Signatures of monitoring in speech and their precursors : a multi-facet outlook.

Lydia Dorokhova

Aix*Marseille université
Socialement engagée

NNT : 2023AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 7 décembre 2023 par

Lydia DOROKHOVA

Signatures of monitoring in speech and their precursors : a multi-facet outlook.

Discipline
Sciences du Langage

École doctorale
ED 356
Cognition, Langage, Education

Laboratoire
Laboratoire Parole et Langage

Composition du jury

Kate WATKINS Full Professor University of Oxford	Rapporteure Présidente du jury
Robert HARTSUIKER Full Professor Ghent University	Rapporteur
Céline AMIEZ Directrice de recherche Université Claude Bernard Lyon 1	Examinatrice
Elin RUNNQVIST Chargée de recherche Aix Marseille Université	Directrice de thèse
Pascal BELIN Professeur des universités Aix Marseille Université	Co-directeur de thèse

Laboratoire Parole et Langage 50 1972 2022
ILCB
anr
int institut de neurosciences de la timone

Affidavit

Je soussignée, DOROKHOVA Lydia, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Runnqvist Elin et Belin Pascal, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Aix-en-Provence le 2/10/2023

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse :

1. 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), tgab038. https://doi.org/10.1093/texcom/tgab038
2. DOROKHOVA, L., MORILLON, B., BAUS, C., BELIN, P., DUBARRY, A.-S., ALARIO, F.-X., & RUNNQVIST, E. (2023). A robust temporal map of speech monitoring from planning to articulation. https://doi.org/10.31234/osf.io/acwqd - preprint

Participation aux conférences et écoles d'été au cours de la période de thèse :

1. 23/04/2022 – Cognitive Neuroscience Society 2022 Conference, San-Francisco, The USA – *EEG signatures of monitoring load and error-related processing in language production reveal dissociable mechanisms.* –Poster (online)
2. 1/06/2022 –International Max Planck Research School Conference - Nijmegen, The Netherlands – *EEG signatures of monitoring load and error-related processing in language production reveal dissociable mechanisms.*– Poster
3. 26/03/2023 – Cognitive Neuroscience Society 2023 Conference, San-Francisco, The USA – *Monitoring : tracing the evolutionary link from movement to cognition.* – Poster
4. 24/10/2023 – The Society for the Neurobiology of Language Conference, Marseille, France – *Linking movement and cognition through the hierarchical build-dup of internal modeling.* – Poster
5. 25/10/2023 – The Society for the Neurobiology of Language Conference, Marseille, France – *Speech monitoring activations in the midcingulate cortex depend on cingulate sulcal morphology.* – Poster

Résumé

La parole est un processus multifacette qui combine rapidement et précisément le traitement cognitif avec les actions motrices. Malgré cette complexité, les erreurs de parole se produisent rarement. Cette précision remarquable peut être attribuée à la présence de l'auto-surveillance (self-monitoring), permettant de détecter et de rectifier rapidement divers types d'erreurs commises ou imminentes. L'objectif principal de cette thèse est d'explorer de manière exhaustive le phénomène de la surveillance de la parole, cherchant à comprendre comment elle fonctionne dans le cadre complexe de la production de la parole et au-delà. La recherche vise à explorer des mécanismes de surveillance distincts, à identifier les moments clés de la mise en œuvre de l'autocontrôle dans la parole, à examiner comment différents niveaux de représentation influencent le processus de surveillance, et à la contextualiser dans le spectre plus large des capacités humaines le long du continuum moteur-cognitif. Dans le premier chapitre, une étude d'IRMf liée à des événements a évalué les rôles de différentes structures cérébrales associées à des cadres théoriques pour les opérations d'auto-surveilance. Les résultats ont indiqué un rôle du cortex temporal dans la détection d'erreurs commises par le biais de la surveillance basée sur la compréhension ; un rôle du cortex frontal médian dans la surveillance basé sur le feedback et un rôle du cervelet dans la modélisation interne de la parole à venir. Un chapitre ultérieur utilisant l'EEG s'est penché sur les aspects temporels du processus de surveillance, en examinant différentes sources de probabilité d'erreur (lexicale et articulatoire-phonétique) ainsi que des erreurs de parole commises. Les résultats ont indiqué que les processus de surveillance interviennent à des stades spécifiques de la production de la parole, en se concentrant sur divers types d'informations pertinentes à chaque étape. Dans le chapitre suivant, une étude d'IRMf a exploré la surveillance générale de l'action en examinant une tâche motrice simple complexifiée par l'introduction d'une règle de surveillance cognitive. L'accent a été mis sur le changement de structures cérébrales associé à l'augmentation des exigences cognitives de la tâche. Les résultats étaient cohérents avec un réseau de contrôle frontal-pariétal-cérébelleux adaptatif et hiérarchique. Enfin, la quatrième étude a examiné la relation entre une spécificité morphologique du cerveau et les activations fonctionnelles dans le cortex cingulaire antérieur (CCA), une zone clé d'intérêt dans la recherche sur la surveillance. Cette étude a révélé que les motifs d'activation déclenchés par les erreurs de parole dépendent des variations sulcales dans le CCA. En résumé, cette exploration multifacette de la surveillance de la parole contribue à notre connaissance de ses multiples mécanismes, de sa dynamique temporelle, de son adaptation aux niveaux de représentation et de sa place le long du continuum des fonctions moteur-cognitives. En utilisant diverses techniques de neuroimagerie et des tâches variées, elle éclaire l'organisation complexe des structures cérébrales impliquées dans le monitoring et sa nature flexible.

Mots clés : monitoring, EEG,fMRI, production de la parole, continuum moteur-cognitif

Abstract

Speech is a multifaceted process that rapidly and precisely combines cognitive processing with motor actions. Despite this complexity, speech errors occur infrequently. This remarkable precision can be attributed to the presence of self-monitoring, allowing to detect and promptly rectify various types of overt or impending errors. The primary focus of this thesis is to comprehensively investigate the phenomenon of speech monitoring, seeking to address how it operates within the complex framework of speech production and beyond. The research aims to explore distinct monitoring mechanisms, identify key moments in the implementation of self-monitoring in speech, examine how different levels of representation influence the monitoring process, and contextualize self-monitoring within the broader spectrum of human capacities along the motor-cognitive continuum. In the first chapter, an event-related fMRI study assessed the roles of different brain structures associated with theoretical frameworks for self-monitoring operations. The findings indicated a role of temporal cortex in detection of overt errors through comprehension-based monitoring; a role of medial frontal cortex in feedback-based control and a role of the cerebellum in the internal modeling of upcoming speech. A subsequent chapter using EEG delved into the temporal aspects of the monitoring process, investigating different sources of error probability (lexical and articulatory-phonetic) as well as overt speech errors. The results indicated that monitoring processes come into play at specific stages of speech production, concentrating on various types of information relevant at each stage. In the next chapter, an fMRI study explored action general monitoring by examining a simple motor task that was complexified by the introduction of a cognitive monitoring rule. The focus was on the shift in brain structures associated with the increase of cognitive demands of the task. The results were consistent with an adaptive and hierarchical fronto-parieto-cerebellar control network. Finally, the fourth study investigated the relationship between a morphological specificity of the brain and functional activations in the anterior cingulate cortex,(ACC), a key area of interest in monitoring research. This study revealed that activation patterns elicited by speech errors depend on sulcal variations in the ACC. In summary, this multifaceted exploration of speech monitoring contributes to our knowledge about its multiple mechanisms, temporal dynamics, adaptability to levels of representation and its place along the continuum of motor-cognitive functions. By employing various neuroimaging techniques and diverse tasks, it sheds light on the complex organization of brain structures involved in monitoring and its flexible nature.

Keywords: monitoring, EEG,fMRI, speech-production, motor-cognitive continuum

A toi, Jérémy,
le plus grand pilier de ma vie,
qui a porté cette thèse sur tes épaules
en me faisant croire qu'elle n'était pas lourde.

Remerciements

La période de ma thèse représente à ce jour plus de 15% de la totalité de ma vie. Comme on fait rarement 2 theses dans une vie cela mérite d'être vu comme une histoire.

Cette histoire a commencé en te rencontrant toi, Elin, peut être la seule, qui à part moi, lira cette page. Depuis le début, j'avais besoin d'une petite chose qui me paraissait très grande : la confiance en moi. Et tu me l'as donnée. C'est ce sentiment de gratitude qui m'a beaucoup porté pendant ces années de thèse et c'est en partie ce sentiment qui m'a accroché lorsque c'était dur. J'ai eu peur de te décevoir et pendant un temps cette peur a constitué un obstacle. J'ai même imprimé certains de tes mails d'encouragements pour les relire afin de me mobiliser. La réalisation de ce manuscrit en est ta réussite. La vulnérabilité que nous avons dévoilée dans les périodes difficiles, nous a aidé à apprendre à nous connaître et à avancer ensemble, en équipe soudée. Je tiens aussi à dire que la dernière année de cette thèse est l'année que j'ai le plus appréciée grâce à cette collaboration. Pour tout cela, et pour tout ce que tu m'as appris aussi bien scientifiquement qu'humainement, je te suis profondément reconnaissante. Tu auras sans doutes d'autres doctorants, mais tu seras pour toujours ma directrice de thèse.

Ensuite cette histoire continue par mon introduction à l'ILCB, la structure dont je suis admirative. Kep Kee Loh, you were the first person I talked to there, and I remember this coincidence being mutual. I admire your passion for science, your hard-working dedication to it, and how accessible you are despite your intelligence. I want to thank you for your support because you provided me with a lot of it during difficult times. I enjoyed my internship in Montréal very much, and even more so, our picnics.

Une autre rencontre majeure pendant ces années écoulées, c'est toi Morgane Peirolo. Je n'ai pas besoin de ma thèse pour te dire tout ce que je pense mais je sais que cela te fera plaisir de voir ton nom ici. Je ne m'attendais pas à me lier aussi fortement et aisément avec quelqu'un sur une période aussi courte. Il fallait que je me reveille à 5h du mat pour arriver a Marseille pour nos sessions de collecte IRMf, mais je garde cela comme un des plus beaux souvenirs parce que j'allais te voir. Merci d'abord d'avoir

accepté de subir cette expérience et surtout merci d'exister. Pour cela tu pourras faire lire ce passage à ta mère, j'en profite ainsi pour la féliciter pour ses excellents muffins.

En parlant des sessions de collecte IRMf, je tiens à te remercier, Jean-Luc Anton. Je te compte parmi les rencontres les plus importantes pour moi. Depuis que je t'ai rencontré, j'ai l'impression que tous les problèmes peuvent se régler simplement par une conversation avec toi. Tu as toujours su me montrer le côté réel des choses en me dénouant de mes émotions et en m'aidant à progresser.

Enfin, ce manuscrit n'aurait probablement pas été le même si le 26 janvier 2023, toi, Benjamin Morillon, ne m'avais pas dit : "Ecris".

Parmi les intemporels, je veux remercier Nadéra Bureau pour tout ce que tu fais déjà en travail 'invisible' et surtout parce que tu m'a apporté beaucoup de soutien et confort pendant ces années. Merci aussi à l'équipe des doctorants LPL et ILCB qui ont fait de cette expérience une joie de vivre. Et plus particulièrement merci à Marie Rebourg et Daria Rousset pour tout ces litres de thé bues ensemble à partager nos joies et nos peines.

Et pour ceux qui sont encore au début de leur chemin : vous êtes forts et capables, ne travaillez pas le weekend et prenez des vraies pauses.

Важно дописать здесь, что эта диссертация существует в большей степени благодаря поддержке моей семьи и друзей. Без моей жизни вне научного круга, я бы уже давно сошла с ума. Жереми, Эстель, мама, папа, Катюша, Инна, спасибо за любовь, поддержку и безграничное понимание и принятие меня во всех моих формах. Я обещаю попытаться наверстать все то время, которое посвятила не вам, а этой работе. Без вас я просто не существую.

Table of Contents

Additional Content

Main content

General Introduction

Monitoring : State of the Art and Outstanding Questions

Humans can perform many tasks flawlessly especially if they practice. For example, we practice walking, manipulating different objects and also speaking every day. Thus, we find ourselves so good in these trite tasks that they appear to look effortless to the point that any error of execution is often spotted and drawn attention to. For example, there are millions of hilarious videos on the Internet where someone's awkward action planning makes them fall. Or also, some slips of the tongue become famous ("Facts are stupid things - stubborn things should I say." R. Reagan), some are questioning ("The decision of one man to launch a wholly unjustified and brutal invasion of Iraq... I mean of Ukraine." G.W. Bush) and some become sensational ("Thank you, I'll take that for my d... baby" Kate Middleton, when sex of her unborn child was not yet known). The reason why so much attention is drawn to a simple error of execution is the rareness of the failure. And the easiest assumption is to consider everyday tasks simple yet it is not the case. Humans coordinate one third of all their muscles to take a single step forward for example, and 100 of muscles need to be coordinated for speaking, all while carrying out other cognitive tasks that make human speech sensed. Consequently, the question is how we are able to perform such complex tasks so efficiently. The question of this dissertation mainly concerns speech and the mechanisms that allow speakers to optimize their speech production performance. Speech is a complex action in which one can distinguish both cognitive and motor components. To accomplish the cognitive components we must find the semantic information that matches an idea we want to express (concept retrieval) and translate it in words (lexical access and selection) and sounds (phonological encoding). For the motor component we must pronounce the sounds correctly, involving the retrieval of an articulatory program and its execution. To make speech flawless both of these components, cognitive and motor, are thought to involve continuous self-monitoring.

This process of self-monitoring is defined as the process of inspecting one's own speech and taking appropriate action when errors are made (Hartsuiker and Kolk 2001).

