE] 0.0, SD 0.5), of which 2,655 (49.2%, i.e., 2,655/5,400, MSE 0.0, SD 0.5) were found in the high monitoring load condition and 893 (16.5%, i.e.,

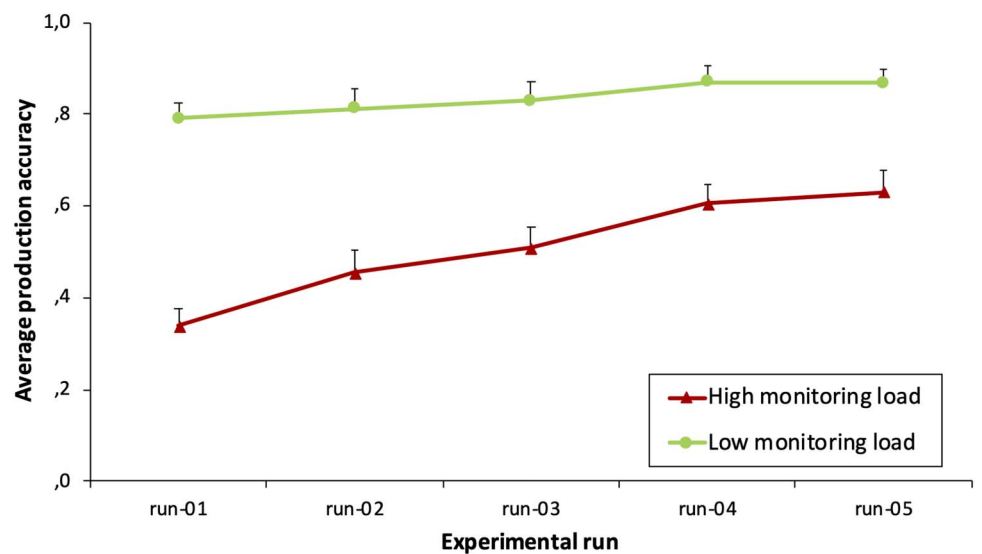


Figure 3. Average production accuracy over the five experimental runs in each monitoring load condition. Error bars represent standard errors of the mean.

893/5,400, *MSE* 0.0, *SD* 0.4) in the low monitoring load condition. This validates that the syllables containing new consonant clusters were more error prone and required more monitoring. A generalized linear mixed model further showed that an interaction between the monitoring load and the number of times an item had been pronounced plays a significant

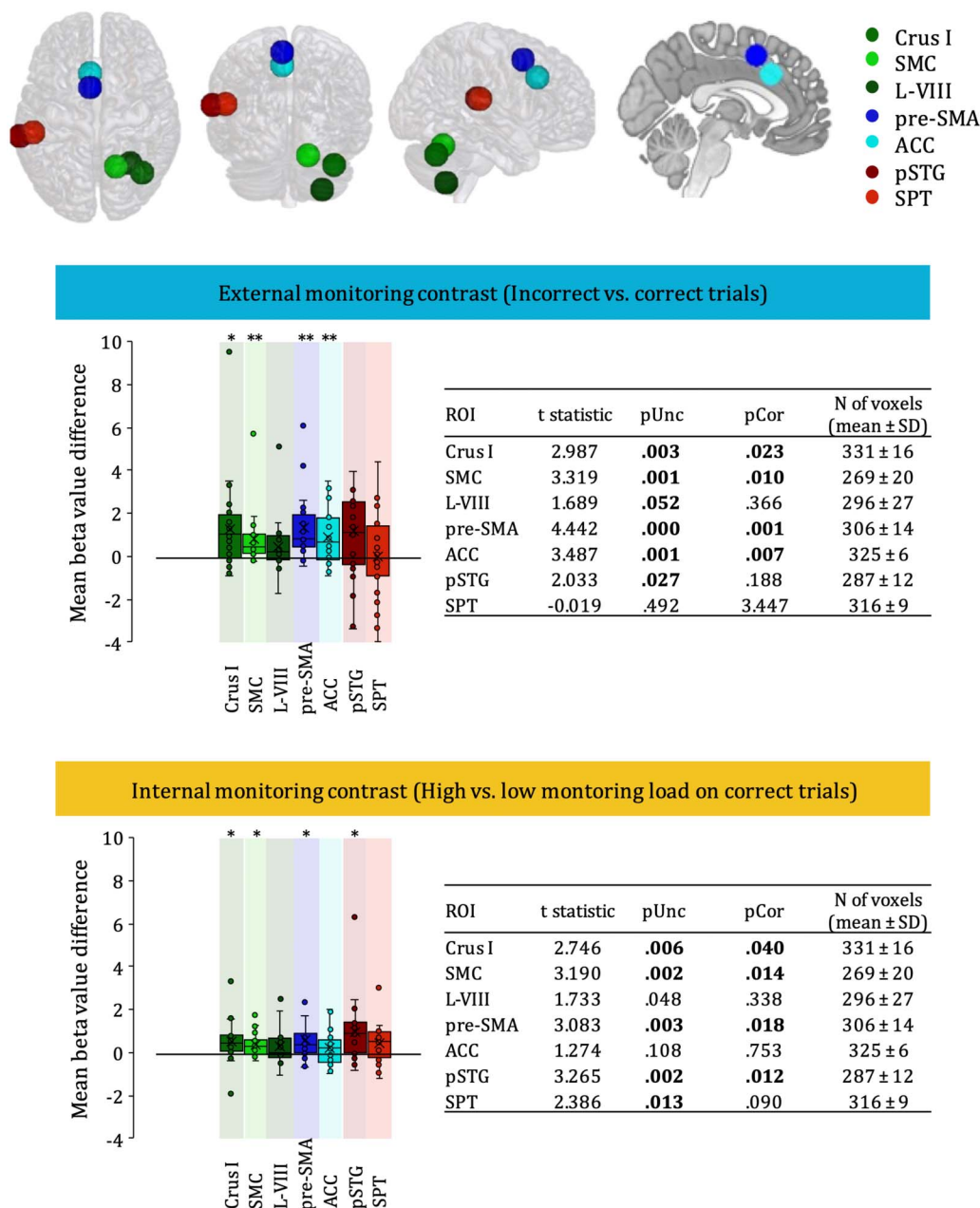


Figure 4. Mean beta value differences in the predefined region of interest (ROI). Top panel represents the locations of the seven predefined regions of interest (ROIs): Crus I, SMC, L-VIII, pre-SMA, ACC, pSTG, and SPT. Middle panel and bottom panel represent the internal monitoring contrast and the external monitoring contrast, respectively. ROIs in medial frontal cortex are represented with blue tones, ROIs in the cerebellum with green tones, and ROIs in temporo-parietal cortex with red tones. The asterisks indicate significant effects <0.05 (*) or <0.005 (**) after applying Bonferroni correction for multiple comparisons. SMC = superior medial cerebellum, L-VIII = lobule VIII, pre-SMA = pre-supplementary motor area, ACC = anterior cingulate cortex, pSTG = posterior superior temporal gyrus, and SPT = Sylvian fissure at the temporo-parietal boundary.

Table 2. Results of the whole-brain analyses of the BOLD response of the (A) external and (B) internal monitoring contrasts.

Region label	Cluster extent	<i>t</i> value of the local maximum	MNI coordinates		
			<i>x</i>	<i>y</i>	<i>z</i>
A. External monitoring (errors vs. correct trials)					
L precentral gyrus	760	6.27	-56	8	36
L postcentral gyrus	73	5.64	-69	-15	24
L supplementary motor area	430	6.18	-2	5	59
L middle temporal gyrus	99	5.55	-62	-10	-1
L insula/inferior frontal gyrus, triangular part	32	4.99	-36	25	2
L inferior parietal gyrus	124	5.62	-36	-45	39
L superior occipital gyrus	96	4.92	-22	-68	39
L inferior occipital gyrus	170	6.05	-46	-68	-16
R inferior occipital gyrus	27	4.58	28	-98	-11
L calcarine fissure	92	6.24	-6	-98	-8
R calcarine fissure	240	5.59	21	-68	6
R cerebellum (VI)	414	6.46	26	-65	-24
R cerebellum (VI)	27	5.51	8	-72	-14
R cerebellum (VIII)	145	5.40	24	-72	-54
B. Internal monitoring (phonotactically illegal vs. legal items on correct trials)					
L inferior frontal gyrus, triangular part	34	5.27	-36	18	9
L supplementary motor area	200	6.13	-6	0	64
L superior temporal gyrus	1120	7.36	-64	-18	9
R temporal pole (superior temporal gyrus)	247	7.91	61	2	-8
L superior parietal gyrus	51	5.12	-16	-78	59
L inferior parietal gyrus	163	5.79	-32	-45	44
L middle occipital gyrus	63	5.26	-26	-72	32
L inferior occipital gyrus	104	5.69	-44	-68	-6
L calcarine fissure	662	6.23	-16	-68	9
L lingual gyrus	39	4.38	-14	-55	-1
R cerebellum (VI)	90	5.50	28	-65	-21
R cerebellum (VIII)	154	5.67	21	-72	-54

Note. Local maxima of blood oxygen level dependent (BOLD) response separated by >8 mm. Regions were automatically labeled using the automated anatomical labeling atlas 3 (Rolls et al., 2020) in SPM. *X*, *y*, *z* = MNI coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively. All peaks are significant at a voxelwise threshold of $p < 0.001$ (extent threshold = 25 voxels) and at a cluster threshold of $p < 0.05$ with a false discovery rate correction for multiple comparisons. L = left, R = right, MNI = Montreal Neurological Institute.

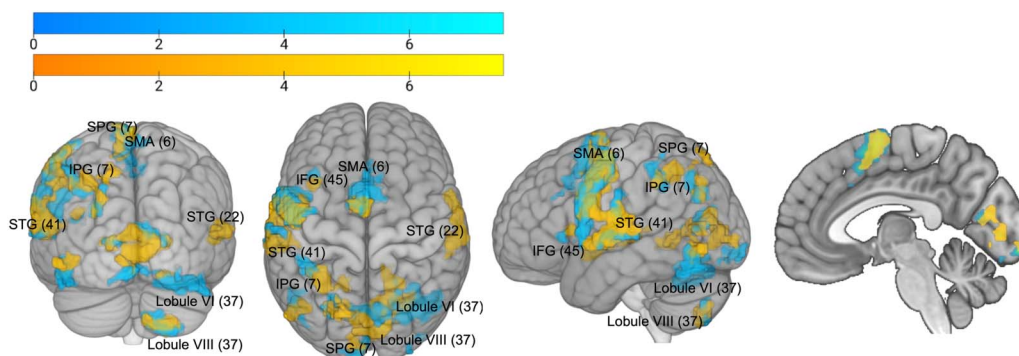


Figure 5. Results of the whole brain analysis. BOLD response of internal monitoring (high vs. low monitoring load for correct trials, yellow) and external monitoring (errors vs. correct trials, blue) contrasts. Overlap across contrasts is visible as green. Statistical *t* maps are overlaid on an MNI template using a voxelwise threshold of $p < 0.001$ and an extent threshold of 25 voxels. Numbers in brackets represent Brodmann areas. IFG = inferior frontal gyrus, IPG = inferior parietal gyrus, SMA = supplementary motor area, SPG = superior parietal gyrus, STG = superior temporal gyrus.

role in predicting production accuracy ($p < 0.001$) and indicates more learning in the high monitoring load condition. Finally, the two conditions differed even in the last experimental run ($p = 2.84e^{-40}$), indicating that the monitoring load remained high for the new consonant combinations throughout the experiment. The results are summarized in Figure 3.