The most intuitive evidence supporting this phenomenon is derived from observations that speakers detect and correct errors by interrupting themselves (Levelt 1983), and can rather accurately report errors committed (Postma and Noordanus 1996). Beyond these type of elementary observations, a large body of research has made a compelling case that speakers can make use of both an external and internal channel to detect and repair speech errors. A first mean of error detection and subsequent correction process involves relying on the external auditory channel. Similarly to how we listen to others, we also hear ourselves speak, and this auditory input can be compared against the intended sequence. When a mismatch occurs, a correction is initiated. This notion is substantiated by studies such as one in which participants' auditory feedback during a Stroop color-naming task was altered to include an alternative response (Lind et al. 2014). Under ideal timing conditions, most of these substitutions went unnoticed, with 85% of these substitutions being perceived as self-produced. This finding indicates that speakers rely on the auditory feedback of their own utterances to construct the sense of what they say, fitting nicely with the idea that auditory feedback is used to monitor speech production. Nonetheless, empirical evidence suggests that speakers do not solely rely on the external sensory channel. For example, halting an erroneous segment takes place either shortly after the error or approximately 500 ms later. That is, error-to-cutoff times display a bimodal distribution (Nooteboom and Quené 2017). This observation suggests the presence of at least two distinct time points at which monitoring processes take place or interact with the production process. The almost instantaneous interval between the error and the correction for the first group of repairs suggests that the process of error detection and repair had occurred internally prior to the error becoming audible (Hartsuiker and Kolk 2001; Levelt 1983). The second group of repairs suggests the reliance on the external sensory channel. Moreover, in the context of typed speech, a dissociation in monitoring processes was exhibited by skilled typists (Logan and Crump 2010). Text input was manipulated to introduce errors or rectify them. Although participants were asked to report errors, they accepted the illusion caused by the modified text input, reporting artificially induced errors as their own. Interestingly, the manipulation did not affect their typing rate; typists were slower after committing an artificially corrected error, yet maintained a consistent speed when an error was artificially introduced. This post-error slowing phenomenon has also been reported in spoken production, where it was linked to the error detection and interpreted as a form of corrective action, because participants who showed larger slowing made fewer errors (Ganushchak and Schiller

2006). If post-error slowing helps preventing future errors, it hints at the existence of an internal error detection channel that directly exploits the slowing. This hypothesis receives further support from the finding that the slowing also occurs on correct trials which are more error-prone (Freund and Nozari 2018). The evidence supporting the use of an internal channel for monitoring gains further strength from the occurrence of pragmatically inappropriate speech, where taboo words are produced less frequently than expected, even in controlled error protocols (Severens et al. 2011). The same tendency concerns non-words. In error-eliciting tasks such as SLIP (Spoonerisms of Laboratory Induced Predisposition, e.g., Baars et al. 1975) where speakers are phonologically primed to commit an error by inverting the initial sounds of a word pair (e.g. *ball toy* ⇒ *tall boy*), the amount of errors is reduced if the outcome of the switched-initials pair is a non-existent word (e.g. *gold zip* ⇒ *zold gip*) (Runnqvist et al. 2021). Moreover, speakers appear to resort to the internal channel when the external one is unavailable. This deduction is supported by studies where auditory feedback is masked, yet participants still accurately identify the errors they have made under conditions of masked auditory feedback (Nooteboom and Quené 2017; Postma and Kolk 1992). These observations indicate the anticipation of possible undesirable outcomes during the process of speech planning.

In summary, there is a widely accepted consensus that individuals actively monitor their speech for errors, and patterns of speech errors underscore the concurrent existence of both internal and external speech monitoring mechanisms. Consequently, the mechanism(s) underlying the ability to detect and correct speech errors should be at the core of any model accounting for language production. Contrary to this expectation, in what follows, I will first briefly describe how the three existing classes of language production models fall short of accounting for speech error patterns and monitoring phenomena solely on the basis of their constitutive components. Next, I will focus on three models designed explicitly to account for self-monitoring during speech production and their proposed neural correlates. Following that, I will delve into recent research endeavors aimed at testing critical predictions derived from these models. Finally, I will elucidate how these studies have generated new constraints and predictions related to self-monitoring during speech production, providing the research questions of the current thesis.

Models of Language Production and Theories of Monitoring Implementation

If one aims to dissect the speech production process into distinct stages, there will be, at the very least, (1) conceptual preparation, (2) lemma retrieval, (3) lemma selection, (4) phonological encoding, (5) phonetic programming, and finally (6) articulation itself (Levelt 1992, 1999; Roelofs 1997; Indefrey and Levelt 2004). Concerning the temporal progression of these stages as well as the amount of interaction between them, there have been three classes of proposals in the literature: (1) sequential modular, (2) interactive cascading and (3) interactive parallel models.

Modular language production models propose a sequential nature of the process, as initially formalized by the works of Lamb and Newell 1966 and Lockwood 1972 as stratificational grammar. This framework posits that language comprises several levels or strata (such as Semantics, Phonology, and Morphology), each with its own structure or syntax. Subsequently, this layered structure was adapted for lexical access models of speech production, involving the retrieval of meaning, word encoding and word production. Modular lexical access models are characterized by a strict unidirectional sequence of modules, where each preceding step must be completed before the subsequent one (Levelt 1992), notably distinguishing lexical processing from encoding (i.e., differentiating meaning from form). Consequently, lemma information, for example, is not influenced by phonological activation.

Further, connectionist Lexical Access models (Dell et al. 1999; Roelofs 1997) embrace the concept of stratification but establish a link between lemma and lexeme, adopting the Spreading-Activation Theory (Dell 1986). This theory posits that higher-level representations directly activate lower-level representations, which in turn reactivate higher levels in cascade, thereby reinforcing overall activation.

Additionally, a third class of models has been proposed to account for the concurrent activation of bottom-up (from concept to encoding) and top-down (from encoding to concept) activations (Strijkers 2016; Fairs et al. 2021). Specifically, the processing of a target word involves a dual-phase mechanism. Initially, the sense of the word, as well as some word knowledge, are activated simultaneously. Their interplay modulates the bottom-up processing by augmenting the sensitivity of automatic target-relevant neural pathways and attenuating the sensitivity of automatic target-irrelevant neural pathways. Following target recognition, a secondary stage of processing, termed 'reverberation', ensues. This phase operates sequentially and serves to integrate the recognized target into the appropriate linguistic and task-specific context, facilitating

the execution of the intended cognitive or behavioral response.

The implementation of monitoring has differed between these proposals. The modular perspective has integrated a monitoring loop based on speech comprehension operating just before articulation for monitoring through the internal channel, and immediately after for monitoring through the external channel. This loop extends to the beginning of the speech planning process, allowing for comprehensive realignment. This implies that a semantic error passes through the lexicalization and encoding stages before being intercepted and corrected. Conversely, an incorrect articulatory command would need to undergo semantic retrieval once again through the monitoring loop. While this proposal can account for virtually any type of speech monitoring, it is not a necessary part of the architecture sustaining modular models of language production.

In contrast, for connectionist models, there is no explicitly designated space for the integration of separate monitoring because the spreading activation itself acts as a safeguard against errors. Put simply, if a segment is incorrect, it probably won't be used due to lack of activation. At the same time, properties inherent to connectionist models such as feedback across levels adeptly account for mixed errors (e.g., both lexical and phonological interference, such as "stupid" for "stubborn," both involving the consonant cluster [st] and both being descriptive adjectives) that are more likely to occur as they receive activation from two levels simultaneously. However, while such models operate effectively without requiring additional processes, they cannot account for the totality of phenomena in language production that require monitoring.

In the context of parallel processing, it is hypothesized that monitoring occurs during a subsequent stage of activation referred to as 'reverberation'. This phase is characterized by the integration of linguistic and contextually relevant information to guide the execution of a desired behavior, such as speech or action. However, a notable gap persists in the current literature, as no precise process has been proposed to elucidate the way monitoring operates during the 'reverberation' stage.

Consequently, these theoretical perspectives propose distinct approaches to the timing and execution of monitoring. Modular models advocate for a dedicated timing mechanism, comprising all potential stages, while the connectionist viewpoint suggests real-time monitoring of stage encoding through mutual reinforcement of activation. Finally, within the realm of parallel processing, it is hypothesised that monitoring takes place during the second stage of activation with no precise account.

In summary, there are broadly three classes of models that cover the entire process

of language production. While two of these models naturally have a way of accounting for certain speech error patterns by integrating spreading activation and feedback, none of the models integrate a monitoring mechanism as a constitutive part. An exception to this is the integrated theory of production and comprehension that was put forward by Pickering and Garrod 2013 and that integrates forward modeling as a means of monitoring upcoming speech at several levels, ranging from conceptual encoding to articulation. However, as it is unclear what type of information processing dynamics this model uses, I will refrain from discussing it here and rather describe it in some more detail in the section dedicated to integration-based monitoring. Similarly, there are models focusing exclusively on the postlexical parts of speech production that integrate well defined mechanisms of control into their architecture through a combination of feedforward and feedback loops (e.g., Guenther et al. 2006; Guenther and Hickok 2016; Hickok 2012; Tourville et al. 2008). While I will leverage the extensive detail found in these models for a portion of the empirical studies, thereby indirectly exploring the applicability of certain mechanisms proposed for post-lexical levels to other levels, the primary focus of this thesis is on theories capable of comprehensively explaining the entire language production process. In what follows, I will briefly discuss several theories have emerged to elucidate the underlying mechanisms of self-monitoring, each rooted in but not necessarily tied to diverse models of language production. Throughout this section I will attempt to establish links to studies that have aimed at uncovering the neural correlates of the proposed monitoring mechanism as this information will be at the center of my empirical research.

Comprehension-Based Monitoring

The comprehension-based monitoring theory stands as one of the most evident and longstanding explanations for the functioning of self-monitoring. This theory posits that individuals utilize the processes involved in speech comprehension to monitor their own inner and audible speech, akin to how they would perceive external speech (Levelt 1983, 1999; Hartsuiker and Kolk 2001; Roelofs 2020).

This model applies seamlessly to both internal and external self-monitoring, differing only in the stage where the monitoring process becomes active. During speech planning, an internal 'phonetic plan' is envisioned to engage speech comprehension processes through an inner loop. In contrast, in the case of audible speech, an external loop interfaces with the speech comprehension system.

When an error is identified through the speech comprehension system, the production halts, and the 'phonetic plan' (just prior to articulation) or the heard speech (following articulation) are sent back to the conceptualizer — the initial stage of the production process — for correction. Subsequently, the speech production process recommences. Consequently, error detection predominantly takes place during later stages of speech production. Remarkably, erroneous units traverse the entire sequential course of speech production, initiating from conceptualization, irrespective of the error's nature. Even in cases of mere mispronunciation, where an incorrect articulatory command is involved, the erroneous unit undergoes stages tied to meaning, such as lemma retrieval.

A meta-analysis (Indefrey and Levelt 2004) involving tasks that manipulated auditory feedback led to the proposal that the bilateral posterior superior temporal gyrus forms the foundation of comprehension-based monitoring. Nevertheless, the validity of the studies forming the basis of this analysis, in terms of indicating monitoring, has been challenged. Altered feedback experiments may not involve genuine speech production errors, revealing processes more related to heightened auditory attention rather than self-monitoring within speech (Meekings and Scott 2021). Furthermore, an updated meta-analysis (Indefrey 2011) displayed involvement of regions beyond the temporal cortex, including the medial frontal cortex and cerebellum. This expansion of regions goes beyond the confines of a strictly comprehension-based framework.

Production-Based Monitoring

The theory of conflict-based monitoring represents a domain-general perspective grounded in the notion that a cognitive system is attuned to the degree of conflict present in any given task. Conflict arises when various alternatives compete, demanding increased effort to resolve the cognitive discord. Consequently, the presence of conflict acts as a direct signal for heightened control, and the level of control is dynamically adjusted accordingly.

In the context of speech, conflict-based monitoring operates when competition emerges at different stages of speech preparation. For instance, if multiple units (e.g., the concepts of BABY and DAUGHTER) are activated with comparable intensity during the conceptualization stage, the overlapping activation signals conflict. This conflict triggers the recognition of the need to alleviate the competing demands. This viewpoint seamlessly accommodates both internal and external monitoring

processes. However, when errors occur, they represent a process breakdown that can incite another layer of conflict tied to expectations. For instance, if an error is detected but remains uncorrected, it gives rise to a conflict first during detection and subsequently when the error is identified in later stages of speech production, such as during articulation or auditory feedback. Hence, multiple sources of conflict can manifest during speech.

Given the domain-general nature of this model, the structures implicated in the process are likewise general. Medial frontal structures, including the pre-supplementary motor area (pre-SMA) and the anterior cingulate cortex (ACC), have been reported to play a role in studies investigating speech errors (Gauvin et al. 2016; Moller et al. 2006). Furthermore, conflict-based monitoring has been effectively tested as part of a language production model in individuals with aphasia (Nozari et al. 2011). Nevertheless, within a framework that entirely omits the concept of feedback correction, certain phenomena like online speech adaptation to altered auditory feedback pose a challenge. Specifically, it's widely acknowledged that when speakers are required to utter a specific target like "bet" but perceive themselves saying "bit" due to modified auditory feedback, they instinctively adjust their speech production based on this altered feedback (Savariaux et al. 1995; Niziolek and Guenther 2013). These instances are closely tied to speech monitoring since speakers occasionally find the need to modify their speech in response to external factors such as interlocutor cues, ambient noise, unexpected environmental changes, and more.

One potential explanation, as proposed by Gauvin and Hartsuiker 2020, involves integrating conflict-based monitoring into both the language production and comprehension streams. This means that a conflict between highly active nodes, whether during the process of response selection or language comprehension, triggers error detection and subsequently initiates correction through a domain-general mechanism. This dual application of conflict-based monitoring provides a plausible route for accounting for instances where speech production adaptation occurs in response to external feedback alterations, as it aligns with the notion of conflict arising from competing demands within the cognitive system.