Using an ROI approach, we examined the beta value differences in seven predefined regions located in the cerebellum, medial frontal, temporal, and temporo-parietal cortices (Table 1) in two contrasts of interest, namely, external monitoring, contrasting incorrect trials with the correct ones, and internal monitoring, contrasting high and low monitoring load on correctly produced trials (Figure 4). Two cerebellar regions, right Crus I and right SMC, were involved both in the contrast targeting external and internal monitoring of speech sequence production (Crus I, internal monitoring: $p = 0.04$, external monitoring: $p = 0.02$; SMC, $p = 0.01$ for each contrast). This was also the case for the pre-SMA in the left medial frontal cortex ($p = 0.02$ in internal and 0.001 in external monitoring). Internal monitoring was further linked to the left pSTG ($p = 0.01$), while external monitoring was linked to the left ACC ($p = 0.007$).

To examine the specificity of the findings from the ROI analyses, we also conducted a whole-brain analysis (Table 2 and Figure 5). After correcting for multiple comparisons, both

Table 3. Summary of the different analyses conducted.

	Analysis	Contrasted variables	Purpose
Behavioral data	Generalized linear mixed model on production accuracy	High vs. low load monitoring items over five runs	Validate monitoring load assumption underlying imaging contrast
fMRI data	Univariate ROI analysis on BOLD response	Errors vs. correct trials	Index external monitoring
		Correct high vs. correct low monitoring load trials	Index internal monitoring
	Univariate whole brain analysis on BOLD response	Errors vs. correct trials Correct high vs. correct low monitoring load trials	Assess specificity of ROI findings for external and internal monitoring contrasts

Note. fMRI = functional magnetic resonance imaging; ROI = region of interest; BOLD = blood oxygen level dependent.

monitoring contrasts revealed activation in cerebellar lobules VI and VIII, as well as SMA in medial frontal cortex and BA22 in temporal cortex. The internal monitoring contrast further confirmed BA 41 activation in the temporal cortex. Both the internal and the external monitoring contrasts also revealed significant clusters in the parietal cortex as well as in the insula. Table 3 summarizes all the analyses that were carried out.

DISCUSSION

The aim of the current study was to investigate the cortico-cerebellar correlates of error monitoring in the speech production of novel syllable sequences. A postlexical variable, phonotactic legality, was manipulated in two conditions that differed in their degree of required monitoring. The critical condition consisted of novel syllable sequences that had to be learned, hence were heavily monitored. The baseline condition consisted of known syllable sequences, hence required less monitoring. Contrasting these two conditions for correct trials provided an index of internal monitoring at the postlexical level. Contrasting errors with correct production indexed external monitoring. In the following, we first summarize the results that were obtained for each broad anatomical ROI and then discuss the implications of these results regarding the relevant variables driving the recruitment of different parts of the cortico-cerebellar monitoring network.

Transmodal and Sensory-Motor Driven Monitoring in Distinct Cerebellar Regions Are Both Triggered in Novel Sound Sequence Production

An important goal of the present study was to better understand how distinct cerebellar regions contribute to speech error monitoring. We were interested in highlighting different cerebellar subcompartments as a proxy for internal modeling. Specifically, we contrasted the hypothesis that cerebellar motor regions are engaged in monitoring of motor behavior (including speech motor aspects) and cerebellar cognitive regions are engaged only in cognition (including cognitive aspects of speech), with the idea that any aspect of speech production would trigger activation of the cognitive or transmodal parts of the cerebellum. In line with the latter, an ROI known to be implicated in a broad number of tasks and cognitive subskills, Crus I, was differentially activated in both contrasts. However, an ROI in the motor region SMC was also differentially activated in both contrasts. Furthermore, in the whole brain analysis, differential activation of cerebellar motor regions lobules VI and VIII, not far from the ROI in the lobule VIII, was observed for both internal and external monitoring. The presence of these motor regions for both internal and external monitoring might indicate sensorimotor driven, hierarchically subordinate speech motor control in addition to the hierarchically superior performance monitoring carried out in Crus I. SMC activation was previously reported for external, but not internal, monitoring (Runnqvist et al., 2021) in a lexical task. However, as the number of errors in that study did not allow distinguishing between fluent and disfluent errors, activation of SMC was likely due to articulatory errors, fitting well with the idea that this region engages in controlling speech motor implementation (e.g., Ackermann et al., 1992). Alternatively or in addition, SMC as well as the motor regions lobules VIII and VI could play a slightly different (not only hierarchically subordinate) function. For instance, the SMC is considered to be involved in the generation of proprioceptive input during internal modeling (as opposed to predictive input through Crus I; e.g., Bosco & Poppele, 2001; Tanaka et al., 2020). Yet an alternative account can be found in the directions into velocities of articulators (DIVA) model, where the superior lateral cerebellum that Crus I belongs to contributes to sensory predictions that form auditory and somatosensory targets. In turn, projections from the SMC are thought to

contribute to precisely timed feedforward commands as evidenced by speech motor deficits such as ataxic dysarthria (e.g., Golfinopoulos et al., 2010). Similarly, lobule VIII has been linked to the implementation of motor sequences and to efficiency in the representation of articulatory patterns as opposed to learning a novel sequence (e.g., Orban et al., 2010; Rauschecker et al., 2008). Interestingly, learning a new motor sequence, thus taxing monitoring heavily, has been found to also engage lobule VI (e.g., Orban et al., 2010). Lobule VI was also differentially activated (though left-lateralized) for internal monitoring of a lexical variable during speech production in addition to Crus I (Runnqvist et al., 2021). A cluster identified by the whole brain analysis peaked in lobule VI, though it likely also comprised the ROI in Crus I. Collectively, these studies suggest that lobule VI might be more cognitive or transmodal than lobule VIII or SMC. Regardless, globally the current results fit the idea of two hierarchically distinct monitoring types being carried out through the cerebellum: (speech) overarching performance monitoring in Crus I (and lobule VI) and subordinate (speech) motor control in SMC (and lobule VIII).

Sensorimotor Driven Feedback Monitoring in the Medial Frontal Cortex

We further looked into medial frontal (pre-SMA, ACC) brain regions to contrast conflict-based monitoring with sensorimotor driven monitoring of articulatory representations or sensory feedback. Here, a conflict-based account of monitoring would predict the involvement of medial frontal structures, in particular the ACC, for both internal and external monitoring. Differential pre-SMA activation was observed for both internal and external monitoring, but the ROI in the ACC was only activated for external monitoring. Regarding the pre-SMA, the whole brain analysis largely confirmed this pattern by revealing significant clusters peaking in the SMA, but presumably comprising the ROI in pre-SMA for both internal and external monitoring (see Figure 5). This pattern is both consistent with the idea of conflict-based monitoring and with the notion that the pre-SMA is involved when motor-related difficulties arise. In favor of the latter, the peak of the external monitoring contrast was slightly more anterior (closer to pre-SMA) compared to the internal monitoring contrast (Figure 5), suggesting a possible dissociation in the relative contribution of pre-SMA and SMA to internal and external monitoring. Concerning the ACC, the present findings are more consistent with a role of the targeted cingulate region in processing feedback as indicated by overt errors (as opposed to performing conflict-based monitoring). This is in line with the ACC exclusively engaging in external monitoring in Runnqvist et al. (2021), who proposed an integrative account of ACC activations for error and feedback monitoring following the proposed vocal control network put forward by Loh et al. (2020). These authors proposed that, across primates, area 44 is in charge of cognitive control of orofacial and non-speech vocal responses, and the mid-cingulate cortex is in charge of analyzing vocal non-speech feedback, driving response adaptation. Furthermore, the cognitive control of human-specific vocalization would require additional recruitment of area 45 and the pre-SMA. In this framework, it would not be the conflict that generates activation in the ACC and pre-SMA, but rather the feedback provided through the articulated error. While the present ACC pattern fits well with this explanation, the pre-SMA pattern would require assuming that the internal monitoring contrast reflects some form of sensory feedback, such as pre-articulatory proprioception. Further supporting the hypothesis of a vocal feedback control network here is activation in BA 45 (see Figure 5) just as in Runnqvist et al. (2021). Thus, the current results support the notion that the medial frontal cortex is involved in sensorimotor driven feedback monitoring. Nevertheless, a caveat to this conclusion is the considerable size and functional heterogeneity of the medial frontal cortex. Due to this, a better understanding of its contributions to monitoring

will likely require both the consideration of a larger set of anatomical subregions and the use of functional localizers.

Predictive Monitoring Supported by Temporo-Parietal Cortices

Finally, we looked into temporo-parietal (pSTG, SPT) brain regions to contrast internal modeling of speech with sensorimotor driven monitoring by auditory feedback. The ROI approach only revealed a differential pSTG activation in the internal monitoring contrast. This result differs from previous results that reported pSTG activation in conditions taxing external but not internal monitoring. However, when completing this picture with the results of the whole brain analysis, one can see that while only the internal monitoring contrast results in differential activation of BA 41, both monitoring contrasts engage BA22 (though with different laterality). A recent meta-analysis (Meekings & Scott, 2021) indicates that active regions in speech error monitoring often differ across studies, and consequently renders the functional role of the pSTG controversial. Clearly, the present results contribute to this impression of disparity, but at the same time also indicate that the temporal cortex engages in error monitoring. A systematic research approach seems necessary to increase a better understanding of the relevant variables governing the location of such monitoring related activations. The whole brain analysis further revealed a significant cluster in the inferior parietal gyrus in both contrasts, possibly comprising part of the SPT ROI. Moreover, a cluster in the superior parietal gyrus was observed in the contrast of internal monitoring. Besides possibly comprising a part of SPT, these parietal clusters are of interest because the parietal cortex has been linked to motor imagery and the retrieval of a stored internal model (e.g., Blakemore & Sirigu, 2003). A functional loop between the parietal cortex and the cerebellum, estimating the current status of the motor system throughout movement execution and allowing for predictive monitoring, has been proposed (e.g., Blakemore & Sirigu, 2003; Ito, 2008; Sokolov et al., 2017). Thus, the combined but differential activation of temporal, parietal, and cerebellar regions for internal and external monitoring suggests that their involvement goes beyond sensory feedback and speech comprehension and is more in line with proposals linking these regions to auditory and somatosensory targets and predictive internal modeling-like mechanisms in speech production (e.g., Guenther et al., 2006; Hickok, 2012, 2014; Tourville & Guenther, 2011).