Integration-Based Monitoring

Another theory of self-monitoring approaches the cognitive process through the lens of internal modeling or predictive coding, known as forward modeling. Originating

from the domain of motor control, this theory has been subsequently extended to encompass diverse cognitive functions, including speech and language. The crux of the theory centers on the existence of predictions concerning the sensory outcomes of actions. These predictions allow the detection of mismatches between the actual and predicted outcomes (Wolpert et al. 1995; Miall and Wolpert 1996). These predictions emerge simultaneously with the formulation of motor programs and are subsequently compared not only against sensory feedback but also against inner perceptions during the action's planning phase, before it materializes overtly.

In the context of speech, predictions could be formulated based on the audition and proprioception of articulatory movements. When the system detects a discrepancy between auditory or proprioceptive predictions and the actual sensory feedback, it triggers a corrective command. Several theoretical models of speech production have integrated forward modeling as a mechanism for self-monitoring (Guenther et al. 2006; Tourville and Guenther 2011; Hickok 2012; Tian and Poeppel 2010).

For higher-level language processes that aren't inherently tied to sensory-motor aspects of speech, one can speculate that similar comparison processes might operate based on more abstract inner representations (Pickering and Garrod 2013). Alternatively, linguistic representations might be distributed and interconnected across all dimensions, enabling sensory-motor attributes to directly influence higher processing levels (Runnqvist et al. 2021; Strijkers 2016; Fairs et al. 2021).

Initially, temporal areas such as the posterior superior temporal gyrus (pSTG) and the superior temporal sulcus (area of Sylvian parietal temporal (SPT)) were proposed to be implicated in these speech-specific mechanisms. For example, in Hickok 2014, the SPT plays a role in comparing auditory and motor targets during the syllabification phase. More recently, the cerebellum's role has garnered support from various studies (Runnqvist et al. 2016; Runnqvist et al. 2021). Prior research had demonstrated that cerebellar activity is influenced by the predictability of self-generated movement outcomes in the motor domain (Blakemore et al. 2001; Imamizu et al. 2000). Studies manipulating auditory feedback have also indicated increased cerebellar activity (Tourville et al. 2008). Furthermore, the involvement of the cerebellum in tasks isolating cognitive processes has prompted the extension of the forward modeling theory to encompass mental activities (Ito 2008). In a linguistic context, this suggests predictions not only for articulatory stages but also for earlier phases like conceptualization and lemma retrieval (Pickering and Garrod 2013).

In summary, the literature presents three distinct types of monitoring from a cognitive standpoint:

(1) **Comprehension-Based Monitoring:** This approach relies on internal and external loops connecting the speech production system with the speech comprehension system. It emphasizes the use of comprehension processes to monitor both inner and audible speech. The strength of the theory lies in its comprehensive explanation of the language production process, encompassing all its stages, and its alignment with behavioral data, including the timing of repairs. Nevertheless, it has a limitation in its inability to accommodate neural activity observed beyond the temporal cortex, as evidenced by numerous monitoring studies (e.g., ACC in Moller et al. 2006; cerebellum in Runnqvist et al. 2021; IFG in Severens et al. 2012).

(2) **Production-Based Monitoring:** This theory operates on the concept of conflict, which quantifies the level of competition occurring at different linguistic levels. Monitoring arises as a response to this conflict, aiding in error detection during speech production. The theory has notable strengths, including its domain-general perspective, alignment with behavioral data, relevance to generalized brain structures, and applicability to pathological conditions. However, it faces limitations when applying the concept of conflict to all linguistic phenomena, particularly in cases involving online speech adaptation, environmental noise, and dynamic environmental changes. These limitations highlight areas where further research and refinement may be needed to enhance the theory's explanatory power and completeness.

(3) **Integration-Based Monitoring:** This perspective stems from the integration between speech production and perception. It enables internal comparisons between predicted percepts and actual motor outcomes, facilitating error detection and correction. The strength of the theory lies in its straightforward application to the motor steps of the language production process, providing a clear and practical framework for understanding how self-monitoring operates in these stages. However, its extension and implementation into higher-level language processing remain somewhat limited and may require further refinement and exploration.

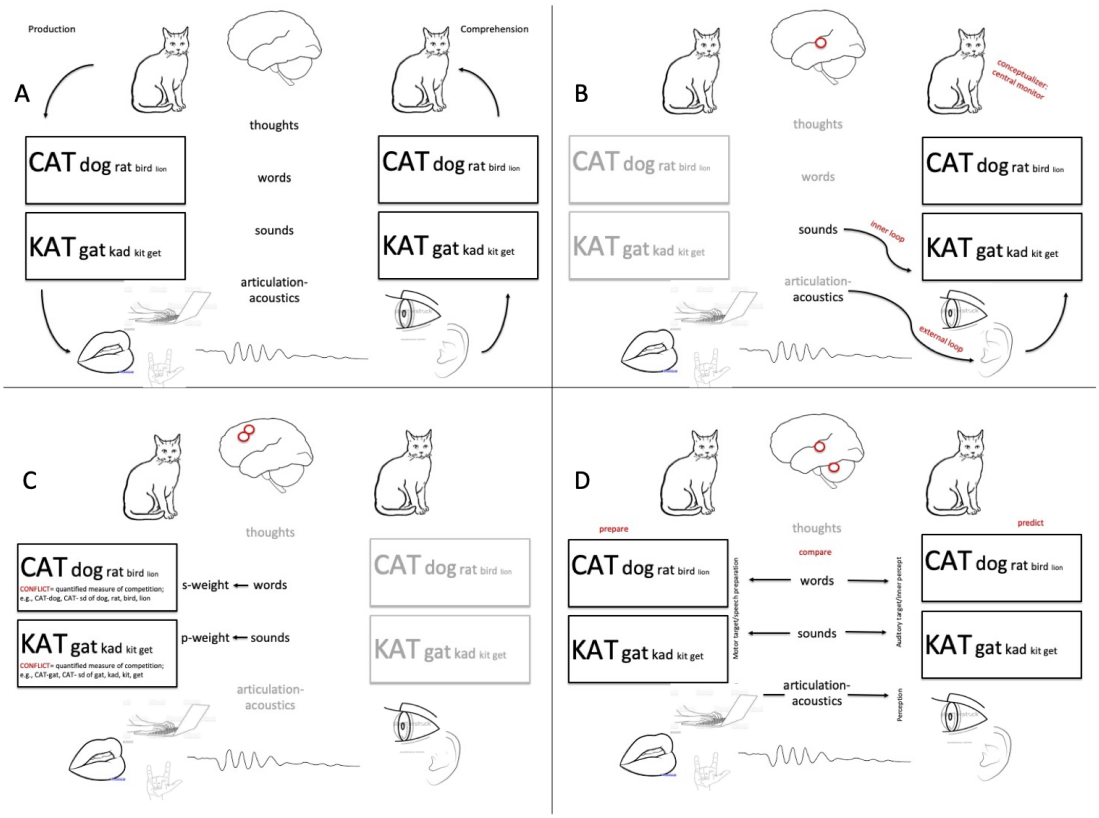


Figure I.1 – Schematic illustration of three types of monitoring models. (Panel A) The basic processes involved in producing and comprehending speech. In the case of production, a concept (cat) triggers activation of lexical (cat, dog, rat bird, lion) and phonological (cat, gat, kad, kit, get) representations. The acoustic signal is the input of speech comprehension that ends with the activation of the corresponding concept. (Panel B) A comprehension-based model of monitoring. (Panel C) A production-based monitoring model in which a domain-general monitoring mechanism operates based on conflict. (Panel D) An integration-based monitoring model: the production system prepares speech guided by predictions or targets that are comprehension-based, and the continuous comparison between both streams provides an error signal upon a mismatch, concretely through a decreased reafference cancellation. From Runnqvist 2023.

Throughout this discourse, it becomes evident that each of these models possesses its own strengths and limitations. However, none of them independently provides a comprehensive explanation for all the phenomena associated with speech errors and their detection. The act of language production is a multifaceted cognitive-motor skill, involving numerous components such as organs, processes, and representations. The complexity of language production makes it a challenge to find a singular model that fully accounts for all aspects of speech errors and their monitoring. For instance, while all three theoretical accounts aim at accounting for monitoring of the entire speech production process, it is possible that they intervene for different dimensions. One dimension that was already discussed is the internal versus external channel

of monitoring. It remains uncertain whether the two monitoring channels are underpinned by identical cognitive processes, implying that the mechanisms diverge only in their engagement timing. Alternatively, internal and external monitoring are distinct processes, each governed by unique underlying mechanisms. Addressing these alternatives was the goal of the first study of the present thesis.

Empirical anchor point of the current thesis: Runnqvist et al. 2021

My involvement in the further discussed study holds profound significance within the context of my thesis. It served as a formative experience in several critical dimensions. Firstly, it introduced me to a new and intricate theoretical domain of speech production. This exposure not only broadened my horizons but also deeply influenced the conceptual underpinnings of my research. My introduction to this new field began with the annotation of speech errors and reaction times, providing me with a practical understanding of speech error patterns and the monitoring process right from the outset. Furthermore, the study in question provided the immediate empirical backdrop for my thesis project. It laid the groundwork for the specific research questions I have pursued, and it has been instrumental in substantiating the hypotheses I aim to test. In essence, my involvement in this study has not only expanded my academic horizons but has also forged an indispensable empirical foundation that anchors and informs the entirety of my thesis.

The investigation of Runnqvist et al. 2021 (Chapter 1 of the present thesis), adopted a comprehensive perspective on theories pertaining to speech monitoring using an event-related fMRI technique.

The study presents three hypotheses about cognitive mechanisms with distinct neural correlates related to speech error monitoring. Concretely, the use of comprehension-based monitoring was indexed through posterior temporal cortex, production-based monitoring through medial frontal cortex, and integration-based monitoring through the cerebellum. Moreover, the probability of committing an error was manipulated (high versus low probability). In that way, monitoring through the inner channel could be indexed by contrasting high and low error probability on correct trials. Finally, the overt errors were also contrasted with correct productions. Given that speakers detect their overt speech errors rather frequently, and that at least a share of these errors are assumed to be detected during articulation, this contrast can be assumed to involve

the external channel to a greater extent. In the fMRI scanner, participants produced word pairs that were primed to elicit a switch of two initial consonants of words in a pair. These errors led to either lexical outcomes (resulting in an existing word pair) or non-lexical outcomes (resulting in a pair of non-words). The assessment of internal monitoring was derived from the comparison between the more error-prone condition (lexical) and the less error-prone condition (non-lexical), specifically focusing on correct trials. Conversely, the evaluation of external monitoring was based on the comparison between overt errors and correct trials. The neural correlates, in this way, functioned as evidence supporting the involvement of mechanisms engaged in internal and external monitoring.

The following key findings emerged from this study:

The temporal cortex emerged as a critical contributor to the process of monitoring of overt speech errors. This empirical evidence substantiates the proposition that the temporal cortex plays a pivotal role in overseeing comprehension-based monitoring, underscoring its involvement in detecting and processing errors during speech production.

The medial frontal cortex (MCF) previously associated with conflict monitoring (e.g. Gauvin et al. 2016; Riès et al. 2011) showed differential brain responses in external and internal speech monitoring. Specifically, only overt errors contrasted with correct trials engaged differentially the ACC and the pre-SMA. Therefore, it suggest that MFC's role may be related to feedback control, rather than conflict as previously thought. This assumption is based on the MFC's role in a broader network for vocal cognitive control, as suggested by Loh et al. 2020. In their work, the authors propose that in primates, the pars opercularis of the inferior frontal gyrus (IFG) governs orofacial and non-speech vocal responses (area 44), while the midcingulate cortex analyzes non-speech vocal feedback. Since the pars triangularis of the IFG (area 45) and pre-SMA are crucial for human speech control, the MCF's involvement aligns with this proposal.

The cerebellum was observed to participate in both internal and external monitoring, with intriguing nuances related to its subregions. Specifically, the posterior cerebellum (Crus I) was associated with both the internal and external monitoring, while the superior medial cerebellum was observed only in condition with overt errors. This distinction highlights the cerebellum's multifaceted role in speech monitoring,

potentially recruiting different cerebellar subregions for distinct functions within this cognitive process.

In summary, the study contributes robust empirical substantiation to the existing hypotheses suggesting the existence of both shared and distinct cognitive mechanisms governing internal and external speech error monitoring. Furthermore, these findings underscore the importance of investigating both facets of speech error monitoring to attain a more holistic comprehension of the intricate cognitive and neural mechanisms underpinning speech production.

Research Questions

While the study of Runnqvist et al. 2021 offers an extensive examination of the neural mechanisms governing speech monitoring and successfully bridges existing theories with these neural mechanisms through the speech-monitoring task, it also opens up several questions. Specifically, the dissociations between the brain structures underpinning the monitoring of lexically driven error probability and overt errors was interpreted in terms of internal and external monitoring, but raises questions about a potential role of timing, levels of representation or a combination of both behind the observed effects. Moreover, the overarching involvement of the cerebellum, a structure typically associated with motor control, suggests that simpler control circuits initially applied to movement have been recycled for the purpose of speech monitoring following a hierarchical buildup. Finally, the involvement of anterior cingulate cortex in contexts restricted to overt error production suggests alternative interpretations of the ACC involvement in terms of vocal feedback control. These inquiries, described in more detail below, are the ones that I aspire to address, at least to some extent, within the scope of my thesis.