Both Internal and External Monitoring and Position Along the Cognitive-to-Motor Continuum Are Relevant in Eliciting Differential Brain Activation

Given the current results, a substantial overlap between internal and external monitoring presents that differs from prior results on the lexical level of internal monitoring (e.g., Runnqvist et al., 2021). This indicates that to some extent, the relevant variable triggering the recruitment of certain brain regions is related to the position of the representation being monitored along the cognitive-to-motor continuum (rather than to whether it takes place during speech planning or upon articulation). However, as differentiation shows, this also indicates that internal and external monitoring might play a role in the use of brain regions sustaining feedback-based mechanisms (see *Sensorimotor Driven Feedback Monitoring in the Medial Frontal Cortex* above). How do the current results fit the inherent assumption of existing speech production models that monitoring relies on a single overarching mechanism applied during different processing stages, before and after articulation (e.g., Levelt et al., 1999; Nozari et al., 2011; Pickering & Garrod, 2013)? The present study shows that this assumption is correct to some extent. Across all contrasts of interest, we obtained evidence for monitoring presumably carried out by predictive internal modeling (cerebellar Crus I) and possibly linked to temporal and parietal cortex (e.g., Clower et al., 2001; Glickstein, 2000; Stockert et al., 2021).

However, when also considering how these findings may fit related prior literature, some of our observations could indicate a sensorimotor driven type of monitoring or control in motor regions of the cerebellum and in the medial frontal cortex. As most models of speech production agree that not all levels of speech are directly linked to sensorimotor aspects (e.g., Dell, 1988; Levelt et al., 1999; but see Fairs et al., 2021, and Strijkers, 2016, for an account of parallel and distributed processing in speech production), it is perhaps not surprising that levels directly connected to sensorimotor aspects of speech, capitalize on monitoring mechanisms based on feedback from these properties (i.e., auditory and proprioceptive feedback). An open question to be addressed in future research is whether the coexistence of an overarching and sensorimotor driven monitoring indicates complementary functions of the brain regions involved (monitoring different representations or triggered to deal with specific needs such as repairing or compensating for overt errors), or rather a hierarchical relationship ranging from implementation control to performance monitoring across nodes of a functionally coupled monitoring network.

ACKNOWLEDGMENTS

Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources. Work for this study was performed in the IRM-INT Center (UMR 7289, CNRS–Aix-Marseille University), platform member of France Life Imaging network. This work, carried out within the Institute of Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government (France 2030), managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

FUNDING INFORMATION

Elin Runnqvist, French National Agency for Research, Award ID: ANR-18-CE28-0013. ILCB, French National Agency for Research, Award ID: ANR-16-CONV-000X/ANR-17-EURE-00XX. France Life Imaging, French National Agency for Research, Award ID: ANR-11-INBS-0006.

AUTHOR CONTRIBUTIONS

Snežana Todorović: Data curation: Lead; Formal analysis: Equal; Visualization: Lead; Writing – original draft: Supporting. **Jean-Luc Anton:** Formal analysis: Equal; Methodology: Supporting; Supervision: Supporting; Writing – original draft: Supporting. **Julien Sein:** Formal analysis: Supporting; Methodology: Supporting; Writing – original draft: Supporting. **Bruno Nazarian:** Methodology: Supporting; Software: Lead; Writing – original draft: Supporting. **Valérie Chanoine:** Data curation: Supporting; Formal analysis: Supporting; Writing – original draft: Supporting. **Birgit Rauchbauer:** Data curation: Supporting; Writing – original draft: Supporting. **Sonja A. Kotz:** Conceptualization: Supporting; Formal analysis: Supporting; Methodology: Equal; Supervision: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting. **Elin Runnqvist:** Conceptualization: Equal; Data curation: Supporting; Formal analysis: Supporting; Funding acquisition: Lead; Methodology: Equal; Project administration: Lead; Supervision: Equal; Writing – original draft: Lead; Writing – review & editing: Lead.

DATA AVAILABILITY STATEMENT

The analysis code and data related to this article are openly available at OpenNeuro (<https://doi.org/10.18112/openneuro.ds004597.v2.0.0>; OpenNeuro Accession Number: ds004597; <https://openneuro.org/datasets/ds004597/versions/2.0.0>).

REFERENCES

- Ackermann, H., Vogel, M., Petersen, D., & Poremba, M. (1992). Speech deficits in ischaemic cerebellar lesions. *Journal of Neurology*, 239(4), 223–227. <https://doi.org/10.1007/BF00839144>, PubMed: 1597689
- Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, 1076(1), 129–143. <https://doi.org/10.1016/j.brainres.2005.11.104>, PubMed: 16480694
- Argyropoulos, G. P. (2016). The cerebellum, internal models and prediction in “non-motor” aspects of language: A critical review. *Brain and Language*, 161, 4–17. <https://doi.org/10.1016/j.bandl.2015.08.003>, PubMed: 26320734
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>, PubMed: 17659998
- Behzadi, Y., Restom, K., Liu, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>, PubMed: 17560126
- Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *NeuroReport*, 12(9), 1879–1884. <https://doi.org/10.1097/00001756-200107030-00023>, PubMed: 11435916
- Blakemore, S.-J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research*, 153(2), 239–245. <https://doi.org/10.1007/s00221-003-1597-z>, PubMed: 12955381
- Bosco, G., & Poppele, R. E. (2001). Proprioception from a spinocerebellar perspective. *Physiological Reviews*, 81(2), 539–568. <https://doi.org/10.1152/physrev.2001.81.2.539>, PubMed: 11274339
- Brett, M., Anton, J. L., Valabrègue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage*, 16(2; S1), Abstract 497.
- Buckner, R. L. (2013). The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron*, 80(3), 807–815. <https://doi.org/10.1016/j.neuron.2013.10.044>, PubMed: 24183029
- Clower, D. M., West, R. A., Lynch, J. C., & Strick, P. L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *Journal of Neuroscience*, 21(16), 6283–6291. <https://doi.org/10.1523/JNEUROSCI.21-16-06283.2001>, PubMed: 11487651
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4–5), 171–178. [https://doi.org/10.1002/\(SICI\)1099-1492\(199706/08\)10:4/5<171::AID-NBM453>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L), PubMed: 9430344
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27(2), 124–142. [https://doi.org/10.1016/0749-596X\(88\)90070-8](https://doi.org/10.1016/0749-596X(88)90070-8)
- Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11), 423–431. [https://doi.org/10.1016/S1364-6613\(00\)01537-0](https://doi.org/10.1016/S1364-6613(00)01537-0), PubMed: 11058820
- de Zubicaray, G. I., Williams, S. C., Wilson, S. J., Rose, S. E., Brammer, M. J., Bullmore, E. T., Simmons, A., Chalk, J. B., Semple, J., Brown, A. P., Smith, G. A., Ashton, R., & Doddrell, D. M. (1998). Prefrontal cortex involvement in selective letter generation: A functional magnetic resonance imaging study. *Cortex*, 34(3), 389–401. [https://doi.org/10.1016/S0010-9452\(08\)70762-7](https://doi.org/10.1016/S0010-9452(08)70762-7), PubMed: 9669104
- Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, 46(1), 39–46. <https://doi.org/10.1016/j.neuroimage.2009.01.045>, PubMed: 19457380
- D’Mello, A. M., Gabrieli, J. D., & Nee, D. E. (2020). Evidence for hierarchical cognitive control in the human cerebellum. *Current Biology*, 30(10), 1881–1892. <https://doi.org/10.1016/j.cub.2020.03.028>, PubMed: 32275880
- Dosenbach, N. U., Koller, J. M., Earl, E. A., Miranda-Dominguez, O., Klein, R. L., Van, A. N., Snyder, A. Z., Nagel, B. J., Nigg, J. R., Nguyen, A. L., Wesevich, V., Greene, D. J., & Fair, D. A. (2017). Real-time motion analytics during brain MRI improve data quality and reduce costs. *NeuroImage*, 161, 80–93. <https://doi.org/10.1016/j.neuroimage.2017.08.025>, PubMed: 28803940
- Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, 36(3), 488–499. <https://doi.org/10.3758/BF03195595>, PubMed: 15641437
- Esteban, O., Ciric, R., Nielson, D., & Goncalves, M. (2022). *TemplateFlow ID: MNI152Nlin2009cAsym* [Software]. Github. <https://github.com/templateflow/tpl-MNI152Nlin2009cAsym>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Concalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111–116. <https://doi.org/10.1038/s41592-018-0235-4>, PubMed: 30532080
- Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., ... Ghosh, S. (2020). *Nipy/Nipype* (Version 1.5.1) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.4035081>
- Esteban, O., Markiewicz, C. J., Goncalves, M., DuPre, E., Kent, J. D., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2020). *fMRIPrep: A robust preprocessing pipeline for functional MRI* (Version 20.2.0) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.4055773>
- Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, 2(3), Article tgab040. <https://doi.org/10.1093/texcom/tgab040>, PubMed: 34296185
- Fiez, J. A. (1996). Cerebellar contributions to cognition. *Neuron*, 16(1), 13–15. [https://doi.org/10.1016/S0896-6273\(00\)80018-5](https://doi.org/10.1016/S0896-6273(00)80018-5), PubMed: 8562076
- Fonov, V. S., Evans, A. C., McKinstry, R. C., Almlí, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47(1), S102. [https://doi.org/10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5)
- Ganushchak, L. Y., & Schiller, N. O. (2008a). Brain error-monitoring activity is affected by semantic relatedness: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, 20(5), 927–940. <https://doi.org/10.1162/jocn.2008.20514>, PubMed: 18201131