Monitoring as a Multifaceted Process: Temporal Dynamics and Levels of Representation

Time Course of Monitoring

Although the differential brain activity observed in Runnqvist et al. 2021 does provide evidence for distinctions in monitoring mechanisms based on the probability of errors and linguistic variables, it does not provide insights into the precise timing of their engagement during speech production. There is a need for a more in-depth

exploration of the timing and sequence of these monitoring stages. As previously deliberated within the **Theories of monitoring implementation** section, it is evident that there exists no universally embraced framework or consensus concerning the entirety of the monitoring implementation process. Nonetheless, it is crucial to recognize that empirical substantiation has surfaced, bolstering each of the proposed mechanisms under distinct circumstances.

One notably influential study in the realm of monitoring implementation is the work by Nooteboom and Quené 2017, which has substantial theoretical implications. This research introduces the concept of a bimodal distribution in cut-off times following the occurrence of an error. According to their findings, there are two distinct channels through which monitoring operates. Specifically, the study reveals that production strings were terminated either approximately 186 milliseconds after the error or around 660 milliseconds following it. The existence of this bimodal distribution strongly suggests the presence of two monitoring channels: one operating before the actual production begins and another after the speech becomes audible to the speaker and possibly to the listener as well.

A potential avenue for addressing further questions on timing of monitoring can be found in the findings of multiple neuroimaging studies that have examined the Error-Related Negativity (ERN) event-related potential (ERP) component. The ERN is a negative-going ERP component that typically emerges at approximately 100 milliseconds following the occurrence of an error. The presence of the ERN at this specific temporal juncture suggests a rapid neural response to the detection of errors in various cognitive tasks. Its timing aligns closely with one of the observed cut-off points in monitoring processes, around 186 milliseconds after an error, as discussed earlier. The ERN component has been observed in several studies, illustrating its presence and relevance across various experimental contexts. For instance, Ganushchak and Schiller 2006 detected this component during a phoneme monitoring task with motor response modality, while in Ganushchak and Schiller 2008b it was observed in a vocal response modality task. In both cases, the ERN exhibited a more pronounced negative amplitude in error trials compared to correct ones. Similarly, Riès et al. 2011 reported the presence of this component in an overt picture naming task, again with a greater negative amplitude on error trials, but also later for errors than for correct trials in task with motor response modality. These findings collectively link the ERN to the process of overt error detection, showcasing its involvement in monitoring during language-related tasks.

Furthermore, experimental findings pertaining to the timing of monitoring processes indicate a secondary process occurring approximately 200 milliseconds after the response. This is evident through the emergence of a subsequent negative component following errors. In studies conducted by Riès et al. 2011, this negativity was identified in a picture naming task with vocal response modality, typically appearing around the 200-millisecond mark. Intriguingly, this component was absent in task that did not involve vocal responses (grammatical gender decision task in Riès et al. 2011), and as demonstrated in research by Masaki et al. 2001 was still present when feedback was obscured by pink noise. Consequently, this component has been associated with error correction phenomena, as a similar component was observed in non-linguistic tasks, such as the Flanker task, specifically in cases where errors were rectified (Fiehler et al. 2005).

Several other investigations have also delved into the time leading up to the outward manifestation of an error. This exploration aims to uncover the processes occurring before an error becomes overt. One notable discovery in this regard is the identification of the ERN mentioned earlier. It's noteworthy that the ERN component that was observed at around 100 ms after the error commission was identified in a distinct time window in the study conducted by Moller et al. 2006. This study employed the SLIP task, involving word pair naming after a speech cue appeared on the screen. Importantly, Moller et al. 2006 reported an increased amplitude of a negative component within the 50 to 150 ms time window following the speech cue for error trials when compared to correct ones. Their research linked this component to processes occurring prior to overt errors and associated it with the conflict arising from competing motor channels before vocalization. Furthermore, within the same study, Moller et al. 2006 identified a second negativity emerging in the time window of 230 ms to 300 ms after the vocalization prompt, attributing it to the same conflict-related function. Moreover, the research conducted by Severens et al. 2011 also examined the effects associated with the cue that instructs participants to speak. In their study, which involved word pairs leading to inappropriate language slips, they observed an augmented negativity in the time window spanning from 550 to 600 ms following the speech cue. It's noteworthy that this negativity emerged in trials that did not ultimately result in errors but were predisposed to elicit inappropriate (taboo) word pairs. As such, this study marks a significant departure by scrutinizing a process not directly tied to the occurrence of an error but rather one that aids in error avoidance. Consequently, the components observed in their study are posited to be linked to

conflict, with a focus more on resolving conflict than detecting it.

In the investigation by Moller et al. 2006, electrophysiological activity linked to the presentation of the target word-pair also diverged between errors and correct trials. Specifically, trials that later resulted in erroneous productions displayed an augmented negativity within the time window spanning from 350 to 600 ms following the presentation of the target pair. This negativity emanated from two distinct sources: the supplementary motor area (SMA) and the middle temporal gyrus (MTG). While the authors associated this activity with conflict, considering the temporal structures and timing involved, this conflict appeared to be linked to phonetic encoding. This was in contrast to the later negative components that Moller et al. 2006 associated with articulatory conflict.

Similarly, another study by Ganushchak and Schiller 2008a documented a negative deflection peaking around 450 ms after the presentation of the target. Furthermore, this negativity exhibited an increase in trials presented with a semantic distractor, suggesting heightened conflict according to the authors.

To consolidate the current body of research, it is evident that multiple critical time points warrant examination to enhance our understanding of the organization of the monitoring process. Previous studies have focused on the moments when a speech target appears, a speech cue is given, and vocalization occurs, all contributing insights into monitoring at various stages. Nevertheless, the existing knowledge base exhibits considerable heterogeneity. While several studies have explored specific time points, there is currently no study that comprehensively addresses all three pivotal events within a single experimental protocol. Consequently, there is a compelling need for alignment and integration of these disparate findings. The goal of the Chapter 2 in this thesis is to achieve this objective. Specifically, this chapter seeks to create a thorough and reliable chronological overview of monitoring in language production by analyzing the entire monitoring process across three time points that have been explored in prior research: the appearance of a to-be-named stimulus, the appearance of a speech cue, and the vocal response. This study also investigates both internal and external monitoring pathways.

Modulations of Monitoring by Level of Representation

The second line of questioning delves into the adaptability and flexibility of the entire monitoring process concerning different levels of representation.

Returning to the well-known examples of slips of the tongue from the first paragraph

of this document, the substitution of "stubborn" with "stupid" could potentially be attributed to the initial consonant cluster [st], presumably constituting a phonological encoding error. This type of error affects the word's form. On the other hand, the replacement of "Iraq" with "Ukraine" appears to be less reliant on the phonology of the words involved, but rather more likely to be semantically connected through the concept of war, rendering the entire sentence fitting in both cases. Similarly, a comparable semantic relationship could be inferred in the third example of a dysfluent [d] preceding "baby," which the media interpreted as standing for "daughter." These latter two examples differ from the first, as they indicate that errors arise from meaning rather than form. Consequently, such slips may occur at varying stages of the speech production process, and the self-monitoring process, as an entity, is capable of identifying and rectifying them. This raises the question of whether diverse errors are subject to monitoring and correction via the same mechanism due to the shared language context, or if monitoring is contingent upon the nature of the represented information. For instance, articulation pertains to the motor aspect of speech, whereas lemma retrieval is more cognitive in nature as it doesn't result in immediate motor actions. Thus, it prompts the inquiry of whether a single process governs the monitoring despite these distinct stages or if the process is influenced by the specific representation being monitored. As previously discussed in **Theories of monitoring implementation** section, numerous models aim to account for the monitoring process in speech, with a general assumption being that they employ the same mechanism across all stages of speech processing. However, none of these models independently offers a comprehensive explanation for all phenomena related to speech errors and their detection. Consequently, there are compelling reasons to consider the interaction of mechanisms during monitoring.

From a theoretical standpoint, most language production models posit that different language levels exhibit varying degrees of direct connection to sensory-motor aspects (e.g. Dell et al. 1999; Levelt 1989). This leads to the idea that levels closely tied to sensory-motor aspects might employ monitoring mechanisms reliant on feedback from these properties, while higher levels may depend on information processing factors such as conflict. Nevertheless, the reviewed models tend to assume the existence of an inner perceptual level or that the same information processing properties are consistently utilized throughout the language production process for monitoring purposes, seemingly rendering perceptual variables irrelevant even in the monitoring of overt errors. Furthermore, certain studies have revealed discrepancies in error

patterns between inner speech and articulated speech. For example, Oppenheim and Dell 2008 demonstrated that in imagined speech, lexical errors occurred more frequently than non-lexical errors, mirroring the pattern observed in overt speech. However, only overt speech exhibited sensitivity to phonetic variables, such as an increase in exchanges of consonant onsets when onsets shared phonetic features. These findings suggest that monitoring at different representational levels may occur at distinct time points and could potentially rely on different mechanisms.

In line with this, it has been observed that error detection during speech planning initiates repairs swiftly, but this differs when errors are externally detected (e.g., Nooteboom and Quené 2017). One interpretation of this discrepancy is that internal repairs rely on the existing activation of the correct candidate, which was initially not selected, whereas repairs for externally detected errors cannot rely on such remaining activation and must be prepared from scratch. Additionally, various neuroimaging studies have provided evidence of distinct neural correlates for internal and external monitoring that could also relate to different levels of representation. For instance, Okada et al. 2018 manipulated speech production, distinguishing between conditions involving only lexical processing (imagined speech) and those engaging motor-phonological processing (speech articulation without phonation). Importantly, none of the conditions involved participants receiving auditory speech feedback. The authors noted increased pSTG activation in the condition requiring motor-phonological processing compared to imagined speech. This was taken to indicate forward modeling, suggesting that overt articulation (the execution of motor speech plans) generates stronger forward predictions than imagined speech.

It is evident that a potential confounding factor arises when trying to differentiate between internal and external monitoring, particularly when assessing cognitive and motor aspects of speech. Previous research has evaluated the examination of internal monitoring during cognitive stages of speech processing, such as pragmatics or lexical processing (e.g., Severens et al. 2012; Runnqvist et al. 2021). Conversely, the assessment of external monitoring has primarily involved the aggregation of multiple observable speech errors, which may contain an indeterminate mix of motor-related issues (Gauvin et al. 2016; Runnqvist et al. 2021). One possibility to differentiate the level of representation is either to manipulate error probability on distinct representational levels and look at the differences in processing on correct trials (e.g., Dorokhova et al. 2023 or Chapter 2) or classify errors based on their functional sources (e.g., Ganushchak and Schiller 2008b).

Further, if comprehension - based, production-based and integration-based monitoring coexist, it is reasonable to surmise that they are specialized for specific representations within the whole complex speech production and monitoring process. Moreover, findings of Todorović et al. 2023 present a substantial overlap of internal and external monitoring that differs from Runnqvist et al. 2021's findings. Importantly, the internal monitoring in Todorović et al. 2023 was manipulated on the articulatory-phonetic level, in contrast to Runnqvist et al. 2021, which concentrated on internal monitoring at the lexical level. The observation of a substantial overlap in brain structures between internal monitoring at the articulatory-phonetic level and external monitoring suggests that the activation of specific brain regions depends on the position of the represented information along the motor-cognitive continuum. Furthermore, the internal and external monitoring differentiation signaled the use of feedback to sustain external monitoring. In light of the aforementioned findings, it becomes evident that a relationship exists between the level of representation and the type of monitoring. Specifically, levels of representation that are closely associated with sensorimotor aspects of speech, such as the articulatory-phonetic level, appear to leverage mechanisms rooted in feedback from these sensorimotor aspects.

In conclusion, it could be hypothesised that different monitoring mechanisms could complement each other, serving the monitoring needs of various types of representations. These representations may align with different levels of language processing. For instance, lexical and grammatical processes might engage separate mechanisms in contrast to post-lexical processes, which involve generating proprioceptive feedback.

To investigate the influence of representation levels, Chapter 2 examined the error probability effect at two distinct levels: lexical and articulatory-phonetic. Additionally, an ongoing fMRI study involving different levels of linguistic representation: semantic and articulatory-phonetic, within a SLIP task was described in Supplementary Chapter A1 in Additional Content of the present thesis. The observed neural correlations across these linguistic representation conditions provide insights into their interplay.

Linking Speech Monitoring with General Action Control through Precursors of Language

Recycling of Mechanisms for the Purpose of Speech Monitoring

Lastly, the involvement of brain structures that were linked to domain general (conflict) or motor-specific (internal modeling) mechanisms in a speech-related task, as it

was put forward by Runnqvist et al. 2021, raises an intriguing question. Is it conceivable that complex cognitive processes, like speech production, recycle mechanisms that originally evolved to support basic motor reflexes?

Previously, motor control prediction models were put forth as a means to explain higher-level cognitive functions. This proposition suggested that the concept of forward modeling could serve as a comprehensive framework for tasks such as action observation and comprehension, mental rehearsal, imitation, and social cognition, as outlined in Wolpert and Flanagan 2001. Furthermore, Pezzulo et al. 2022 suggests that the evolution of cognitive capabilities stems from the internalization of control, essentially transforming the control of external states in the real world into self-regulation of behavior. In other words, suggesting an evolutionary link from simple motor reflexes to high-level cognitive processing (also Friston 2011). Additionally, Ito 2008 introduces an adaptation of forward and inverse models for both voluntary movements and mental activities, proposing that this same process operates universally to fulfill motor and cognitive needs through an adaptive network in the brain. Looking at it from another angle, an expanding body of research has delved into the connection between prediction and cognition. These studies have revealed the existence of fronto-temporo-parietal or cortico-cerebellar brain networks that exhibit sensitivity to variations in predictability within cognitive tasks, even in cases where these tasks have minimal or no motor-related elements (e.g., Caucheteux et al. 2023; Runnqvist et al. 2016; Runnqvist et al. 2021).