- Ganushchak, L. Y., & Schiller, N. O. (2008b). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, 39(1), 395–405. <https://doi.org/10.1016/j.neuroimage.2007.09.001>, PubMed: 17920932
- Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage*, 126, 96–105. <https://doi.org/10.1016/j.neuroimage.2015.11.037>, PubMed: 26608243
- Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, 3(1), Article 17. <https://doi.org/10.5334/joc.81>, PubMed: 32944680
- Glickstein, M. (2000). Functional localisation in the cerebral cortex and cerebellum: Lessons from the past. *European Journal of Morphology*, 38(5), 291–300. <https://doi.org/10.1076/ejom.38.5.291.7364>, PubMed: 11151041
- Golfopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, 52(3), 862–874. <https://doi.org/10.1016/j.neuroimage.2009.10.023>, PubMed: 19837177
- Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, 5, 13. <https://doi.org/10.3389/fninf.2011.00013>, PubMed: 21897815
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>, PubMed: 19573611
- Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum. *Elife*, 7, Article e36652. <https://doi.org/10.7554/eLife.36652>, PubMed: 30106371
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301. <https://doi.org/10.1016/j.bandl.2005.06.001>, PubMed: 16040108
- Hansen, S. J., McMahon, K. L., & de Zubicaray, G. I. (2019a). Neural mechanisms for monitoring and halting of spoken word production. *Journal of Cognitive Neuroscience*, 31(12), 1946–1957. https://doi.org/10.1162/jocn_a_01462, PubMed: 31418336
- Hansen, S. J., McMahon, K. L., & de Zubicaray, G. I. (2019b). The neurobiology of taboo language processing: fMRI evidence during spoken word production. *Social Cognitive and Affective Neuroscience*, 14(3), 271–279. <https://doi.org/10.1093/scan/nsz009>, PubMed: 30715549
- Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42(2), 113–157. <https://doi.org/10.1006/cogp.2000.0744>, PubMed: 11259106
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145. <https://doi.org/10.1038/nrn3158>, PubMed: 22218206
- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29(1), 2–20. <https://doi.org/10.1080/01690965.2013.834370>, PubMed: 24489420
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, 8(9–10), 2379–2382. <https://doi.org/10.1097/00001756-199707070-00055>, PubMed: 9243644
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>, PubMed: 15037128
- Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4), 304–313. <https://doi.org/10.1038/nrn2332>, PubMed: 18319727
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>, PubMed: 12377157
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6), PubMed: 11516708
- Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, 49(9), 2449–2461. <https://doi.org/10.1016/j.cortex.2012.12.012>, PubMed: 23318086
- Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, 1(1), 76–85. <https://doi.org/10.1137/0701007>
- Lesage, E., Hansen, P. C., & Miall, R. C. (2017). Right lateral cerebellum represents linguistic predictability. *Journal of Neuroscience*, 37(26), 6231–6241. <https://doi.org/10.1523/JNEUROSCI.3203-16.2017>, PubMed: 28546307
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4), PubMed: 6685011
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>, PubMed: 11301520
- Loh, K. K., Procyk, E., Neveu, R., Lamberton, F., Hopkins, W. D., Petrides, M., & Amiez, C. (2020). Cognitive control of orofacial motor and vocal responses in the ventrolateral and dorsomedial human frontal cortex. *Proceedings of the National Academy of Sciences*, 117(9), 4994–5005. <https://doi.org/10.1073/pnas.1916459117>, PubMed: 32060124
- McCall, J. D., DeMarco, A. T., Mandal, A. S., Fama, M. E., van der Stelt, C. M., Lacey, E. H., Laks, A. B., Snider, S. F., Friedman, R. G., & Turkeltaub, P. E. (2023). Listening to yourself and watching your tongue: Distinct abilities and brain regions for monitoring semantic and phonological speech errors. *Journal of Cognitive Neuroscience*, 35(7), 1169–1194. https://doi.org/10.1162/jocn_a_02000, PubMed: 37159232
- McCloskey, D. I. (1981). Corollary discharges: Motor commands and perception. In *Comprehensive physiology* (pp. 1415–1447). American Physiological Society.
- Medina, J. F., & Mauk, M. D. (2000). Computer simulation of cerebellar information processing. *Nature Neuroscience*, 3(11), 1205–1211. <https://doi.org/10.1038/81486>, PubMed: 11127839
- Meekings, S., & Scott, S. K. (2021). Error in the superior temporal gyrus? A systematic review and activation likelihood estimation meta-analysis of speech production studies. *Journal of Cognitive Neuroscience*, 33(3), 422–444. https://doi.org/10.1162/jocn_a_01661, PubMed: 33326327
- Middleton, F. A., & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, 266(5184), 458–461. <https://doi.org/10.1126/science.7939688>, PubMed: 7939688

- Moberget, T., Gullersen, E. H., Andersson, S., Ivry, R. B., & Endestad, T. (2014). Generalized role for the cerebellum in encoding internal models: Evidence from semantic processing. *Journal of Neuroscience*, *34*(8), 2871–2878. <https://doi.org/10.1523/JNEUROSCI.2264-13.2014>, PubMed: 24553928
- Moberget, T., & Ivry, R. B. (2016). Cerebellar contributions to motor control and language comprehension: Searching for common computational principles. *Annals of the New York Academy of Sciences*, *1369*(1), 154–171. <https://doi.org/10.1111/nyas.13094>, PubMed: 27206249
- Moeller, S., Yacoub, E., Olfman, C. A., Auerbach, E., Strupp, J., Harel, N., & Ugurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, *63*(5), 1144–1153. <https://doi.org/10.1002/mrm.22361>, PubMed: 20432285
- Motley, M. T., Camden, C. T., & Baars, B. J. (1981). Toward verifying the assumptions of laboratory-induced slips of the tongue: The output-error and editing issues. *Human Communication Research*, *8*(1), 3–15. <https://doi.org/10.1111/j.1468-2958.1981.tb00652.x>
- Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 578–594. [https://doi.org/10.1016/S0022-5371\(82\)90791-5](https://doi.org/10.1016/S0022-5371(82)90791-5)
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™/A lexical database for contemporary French: LEXIQUE™. *L'Année Psychologique*, *101*(3), 447–462. <https://doi.org/10.3406/psy.2001.1341>
- New, B., & Spinelli, E. (2013). Diphones-fr: A French database of diphone positional frequency. *Behavior Research Methods*, *45*(3), 758–764. <https://doi.org/10.3758/s13428-012-0285-y>, PubMed: 23239073
- Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19–35. <https://doi.org/10.1016/j.jml.2017.01.007>
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. <https://doi.org/10.1016/j.cogpsych.2011.05.001>, PubMed: 21652015
- Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, *98*(1), 112–117. <https://doi.org/10.1016/j.bandl.2006.04.006>, PubMed: 16716388
- Okada, K., Matchin, W., & Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin & Review*, *25*(1), 423–430. <https://doi.org/10.3758/s13423-017-1284-x>, PubMed: 28397076
- Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. <https://doi.org/10.1016/j.cognition.2007.02.006>, PubMed: 17407776
- Orban, P., Peigneux, P., Lungu, O., Albouy, G., Breton, E., Laberrenne, F., Benali, H., Maquet, P., & Doyon, J. (2010). The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. *NeuroImage*, *49*(1), 694–702. <https://doi.org/10.1016/j.neuroimage.2009.08.055>, PubMed: 19732838
- Peterburs, J., & Desmond, J. E. (2016). The role of the human cerebellum in performance monitoring. *Current Opinion in Neurobiology*, *40*, 38–44. <https://doi.org/10.1016/j.conb.2016.06.011>, PubMed: 27372055
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. <https://doi.org/10.1017/S0140525X12001495>, PubMed: 23789620
- Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, *39*(4), 375–392. <https://doi.org/10.1177/002383099603900403>
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>, PubMed: 23994314
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (Eds.). (2001). Modulation of movement by the cerebellum: Circuits within the cerebellum. In *Neuroscience* (2nd ed.). Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK10865/>
- Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*, *29*(11), 1231–1242. <https://doi.org/10.1002/hbm.20460>, PubMed: 17948887
- Riès, S., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, *23*(6), 1419–1436. <https://doi.org/10.1162/jocn.2010.21467>, PubMed: 20350170
- Roelofs, A. (2020). Self-monitoring in speaking: In defense of a comprehension-based account. *Journal of Cognition*, *3*(1), Article 18. <https://doi.org/10.5334/joc.61>, PubMed: 32944681
- Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J., & Joliot, M. (2020). Automated anatomical labeling atlas 3. *NeuroImage*, *206*, Article 116189. <https://doi.org/10.1016/j.neuroimage.2019.116189>, PubMed: 31521825
- Runnqvist, E. (2023). Self-monitoring: The neurocognitive basis of error monitoring in language production. In R. J. Hartsuiker & K. Strijkers (Eds.), *Language production* (pp. 168–190). Routledge. <https://doi.org/10.4324/9781003145790-7>
- Runnqvist, E., Bonnard, M., Gauvin, H. S., Attarian, S., Trébuchon, A., Hartsuiker, R. J., & Alario, F.-X. (2016). Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex*, *81*, 203–214. <https://doi.org/10.1016/j.cortex.2016.05.008>, PubMed: 27249802
- Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., Sein, J., Anton, J.-L., Dorokhova, L., Belin, P., & Alario, F.-X. (2021). Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications*, *2*(2), Article tgab038. <https://doi.org/10.1093/texcom/tgab038>, PubMed: 34296182
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughhead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>, PubMed: 22926292
- Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience*, *27*(4), 819–831. https://doi.org/10.1162/jocn_a_00737, PubMed: 25313656

- Severens E., Kühn, S., Hartsuiker, R. J., & Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Social Cognitive and Affective Neuroscience*, 7(4), 431–435. <https://doi.org/10.1093/scan/nsr030>, PubMed: 21609970
- Shergill, S. S., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2000). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry*, 57(11), 1033–1038. <https://doi.org/10.1001/archpsyc.57.11.1033>, PubMed: 11074868
- Smyser, C. D., Snyder, A. Z., & Neil, J. J. (2011). Functional connectivity MRI in infants: Exploration of the functional organization of the developing brain. *NeuroImage*, 56(3), 1437–1452. <https://doi.org/10.1016/j.neuroimage.2011.02.073>, PubMed: 21376813
- Sokolov, A. A., Miall, R. C., & Ivry, R. B. (2017). The cerebellum: Adaptive prediction for movement and cognition. *Trends in Cognitive Sciences*, 21(5), 313–332. <https://doi.org/10.1016/j.tics.2017.02.005>, PubMed: 28385461
- Stockert, A., Schwartze, M., Poeppel, D., Anwander, A., & Kotz, S. A. (2021). Temporo-cerebellar connectivity underlies timing constraints in audition. *Elife*, 10, Article e67303. <https://doi.org/10.7554/eLife.67303>, PubMed: 34542407
- Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, 44(2), 489–501. <https://doi.org/10.1016/j.neuroimage.2008.08.039>, PubMed: 18835452
- Strick, P. L., Dum, R. P., & Fiez, J. A. (2009). Cerebellum and non-motor function. *Annual Review of Neuroscience*, 32, 413–434. <https://doi.org/10.1146/annurev.neuro.31.060407.125606>, PubMed: 19555291
- Strijkers, K. (2016). Can hierarchical models display parallel cortical dynamics? A non-hierarchical alternative of brain language theory. *Language, Cognition and Neuroscience*, 31(4), 465–469. <https://doi.org/10.1080/23273798.2015.1096403>
- Tanaka, H., Ishikawa, T., Lee, J., & Kakei, S. (2020). The cerebro-cerebellum as a locus of forward model: A review. *Frontiers in Systems Neuroscience*, 14, 19. <https://doi.org/10.3389/fnsys.2020.00019>, PubMed: 32327978
- Teghipco, A., Okada, K., Murphy, E., & Hickok, G. (2023). Predictive coding and internal error correction in speech production. *Neurobiology of Language*, 4(1), 81–119. https://doi.org/10.1162/nol_a_00088, PubMed: 37229143
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>, PubMed: 23667281
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>, PubMed: 20378467
- Volfart, A., McMahon, K. L., Howard, D., & de Zubicaray, G. I. (2022). Neural correlates of naturally occurring speech errors during picture naming in healthy participants. *Journal of Cognitive Neuroscience*, 35(1), 111–127. https://doi.org/10.1162/jocn_a_01927, PubMed: 36306259
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882. <https://doi.org/10.1126/science.7569931>, PubMed: 7569931
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>, PubMed: 11293691