Given the prior cited works, it is worth considering the concept of self-monitoring as a potential overarching cognitive ability, extending beyond the confines of language. This proposition opens up intriguing avenues for exploration. One compelling avenue involves investigating whether self-monitoring functions as a shared cognitive process across a spectrum of tasks. For instance, does the way we monitor our actions while driving align with how we oversee our speech during a conversation? Similarly, can the monitoring of our movements during sports be likened to the monitoring of our articulation while reciting a poem? A comparison between these two scenarios provides insight into this concept. In the context of both driving and conversation, a continuous flow of multiple cues demands rapid and precise reactions, necessitating seamless coordination and quick adjustments. Conversely, the second example, which involves the recitation of a poem, represents a learned performance where execution takes precedence. These elementary examples serve to illustrate that speech can be perceived as a commonplace cognitive task akin to various other activities in our daily

lives.

This idea finds resonance in the fact that two out of the three discussed **Theories of Monitoring Implementation** (forward and conflict-based theories) are not limited to language processing alone. In line with this view, several speech monitoring models capitalized on the non-specificity of language processing over other functions by implementing domain−general theories into current speech models. The conflict-based monitoring theory was successfully implemented into model of speech production (Nozari et al. 2011) claiming the domain-general process of conflict detection to be used for speech monitoring. As well as the internal modeling theory in computational model by Pickering and Garrod 2013 which modelizes both language production and comprehension.

Furthermore, empirical evidence from neuroimaging studies lends support to the notion that non-linguistic brain networks can be repurposed for the monitoring process. This concept of repurposing is exemplified by the presence of components such as the ERN, which has been observed in non-linguistic contexts as well (e.g., Fiehler et al. 2005). The ERN is known to be elicited in the ACC, which is the primary locus of conflict monitoring (e.g., Botvinick et al. 1999), but also of feedback processing (Loh et al. 2020) as well as the encompassing entity of several motor areas (e.g., Loh et al. 2018) Furthermore, task-related activation for feedback and motor processing within this region has been found to depend on its morphological variability.

In light of these findings, we aimed to explore the domain generality of the ACC region by investigating whether this morphological variability influences linguistic tasks, similar to its observed impact on motor activation (Loh et al. 2018) and feedback processing (Loh et al. 2020). To accomplish this, Chapter 4 involved analyzing individual anatomical scans to categorize them morphologically and then examining the relationship between these morphological classifications and functional activations, specifically focusing on the contrast between correct responses and overt errors in the SLIP task.

Association of motor and cognitive processing as a hierarchical buildup

Moreover, empirical findings support a continuity in the internal modeling used in both the motor and cognitive domains under the form of hierarchical buildup. For example, patterns of hierarchical motor and executive control have been observed in the frontal and parietal cortices, as demonstrated by researchers such as Choi et al. 2018 and Grafton and Hamilton 2007. These patterns also extend to the cerebellum,

as evidenced by the work of D'Mello et al. 2020.

Relatedly, Grafton and Hamilton 2007 established a motor hierarchy that spans from the lowest to the highest levels of complexity. They applied this hierarchy to the realm of action observation, specifically focusing on grasping movements. Their study identified a distributed set of brain regions that exhibited varying levels of activation based on the behavioral motor complexity and its relation to the final outcome of a movement.

Our next hypothesis aims to relate to the varying levels of linguistic and non-linguistic representations, suggesting that monitoring operates in a hierarchical manner, spanning from primary to transmodal processing. This could explain the differing brain regions activated in monitoring-related tasks across the temporo-parietal cortex, medial frontal cortex, and cerebellum. This hierarchical gradient may encompass sensory-motor representations to more abstract and associative ones, aligning with proposals made for the temporal lobe by Mesulam 1998. For example, within the temporal cortex, this gradient might range from basic auditory functions (primary auditory area) to higher-level functions such as speech comprehension (Wernicke's area) and lexical and semantic processing (mid-temporal gyrus). Similarly, within the frontal lobe, it could extend from basic motor functions (primary motor area) to more complex motor behaviors like cognitive control and speech production (Broca's region, Badre and D'Esposito 2009). Recent descriptions of gradients in the cerebellum, moving from basic motor processing to transmodal processing, further support this hypothesis (Guell et al. 2018; D'Mello et al. 2020).

In summary, prior studies have provided valuable insights into the realms of both motor and mental actions, particularly with regard to internal modeling. These studies have also compellingly argued that the brain's organization follows functional hierarchical gradients, as exemplified by the research of Huntenburg et al. 2018 and Guell et al. 2018, and that similar hierarchical patterns are present in the domain of action control.

In this context, our primary objective was to establish a link between internal modeling and hierarchical processing theories. Our approach to achieving this goal involved investigating whether alterations in cognitive demands would lead to observable shifts in the hierarchical arrangement of frontal, parietal, and cerebellar regions during self-generated actions. To carry out this investigation, we conducted a study in Chapter 3. In this study, we designed a motor task that required the participation of various body parts, and we utilized fMRI scanning to examine the corresponding

brain activity. Participants were given specific instructions, either to perform a simple motor movement or to engage in a movement that included an additional monitoring rule, thereby introducing a greater level of complexity to the task. Our analysis focused on discerning the differences in brain regions activated during isolated motor movements and those activated during monitored movements. We also explored how these patterns of brain activation corresponded to existing empirical evidence that supports the hierarchical organization of the cortex and cerebellum.

1 Cerebellar and cortical correlates of internal and external speech error monitoring. An fMRI study.

Runnqvist Elin
Chanoine Valérie
Strijkers Kristof
Pattamadilok Chotiga
Bonnard Mireille
Nazarian Bruno
Sein Julien
Anton Jean-Luc
Dorokhova Lydia
Belin Pascal
Alario François - Xavier

Outline

Cerebral Cortex Communications, 2021, **2**, 1–12

doi: 10.1093/texcom/tgab038
Original Article

ORIGINAL ARTICLE

Cerebellar and Cortical Correlates of Internal and External Speech Error Monitoring

Elin Runnqvist[1], Valérie Chanoine[1,2], Kristof Strijkers[1], Chotiga Pattamadilok[1], Mireille Bonnard[3], Bruno Nazarian[4,5], Julien Sein[4,5], Jean-Luc Anton[4,5], Lydia Dorokhova[1], Pascal Belin[5] and F.- Xavier Alario[6]

[1]Aix-Marseille Université, CNRS, LPL, Aix-en-Provence 13100, France, [2]Institute of Language, Communication and the Brain, Aix-en-Provence 13100, France, [3]Aix-Marseille Université, Inserm, INS 13005, Marseille, France, [4]Centre IRM, Marseille 13005, France, [5]Aix-Marseille Université, CNRS, INT 13005, Marseille, France and [6]Aix-Marseille Université, CNRS, LPC 13331, Marseille, France

Address correspondence to Elin Runnqvist. Email: elin.runnqvist@univ-amu.fr

F.-Xavier Alario and Pascal Belin share senior authorship of the study

Abstract

An event-related functional magnetic resonance imaging study examined how speakers inspect their own speech for errors. Concretely, we sought to assess 1) the role of the temporal cortex in monitoring speech errors, linked with comprehension-based monitoring; 2) the involvement of the cerebellum in internal and external monitoring, linked with forward modeling; and 3) the role of the medial frontal cortex for internal monitoring, linked with conflict-based monitoring. In a word production task priming speech errors, we observed enhanced involvement of the right posterior cerebellum for trials that were correct, but on which participants were more likely to make a word as compared with a nonword error (contrast of internal monitoring). Furthermore, comparing errors to correct utterances (contrast of external monitoring), we observed increased activation of the same cerebellar region, of the superior medial cerebellum, and of regions in temporal and medial frontal cortex. The presence of the cerebellum for both internal and external monitoring indicates the use of forward modeling across the planning and articulation of speech. Dissociations across internal and external monitoring in temporal and medial frontal cortex indicate that monitoring of overt errors is more reliant on vocal feedback control.

Key words: cerebellum, error monitoring, fMRI, forward modeling, speech production

Introduction

Several phenomena indicate that speakers inspect their utterances for errors. The most obvious evidence for this is that speakers can interrupt and correct themselves (self-repairs, Levelt 1983) or accurately report having committed an error (Postma and Noordanus 1996). Errors are sometimes interrupted or repaired almost immediately after they start to be pronounced, at a velocity indicating that error detection and repair had already been prepared internally, before the error was even audible (Levelt 1983; Hartsuiker and Kolk 2001). Moreover, certain types of errors, such as taboo or nonwords, occur below chance when they would be considered as inappropriate utterances (Baars et al. 1975; Nooteboom and Quené 2008). This indicates that the monitor can filter out impending errors before articulation, thus lending further support to the notion that monitoring may also take place internally. Despite the

Received: 30 April 2021; **Revised:** 26 May 2021; **Accepted:** 26 May 2021

consensus regarding the existence of both inner and external error monitoring processes, their cognitive and neural basis remains contentious (see Lind and Hartsuiker 2020; Gauvin and Hartsuiker 2020; and Nozari 2020 for reviews). Here we aimed at better characterizing the presence of 3 different monitoring mechanisms invoked to account for both inner and external monitoring, namely 1) "comprehension-based monitoring" with neural correlates in temporal cortex, 2) "forward modeling" with neural correlates in the cerebellum, and 3) "conflict-based monitoring" with neural correlates in medial frontal cortex. To this end, we used event-related functional magnetic resonance imaging (fMRI) during an overt production task eliciting speech errors.

Temporal Cortex and Comprehension-Based Monitoring

An influential view has been that speakers rely on speech comprehension processes to detect errors (Levelt 1983; Levelt et al. 1999; Hartsuiker and Kolk 2001; Roelofs 2020). A speaker's own phonologically encoded internal representations and audible speech utterances would be the input of an inner and external channel, respectively, feeding into the very processing loops used when perceiving speech produced by others. This cognitive account fitted nicely with the neurobiological proposal linking monitoring processes to activity in regions of the auditory cortex (Indefrey and Levelt 2004), which was based on the observation of enhanced bilateral activation of posterior superior temporal gyrus (pSTG) in conditions requiring increased speech monitoring (e.g., manipulated auditory feedback, Hirano et al. 1997; auditory hallucinations, Shergill et al. 2000). Other models implement the reliance on speech perception for error detection as a feedback circuit comparing auditory perception with an internal auditory target, and the proposed locus of this comparison is also pSTG (e.g., Golfinopoulos et al. 2010) or the neighboring region sylvian fissure at the parietotemporal boundary (SPT) (e.g., Hickok 2012). However, a recent review and meta-analysis of 17 studies argued to support the implication of the pSTG in monitoring concluded that existing neuroimaging evidence is insufficient to make such an argument (e.g., Meekings and Scott 2021). In particular, there was a mismatch between the pSTG regions proposed as responsible for error detection in the previous literature and the regions identified in an activation likelihood estimate analysis. Also, the studies themselves were found to be methodologically and theoretically inconsistent with one another. In addition, none of the studies on which the models were built was actually based on natural speech errors, but rather on feedback alterations. Hence, it remains an open question whether the pSTG has a role in the monitoring of true speech errors.

Cerebellum and Forward Modeling

The involvement of the cerebellum has been reported in studies involving manipulations of participants' auditory feedback to their own speech (e.g., distorted or noisy feedback, Christoffels et al. 2007; Tourville et al. 2008), verbal fluency (e.g., produce as many words as possible beginning with "s," Leggio et al. 2000), and error priming during speech production (e.g., "tax coal" priming the target "cap toast" into the error "tap coast," Runnqvist et al. 2016). To understand this cerebellar involvement for speech production, one can turn to what is known about the monitoring of nonverbal actions. The cerebellum has been ascribed a crucial role in the monitoring of motor actions through the theoretical construct of forward modeling (also labeled "internal modeling" or "predictive coding"). In a forward modeling framework, the correction of motor commands is ensured by producing expectations of the commands' sensory consequences before their output is effective as physical actions (i.e., through corollary discharges or efference copies; McCloskey 1981; Jeannerod 1988; Wolpert et al. 1995). Cerebellar activity, particularly in the posterior lobules, is modulated by the predictability of the consequences of self-generated movements (Imamizu et al. 2000; Blakemore et al. 2001). Hence, the cerebellum has been proposed as an important center of this forward modeling of motor actions (Imamizu et al. 2000; Blakemore et al. 2001; Miall and King 2008).

The hypothesis of cerebellar forward modeling has also been incorporated into theories and empirical investigations of mental activities, including language processing (Ito 2008; Strick et al. 2009; Desmond and Fiez 1998; Pickering and Garrod 2013; Hickok 2012; Lesage et al. 2017; Argyropoulos 2016). For example, Ito (2008) proposed to extend the domain of forward models from sensorimotor actions to mental activities based on a review of anatomical (i.e., appropriate neural wiring between the cerebellum and the cerebral cortex), functional (appropriate mental activity in the cerebellum), and neuropsychological data (the association of some mental disorders with cerebellar dysfunction). In line with this proposal, it has been shown that a gradient within the posterolateral cerebellum supports cognitive control of both concrete, proximal actions (motor-adjacent subregions) and abstract future processing (motor-distal subregions, e.g., D'Mello et al. 2020). Several theoretical models of the motor control of speech incorporate some form of forward modeling (i.e., Guenther et al. 2006; Tourville and Guenther 2011; Hickok 2012, 2014; Tian and Poeppel 2010). For example, Golfinopoulos et al. (2010) propose that auditory feedback control would be complemented by a cerebellar module (superior lateral cerebellum) and a feedforward control subsystem mediated by a transcerebellar pathway (anterior paravermal parts of the cerebellum). Hickok (2012) proposes that the cerebellum is in charge of the comparison (coordinate transform) between auditory and motor targets at the phonetic encoding stages of speech production. The integration of the cerebellum in these models is based on evidence from feedback manipulations as discussed previously (e.g., Ghosh et al. 2008) and on the role of the cerebellum in ataxic dysarthria studies (e.g., Ackermann et al. 1992). A less explored hypothesis states that linguistic levels of processing that are beyond speech motor control are also monitored through forward models (Pickering and Garrod 2013). Furthermore, this psycholinguistic proposal has not been neurobiologically specified. However, given the increasing evidence of a role of the cerebellum in cognitive processing, an extension of the mechanisms operating on speech motor aspects to language processing proper is conceivable. One study has reported an increase in the production of phonological substitution errors after repetitive transcranial magnetic stimulation to the right posterolateral cerebellar Crus I (e.g., Runnqvist et al. 2016). Hence, this study suggests a direct involvement of the posterior cerebellum in speech monitoring beyond articulatory aspects. However, among others, open questions that remain are whether this type of monitoring is applied during planning or articulation and whether the same or different parts of the cerebellum would be involved for monitoring inner versus overt speech.

Medial Frontal Cortex and Conflict-Based Monitoring

The involvement of several areas in the medial frontal cortex such as the presupplementary motor area (pre-SMA) and the

anterior cingulate cortex (ACC) has been reported in studies investigating error related processing in language production (Gauvin et al. 2016; De Zubicaray et al. 2001; Möller et al. 2007). These areas are the same ones that have been linked to error detection and conflict monitoring in domains other than language, such as in cognitive control (Botvinick et al. 2001; Nachev et al. 2005). The conflict monitoring theory holds that medial frontal structures constantly evaluate current levels of conflict and that, when a conflict threshold is passed, they relay this information on to other regions in frontal cortex responsible for control, triggering them to adjust the strength of their influence on processing. A need for greater control is thus indicated by the occurrence of conflict itself. Such theory can account both for inner and external monitoring through a single mechanism operating on a continuum of conflict on which overt errors would be the most extreme case.

The idea of conflict monitoring as a means of preventing and detecting errors has been incorporated into a model of language production (Nozari et al. 2011) that successfully simulated error detection performance in aphasic patients. Moreover, a few studies have obtained evidence for an involvement of the ACC and pre-SMA also on correctly named trials in tasks involving the presence of explicit conflict in the stimulus to be processed for language production (e.g., semantic interference inflicted by the categorical relationship between a picture to be named and a (near-) simultaneously presented distractor; De Zubicaray et al. 2001; Abel et al. 2012). However, the available evidence only bears on the involvement of medial frontal cortex in the processing of overt errors or of conflict of the type requiring the exclusion of a competing response that is directly present in the stimulus. Hence, in the context of a task without explicit conflict in the stimulus, it remains an open question whether the medial frontal cortex has a role for monitoring in the absence of overt errors.

The Current Study

In short, 3 hypotheses about cognitive mechanisms with distinct neural correlates can be distilled from the literature related to internal and external speech error monitoring, namely comprehension-based monitoring through posterior temporal cortex, forward modeling through the cerebellum, and conflict-based monitoring through medial frontal cortex. As evidenced by our review of the literature, many questions regarding the circumstances in which these mechanisms may be at play remain open. Here we sought to fill some of these gaps by providing independent empirical support for 1) a role of the temporal cortex in the monitoring of true speech errors; 2) an involvement of the cerebellum in inner and/or external monitoring, possibly recruiting different parts of the cerebellum for different functions (posterior for speech planning and superior medial for articulation); and 3) a role of the medial frontal cortex for inner monitoring (in the absence of overt errors). We addressed these hypotheses through an event-related fMRI study designed to examine both internal and external speech error monitoring, with a zoom on temporal, cerebellar, and medial frontal regions linked to the different monitoring mechanisms discussed above.

Eleven regions of interest (ROI) were selected within these 3 broad anatomical regions (Table 1), corresponding to Montreal Neurological Institute (MNI) coordinates reported in theoretically relevant meta-analyses, models, or studies eliciting natural speech errors. In particular, our ROIs in temporal cortex correspond to pSTG regions proposed to underlie the auditory target in the DIVA model (e.g., Golfinopoulos et al. 2010) and to the region SPT that corresponds to the coordinate transform between auditory and motor targets in the HSFC model (e.g., Okada and Hickok 2006). For the cerebellum we selected 2 right posterior coordinates linked to (cognitive aspects of) language processing in the meta-analysis of Stoodley and Schmahmann (2009) as well as the coordinates corresponding to the superior medial cerebellum linked to the articulatory aspects of speech in the DIVA model (e.g., Golfinopouolous et al. 2010). Finally, for medial frontal cortex we selected the coordinates reported for ACC and pre-SMA in Gauvin et al. (2016), being the only previous study that directly contrasted overt natural speech errors and correct trials. For estimates on the right ACC and pre-SMA we used the coordinates reported by the meta-analysis of Hester et al. (2004) stemming from nonlinguistic error-related processing.

Twenty-four healthy volunteers, native speakers of French, performed an error eliciting production task while undergoing blood-oxygen-level-dependent (BOLD) imaging. Based on evidence that a majority of overt errors involve error detection and hence monitoring (Gauvin et al. 2016), external monitoring was indexed by contrasting correct trials and trials with errors. Extending previous work, internal monitoring was indexed on correct trials by manipulating the likelihood of committing an error and hence the load on speech monitoring mechanisms in 2 conditions. This was achieved by priming spoonerisms that for half of the trials would result in lexical errors (e.g., "tap coast" for the target "cap toast") and the other half in nonlexical errors (e.g., "*sost *pon" for the target "post son," Fig. 1). Speakers are more error-prone when lexical rather than nonlexical errors are primed (Nooteboom and Quené 2008; Oppenheim and Dell 2008). This effect seems to be caused by a combination of context biases (inappropriate production candidates are more easily discarded, e.g., Hartsuiker et al. 2005) and of the interactive activation dynamics inherent to speech preparation (the lexical competitor would count on both a phonological and lexical source of activation compared with the nonlexical one, e.g., Dell 1986). Regardless of the cause of the effect, the rationale here is that to-be-articulated words with higher error probability should reveal an enhanced involvement of the inner monitor (Severens et al. 2012). Hence, lexical versus nonlexical error priming was contrasted to index internal monitoring.

Methods

Participants

The study received appropriate ethical approval (filed under id "EudraCT: 2015-A00845-344" at the regional ethical committee "Comité de Protection des Personnes Sud Méditerranée I"). Twenty-eight (18 females, 10 males) right-handed native speakers of French participated in exchange for monetary compensation. Four participants (4 males) were excluded from the analyses: 3 because of excessive head movements during the acquisition and 1 because of a misunderstanding of the task. The average age of the remaining 24 participants was 23.8 (SD 3.2). No participant reported any history of language or neurological disorders.

Materials

Target stimuli were 320 printed French nouns (those used in Runnqvist et al. 2016) to be presented in pairs. For illustrative purposes, the examples in the text are given in English. To control for differences due to physical variance of stimuli, the same words were used across participants and conditions (albeit

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

			Comprehension-based Inner monitoring (inner speech) and external monitoring (audible speech)	Forward modeling Inner monitoring (linguistic dimensions) and/or external monitoring (motor dimensions)	Conflict-based Inner monitoring (impending errors) and/or external monitoring (overt errors)
Medial frontal cortex	ACC	roi1_ACC_L roi2_ACC_R			(−6, 20, 34) Gauvin et al. 2016 (1, −14, 39) Hester et al. 2004*
	Pre-SMA	roi3_Pre-SMA_L roi4_Pre-SMA_R			(−6, 8, 49) Gauvin et al. 2016 (11, −9, 53) Hester et al. 2004*
Cerebellum	Posterior	roi5_RCB1_R roi6_RCB2_R		(37.9, −63.7, −29.7) (12.5, −86.1, −32.9) Stoodley and Schmahmann 2009*	
	Superior medial	roi7_SMC_L roi8_SMC_R		(−18, −59, −22) (16, −59, −23) Golfinopoulos et al. 2010*	
Temporal cortex	SPT	roi9_SPT_L		(−54, −30, 14) Okada and Hickok 2006	
	pSTG	roi10_pSTG_L roi11_pSTG_R	(−64.6, −33.2, 13.5) (69.5, −30.7, 5.2) Golfinopoulos et al. 2010*		

Asterisks indicate meta-analysis or model-based coordinates.

combined differently to prime lexical and nonlexical errors, e.g., "mole sail," "mole fence"). Exchanging the first letters of these combinations would result in a new word pair in one case ("sole mail," lexical error outcome) and in a nonword pair in the other case ("fole mence," nonlexical error outcome). All combinations for which the exchange of initial phonemes resulted in new word pairs (mole sail) were used also in reversed order (sole mail). An orthographic criterion was used for selecting stimuli. To control for the variable of phonetic distance of the word pair onsets across the conditions of interest, these were coded for the degree of shared phonetic features (place and manner of articulation plus voicing), being assigned a number ranging from 0 (phonetically distant words) to 2 (phonetically close words). This was deemed necessary because with decreasing phonetic distance between onsets speakers are more likely to exchange onsets (e.g., Nooteboom and Quené 2008). We also included this variable in all analyses and we report the corresponding results in the supplementary information (Supplementary Tables 2–4 and Supplementary Fig. 1). A total of 102 pairs shared 0 features, 161 pairs shared 1 feature, and 57 pairs shared 2 features. The stimuli across the lexical and nonlexical conditions did not differ in the average amount of shared features (lexical 0.9 shared features vs. nonlexical 0.8 shared features, $P = .47$). The words in the target pairs were selected with the criterion that they should be semantically unrelated. A given participant was only presented with one combination for each word (lexical or nonlexical outcome) and was only presented with one of the words differing in only the first sound (mole or sole). During the experiment, 3 priming word pairs preceded each target word pair. The first 2 shared the initial consonants, and the third pair had further phonological overlap with the error being primed ("sun mall"—"sand mouth"—"soap mate"—"mole sail"). To induce errors, the order of the 2 initial consonants (/s/ and /m/) is different for the primes and the target. Participants were also presented with 140 filler pairs that had no specific relationship to their corresponding target pairs. One to 3 filler pairs were presented before each prime and target sequence. Thus, each participant was presented with 460 unique word combinations (80 targets of which 40 lexical and 40 nonlexical error outcome, 240 primes and 140 fillers). Each participant completed 6 experimental runs in which word pairs were repeated 3 times in different orders. Eight lists with a different randomization of the stimuli sequences were created.

Procedure

Word pairs remained on the screen for 748 ms. Words presented for silent reading were followed by a blank screen for 340 ms. All targets and 40% of the filler items were followed by a question mark for 544 ms, replaced by an exclamation mark presented 544 ms after the presentation of the question mark and remaining for 1020 ms. Before the next trial started there was a blank screen for 544 ms in the case of filler production trials and jittered between 544 and 1564 in the case of target production trials. The jittered inter stimulus interval was generated according to an exponential function and randomized across runs (e.g., Henson 2007). Participants were instructed to silently read the word pairs as they appeared, naming aloud the last word pair they had seen whenever a question mark was presented and before the appearance of an exclamation mark. Stimulus presentation and recording of productions to be processed offline were controlled

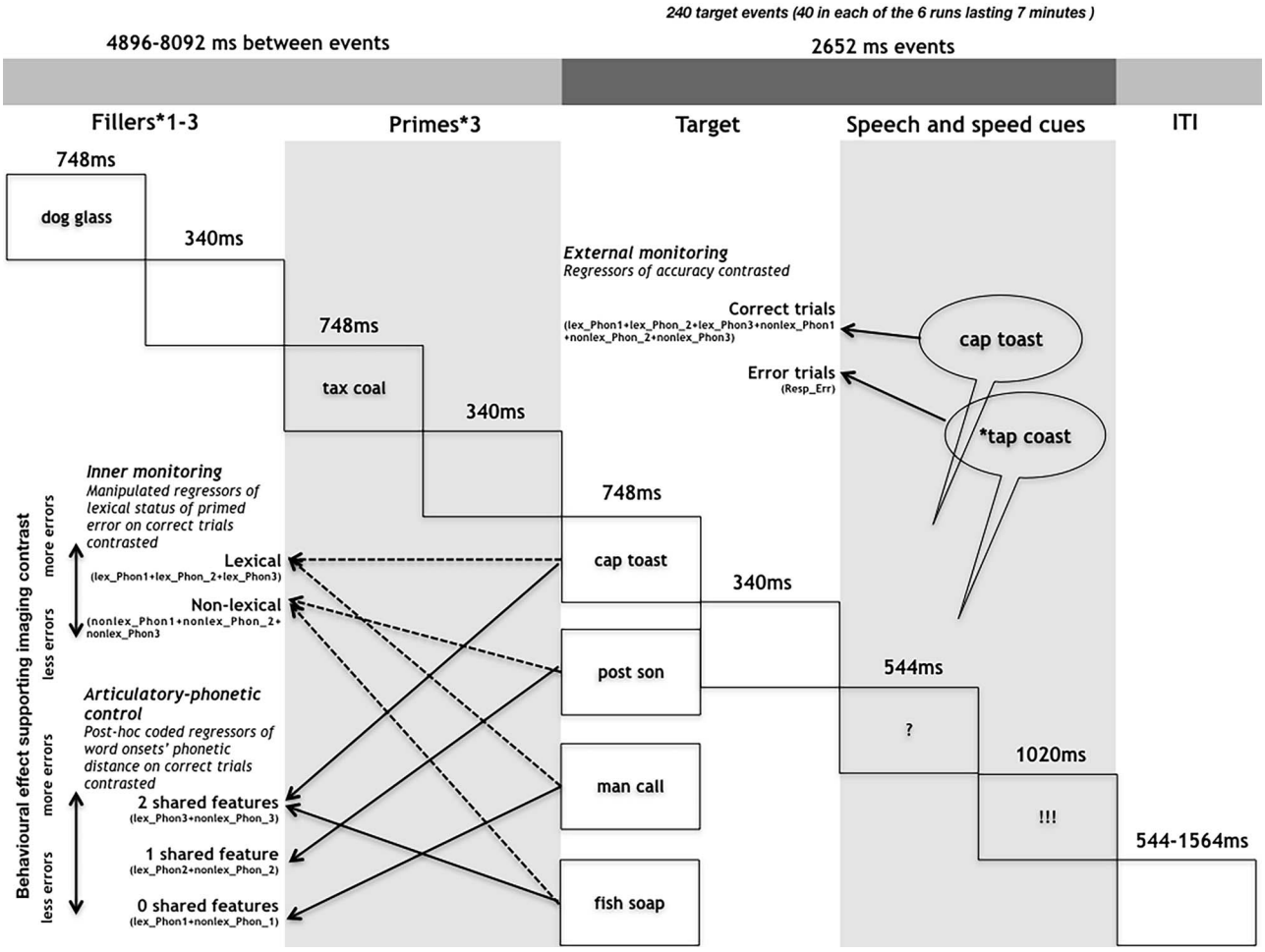


Figure 1. Depiction of the experimental design and procedure.

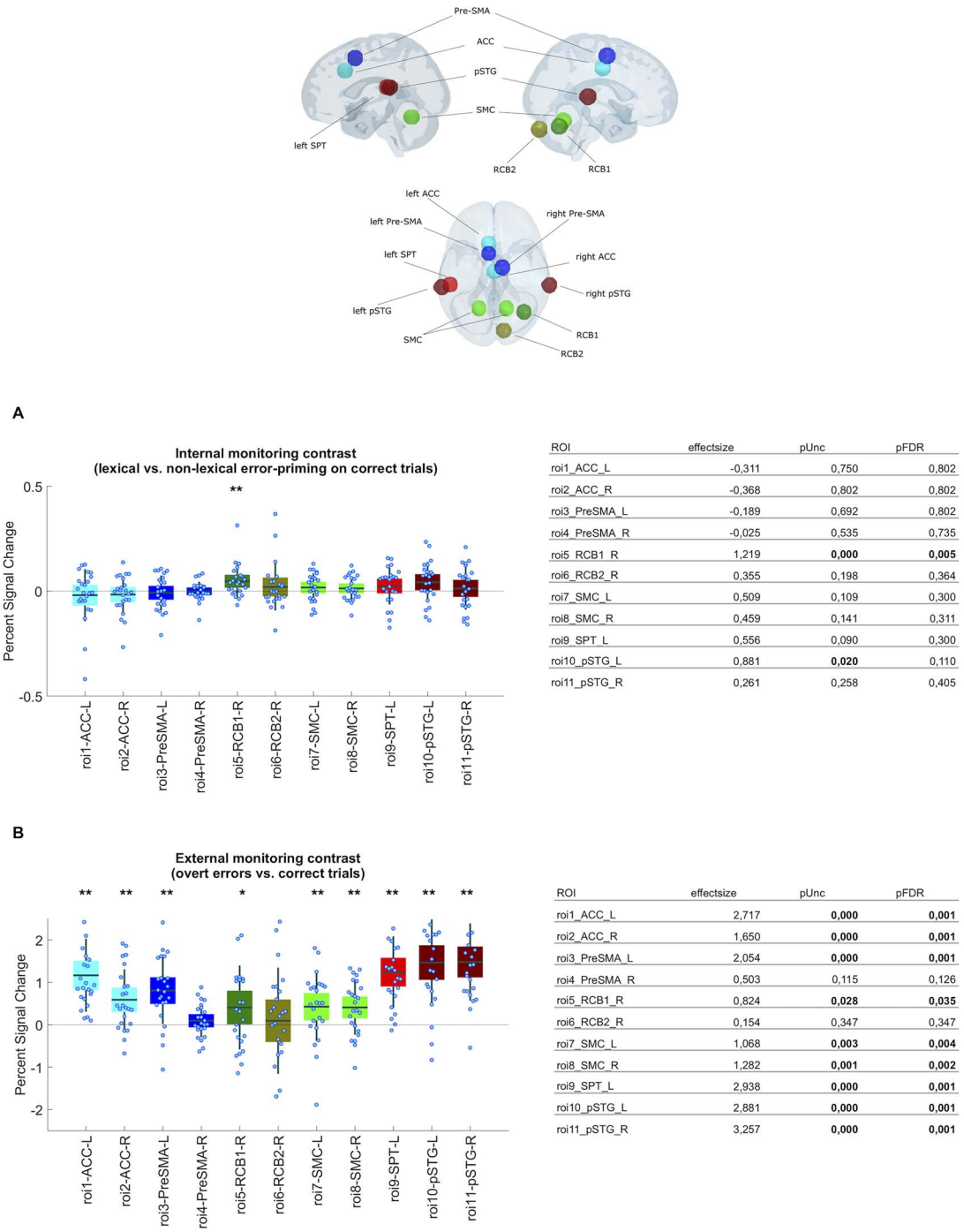
by a custom-made presentation software compiled using the LabVIEW development environment (National Instruments).

MRI Data Acquisition

Data were collected on a 3-Tesla Siemens Prisma Scanner (Siemens, Erlangen, Germany) at the Marseille MRI Center (Centre IRM-INT@CERIMED, UMR7289 CNRS & AMU) using a 64-channel head coil. Functional images (EPI sequence, 54 slices per volume, multi-band accelerator factor 3, repetition time = 1.224 s, spatial resolution = $2.5 \times 2.5 \times 2.5$ mm, echo time = 30 ms, flip angle = 65°) covering the whole brain were acquired during the task performance. Whole-brain anatomical MRI data were acquired using high-resolution structural T_1-weighted image (MPRAGE sequence, repetition time = 2.4 s, spatial resolution = $0.8 \times 0.8 \times 0.8$ mm, echo time = 2.28 ms, flip angle = 8°) in the sagittal plane. Prior to functional imaging, fieldmap image (dual echo gradient-echo acquisition, repetition time = 7.06 s, spatial resolution = 2.5 mm^3, echo time = 59 ms, flip angle = 90°) was also acquired.

Behavioral Data Processing and Analyses

A person naïve to the purpose of the experiment transcribed all spoken productions and inspected and codified vocal response onsets of all individual recordings using Check-vocal (Protopapas 2007). Check-vocal is a software that allows for semiautomatic codification of the response accuracy and timing based on 2 sources of information: the speech waveform and the spectrogram. The transcriptions were scored as correct, dysfluencies, partial responses (e.g., only 1 word produced), full omissions, and erroneous productions. Errors were classified as "priming-related errors" or "other errors." Priming-related errors included full exchanges (*mill pad* => *pill mad*), anticipations (*mill pad* => *pill pad*), perseverations (*mill pad* = > *mill mad*), repaired and interrupted exchanges (*mill pad* => *pi . . . mill pad*), full and partial competing errors (*mill pad* => *pant milk/pant pad*), and other related errors (*mill pad* => *mad pill*). Other errors included diverse phonological substitutions that were unrelated to the priming manipulation (e.g., *mill pad* => *chill pant/gri..mill pad/ . . . pant*). To assess the presence of a lexical bias and validate our assumption of a difference in monitoring load across our experimental conditions, errors were analyzed using the lme4 package (Bates et al. 2015) in R version 3.2.2 (R Development Core Team 2015). We used generalized linear mixed models (GLMM) with a binomial link function (Jaeger 2008), estimating the conditional probability of a response given the random effects and covariate values. For completeness, response times were also analyzed though we did not have any specific predictions for these. This was done using linear mixed models (LMM), estimating the influence of fixed and random covariates on the response. The summary output of the GLMM function of lme4 in R provides *P* values based on asymptotic Wald tests, which is common practice for generalized linear models (e.g., Bolker et al. 2009). In contrast, the summary



ROI	effectsize	pUnc	pFDR
roi1_ACC_L	-0,311	0,750	0,802
roi2_ACC_R	-0,368	0,802	0,802
roi3_PreSMA_L	-0,189	0,692	0,802
roi4_PreSMA_R	-0,025	0,535	0,735
roi5_RCB1_R	1,219	**0,000**	**0,005**
roi6_RCB2_R	0,355	0,198	0,364
roi7_SMC_L	0,509	0,109	0,300
roi8_SMC_R	0,459	0,141	0,311
roi9_SPT_L	0,556	0,090	0,300
roi10_pSTG_L	0,881	**0,020**	0,110
roi11_pSTG_R	0,261	0,258	0,405

ROI	effectsize	pUnc	pFDR
roi1_ACC_L	2,717	**0,000**	**0,001**
roi2_ACC_R	1,650	**0,000**	**0,001**
roi3_PreSMA_L	2,054	**0,000**	**0,001**
roi4_PreSMA_R	0,503	0,115	0,126
roi5_RCB1_R	0,824	**0,028**	**0,035**
roi6_RCB2_R	0,154	0,347	0,347
roi7_SMC_L	1,068	**0,003**	**0,004**
roi8_SMC_R	1,282	**0,001**	**0,002**
roi9_SPT_L	2,938	**0,000**	**0,001**
roi10_pSTG_L	2,881	**0,000**	**0,001**
roi11_pSTG_R	3,257	**0,000**	**0,001**

Figure 2. Percent signal change in the 11 predefined ROI (location in the brain in top central panel) for (*A*) the internal monitoring contrast and (*B*) the external monitoring contrast. ROIs in medial frontal cortex are represented with blue tones, ROIs in the cerebellum in green tones and ROIs in temporal cortex in red tones. The asterisks indicate significant effects <0.05 (*) or <0.005 (**) after correcting for multiple comparisons using FDR.

output of the LMM function only provides *t*-values. Consequently, we report *P* values for error rates and *t*-values for response times. Following common practice (e.g., Fisher 1925), we take *t*-values to approximate *z*-scores and assume that absolute values above 1.96 reflect significant effects.

To assess the effect of the manipulated variable lexical status of primed errors and the control variable phonetic distance of the word pair onsets on priming-related errors, separate models were fitted for the 2 variables. The models included crossed random effects for subjects and items and the fixed factor lexicality or phonetic distance. Additional models including the same fixed and random variables were conducted on all errors and are reported in the Supplementary Tables 1–3. A histogram visualization of the response time data indicated a non-normal

distribution. Therefore, log-transformed response times were modeled with mixed linear models. All models included the crossed random factors subject and item. For correct trials a first model included the fixed factor lexicality. Another model included the fixed factor shared phonetic features. A final model on all responses (i.e., both correct and incorrect trials) included the fixed factor accuracy.

Image Processing and Analyses

The fMRI data were preprocessed and analyzed using the Statistical Parametric Mapping software (SPM12, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) on MATLAB R2018b (Mathworks Inc., Natick, MA). The anatomical scan was spatially normalized to the avg152 T_1-weighted brain template defined by the Montreal Neurological Institute using the default parameters (nonlinear transformation). The Fieldmap images were used during the realign and unwarp procedure for distortion and motion correction. Functional volumes were spatially realigned and normalized (using the combination of deformation field, coregistered structural and sliced functional images) and smoothed with an isotropic Gaussian kernel (full-width at half-maximum = 5 mm). The Artefact Detection Tools (ART) implemented in the CONN toolbox (www.nitrc.org/projects/conn, RRID:SCR_009550) was used to define the regressors of no interest related to head movements and functional data outliers (see next section). Automatic ART-based identification of outlier scans used a 97th percentiles superior to normative samples in the definition of the outlier thresholds (global-signal z-threshold of 5 and subject-motion threshold of 0.9 mm).

For the univariate analysis on the whole brain, a general linear model (GLM) was generated for each subject. The GLM included, for each of the 6 runs, 7 regressors modeling response accuracy, lexical status of error priming and phonetic distance of target pair onsets: Resp_ER, lex_Phon1_CR, lex_Phon2_CR, lex_Phon3_CR, nonlex_Phon1_CR, nonlex_Phon2_CR, nonlex_Phon3_CR (CR for correct responses and ER for errors). For the contrast targeting internal monitoring, we contrasted lex_Phon1_CR, lex_Phon2_CR, and lex_Phon3_CR, with nonlex_Phon1_CR, nonlex_Phon2_CR, and nonlex_Phon3_CR. For the contrast targeting external monitoring, we contrasted Resp_ER with lex_Phon1_CR, lex_Phon2_CR, lex_Phon3_CR, nonlex_Phon1_CR, nonlex_Phon2_CR, and nonlex_Phon3_CR. For the articulatory-phonetic control, we contrasted lex_Phon1_CR and nonlex_Phon1_CR with lex_Phon3_CR, and nonlex_Phon3_CR. In the GLM, the regressors of no interest were also included using an ART text file per subject (each file described outlier scans from global signal and head movements from ART). Regressors of interest were convolved with the canonical hemodynamic response function, and the default SPM autoregressive model AR(1) was applied. Functional data were filtered with a 128 s high-pass filter. Statistical parametric maps for each experimental factor and each participant were calculated at the first level and then entered in a second-level 1-sample t-test analysis of variance (random effects analysis or RFX). All statistical comparisons were performed with a voxelwise threshold of $P < .001$ and a cluster extent threshold of 25 voxels. For the univariate analysis on ROIs, 11 anatomical ROIs were created based on the previous literature (Table 1). ROIs with a MNI coordinates center and a 10-mm-radius were created using the MarsBar SPM toolbox (Brett et al. 2002) and applying a mask that only extracted voxels pertaining to gray matter. For a given ROI mask and on the basis on unsmoothed functional images, we extracted each subject's percent signal changes using MarsBar software (http://marsbar.sourceforge.net/). Percent signal changes were computed from canonical events using a MarsBar's function called "event_signal" (with "max abs" option) and averaged across voxels within a ROI. From each contrast ("internal monitoring", "external monitoring," and "articulatory-phonetic control"), we obtained a vector of 24% signal changes (1 per subject) per ROI (n = 11). For each ROI, we performed permutation tests (from Laurens R Krol, see https://github.com/lrkrol/permutationTest) to compare the distribution of the percent signal changes to the null hypothesis (normal distribution). Statistical tests were conducted using 2000 permutations and false discovery rate (FDR) was used to correct for multiple comparisons (Benjamini and Hochberg 1995).

Results

Out of the 5760 target trials across all participants, 706 resulted in errors (12.3%, mean standard error (MSE) 0.4, SD 32.8), of which 155 (2.7%, MSE 0.2, SD 16.2) were related to the priming manipulation. For the subset of 155 priming-related errors, more errors were made in the lexical outcome condition (3.9%, MSE 0.4, SD 19.4) than in the nonlexical outcome condition (1.5%, MSE 0.2, SD 11.9; $P < .001$; Table 2 A). This validates the assumption that, also in the present dataset, the lexical condition was more error prone and required more monitoring. As in the previous literature, no significant differences were observed in the response times between the lexical (419 ms) and nonlexical (417 ms) outcome conditions (e.g., Hartsuiker et al. 2005; Runnqvist et al. 2016). Replicating previous findings (e.g., Gauvin et al. 2016), correct trials (418 ms) were produced faster than trials with errors (506 ms; Table 2 B).

Using MNI coordinates reported in the previous literature (Table 1), we examined percent signal change for our 2 contrasts in 11 predefined ROIs located in temporal, cerebellar, and medial frontal regions. A ROI in the right posterior cerebellum was involved both in the contrast targeting external monitoring ($q = 0.035$, $d = 0.82$) and in the internal monitoring of words ($q = 0.005$, $d = 1.22$; Fig. 2). Furthermore, external monitoring was also linked to bilateral superior medial cerebellum (left $q = 0.004$, $d = 1.07$; right $q = 0.002$, $d = 1.28$), bilateral ACC (left $q < 0.001$, $d = 2.72$; right $q < 0.001$, $d = 1.65$), left pre-SMA ($q < 0.001$, $d = 2.05$), region SPT ($q < 0.001$, 2.94), and bilateral pSTG (left $q < 0.001$, $d = 2.88$; right $q < 0.001$, $d = 3.26$).

To follow up on the potential differences in internal and external monitoring, we directly compared the external monitoring contrast with the internal monitoring contrast. The effects were larger for the former compared to the latter in bilateral superior medial cerebellum (left $q = 0.022$, $d = -0.94$; right $q = 0.003$, $d = -1.14$), bilateral ACC (left $q < 0.001$, $d = -2.41$; right $q < 0.001$, $d = -1.51$), left pre-SMA ($q < 0.001$, $d = -1.88$), region SPT ($q < 0.001$, $d = -2.61$), and bilateral pSTG (left $q < 0.001$, $d = -2.56$; right $q < 0.001$, $d = -2.90$).

To examine the specificity of the findings from the ROI analyses, we also conducted a whole-brain analysis (Table 4 and Fig. 3). In the internal word monitoring contrast, only the BOLD response of a cluster in the left posterior cerebellum (lobule VI) survived the correction for multiple comparisons. For the contrast targeting external monitoring, significant clusters of differential BOLD response were observed in frontal, medial frontal, temporal, insular, and parietal regions in cortex as well as regions in basal ganglia. Table 3 summarizes all the analyses that were carried out.

In summary, both the contrast targeting internal monitoring of words and the contrast targeting external monitoring of errors

Table 2. Summary of the GLMM of priming-related errors (A) and the LMMs on response times (RTs) (B)

A: Errors	**Effect estimate**	**Std.err.**	**z-value**	***P* value**
Intercept	−3.70	0.20	−18.40	<.001
Lexical status (nonlexical)	−1.06	0.22	−4.85	<.001
B: RTs	**Effect estimate**	**In ms**	**Std.err.**	**t-value**
Intercept	5.98	419	0.04	147 0.77
Lexical status (nonlexical)	−0.003	−2	−0.01	0.269
Intercept	6.17	504	0.04	149.99
Accuracy (error)	−0.19	−86	0.01	−13.26

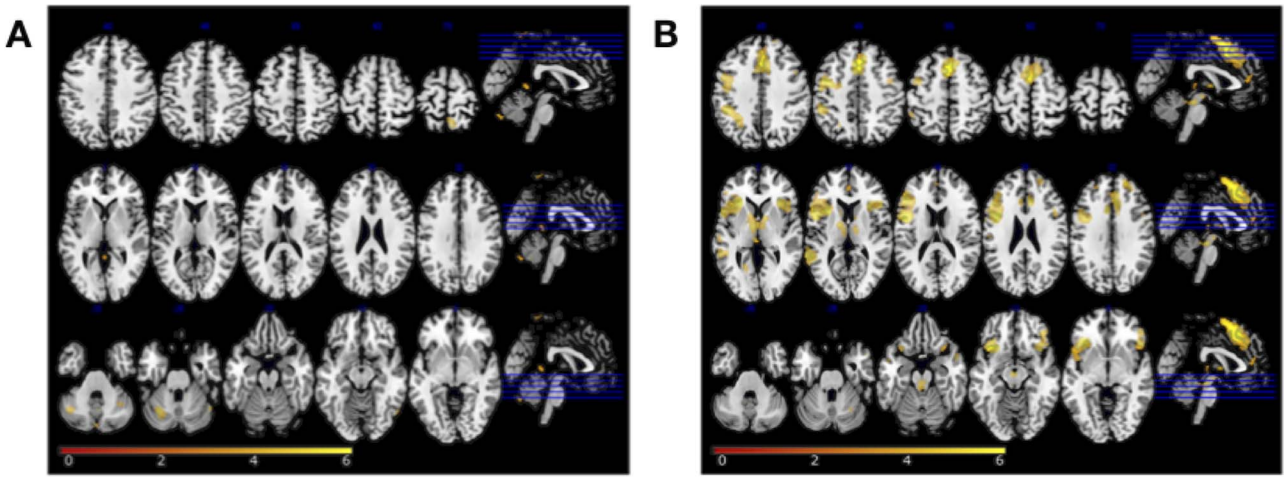


Figure 3. RFX results on the BOLD response of internal monitoring (lexical vs. nonlexical error priming; panel A) and external monitoring (errors vs. correct trials; panel B). Statistical *t*-maps are overlaid on MNI cortex slices (5 axial slices and 1 sagittal slice per line) using a voxelwise threshold of $P < .001$ and an extent threshold of 25 voxels.

Table 3. Summary of the different analyses conducted

	Analysis	Contrasted variables	Purpose
Behavioral data	Generalized mixed linear model on priming-related errors	Lexical versus nonlexical error priming	Validate monitoring load assumption underlying imaging contrast
	LMM on response times	Lexical versus nonlexical error priming Errors versus correct trials	
Brain data	Analysis on percent signal change in 10 mm spherical predefined ROI	Lexical versus nonlexical error priming	Index internal monitoring
		Errors versus correct trials	Index external monitoring
		(lexical vs. nonlexical error priming) versus (errors vs. correct trials)	Compare internal and external monitoring
	Univariate whole brain analysis on BOLD response	Lexical versus nonlexical error priming	Assess specificity of ROI findings
		Errors versus correct trials	

revealed a differential percent signal change in the right posterior cerebellum. The latter contrast also revealed a differential percent signal change in superior medial cerebellum and of temporal and medial frontal regions.

Discussion

In this study, we explored the neural basis of the cognitive mechanisms that allow speakers to monitor their speech, both internally during planning and externally during articulation. Concretely, we aimed at answering 1) whether the pSTG has a role in the monitoring of actual speech errors indicating comprehension-based monitoring; 2) whether the cerebellum would be involved in inner monitoring (posterior) and/or external monitoring (superior medial), indicating forward modeling; and 3) whether the medial frontal cortex would be involved in the presence of inner monitoring load, indicating conflict-based monitoring. In the following we discuss how the results answered these questions.

Temporal Cortex and Monitoring of Speech Errors

All 3 ROIs in temporal cortex (bilateral pSTG and SPT) showed a differential percent signal change for speech errors compared

Table 4. Results of the whole-brain analyses of the BOLD response of the external (A) and internal (B) monitoring contrasts

A. External monitoring (errors vs. correct trials)			MNI coordinates		
Region label	Extent	*t*-value	x	y	z
L superior medial gyrus	1244	8.064	−3	22	48
L posterior–medial frontal	1244	5.908	−3	2	63
R ACC	1244	6.373	8	27	28
L ACC	36	3.970	0	44	13
L inferior frontal gyrus (pars opercularis)	2343	7.592	−48	4	16
L inferior frontal gyrus (pars orbitalis)	2343	7.265	−35	29	−2
R inferior frontal gyrus (pars orbitalis)	542	5.685	45	39	−10
R inferior frontal gyrus (pars triangularis)	51	5.742	48	14	28
R superior frontal gyrus	135	5.538	20	52	36
L precentral gyrus	2343	6.923	−45	−4	51
L middle temporal gyrus	195	7.248	−58	−46	8
L middle temporal gyrus	65	4.829	−60	−21	1
L thalamus proper	349	6.590	−10	−6	6
R pallidum	349	5.354	13	4	3
Brain stem	71	6.120	0	−29	−17
L dorsal caudal	28	4.828	−5	−14	−12
R insula lobe	542	5.756	35	24	3
L inferior parietal lobule	334	5.662	−45	−41	41
R cerebellum (VI)	26	4.109	30	−61	−30
B. Internal monitoring (lexical vs. nonlexical error priming)					
L cerebellum (VI)	160	4.994	−28	−66	−27
R cerebellum (VIII)	36	3.917	35	−51	−42
R precuneus	51	4.644	5	−56	71

Local maxima of BOLD response separated by >20 mm. Regions were automatically labeled using the Anatomy Toolbox atlas. x, y, and z = MNI coordinates in the left–right, anterior–posterior and inferior–superior dimensions, respectively. All peaks are significant at a voxelwise threshold of $P < .001$ (extent threshold = 25 voxels). Peaks that are significant at a cluster threshold of $t < .05$ with an FDR correction for multiple comparisons are marked with bold fonts. L = left, R = right.

to correct trials. Hence, some form of comprehension-based monitoring likely takes place in the case of overt speech errors (or more strongly for speech errors compared with correct utterances). The current study cannot answer whether such comprehension-based monitoring is carried out through speech comprehension processes directly, through feedback control processes (comparing auditory percepts and targets) or in the form of increased response for unexpected input (and thus connected to the cerebellar forward modeling that will be discussed later on). Importantly, this is the first study showing a role of pSTG/SPT for an overt speech production task involving the articulation of natural speech errors. The whole-brain analysis of the BOLD response for the contrast targeting external monitoring revealed 2 clusters peaking in the left middle temporal gyrus and in the left inferior parietal lobule, respectively, given their extent, that likely comprise the voxels targeted by the ROI coordinates. Thus, the whole-brain analysis seems to further confirm the ROI results and also sheds light on the fact that these results are not very specific as rather large portions of temporal and parietal cortex are differentially active for errors compared to correct trials.

Cerebellum Involved in both Internal and External Monitoring

The contrast targeting internal monitoring showed a differential percent signal change in a region in the right posterior cerebellum that has been attributed an important role in the forward modeling of self-generated actions (e.g., Imamizu et al. 2000; Blakemore et al. 2001; Ito 2008; Miall and King 2008; Strick et al. 2009). To our knowledge, this is the first time that the involvement of the cerebellum in the internal monitoring of an unambiguously linguistic aspect of language production has been reported. While previous studies have reported an involvement of the cerebellum for articulatory–acoustic aspects of speech, here the involvement was modulated by lexical information, a level of language processing that is distinct from the sensory–motor aspects of speech. One possibility is that this occurs because in language use sound and meaning always cooccur. Over time, this arguably leads the 2 dimensions to form an interconnected distributed representation (Strijkers 2016; Fairs et al. 2021). This holistic format of linguistic representations would entail that sound and meaning dimensions would become active in parallel both when producing and understanding

speech, hence over time also sharing processing dynamics. In this way, motor control processes could be directly applied to any level of language processing. Another, not mutually exclusive, possibility is that all self-generated actions, whether motor or mental, may be supervised through forward modeling enabled by cerebellar connections to different areas of cortex (Ito 2008; Strick et al. 2009). The cerebellum would generate the prediction of the sensory or mental consequences of the action (efference copying), whereas the cortical region in question would be in charge of inhibiting the neural response that the action is expected to generate. In the case of language, the modeling of different levels of linguistic representation might result in reafference cancelation in different areas of cortex. Regardless the exact mechanism, the link between cerebellum activity and a processing level in principle distant from articulation calls for an extended role of the cerebellum (i.e., beyond speech motor control) in current models (Golfinopoulos et al. 2010; Hickok 2012).

Secondly, the contrast targeting external monitoring showed a differential percent signal change of the same right cerebellar region as internal monitoring and also a differential percent signal change bilaterally of the ROIs located in the superior medial cerebellum. This latter region has been linked to articulatory difficulties such as ataxic dysarthria and hence speech motor control troubles. An interesting possibility is that the posterior cerebellar activation might be especially due to the lexical and fluent errors (being more similar to the effect of the inner monitoring contrast) and the superior medial cerebellar activation might be especially due to nonlexical or more dysfluent errors. Unfortunately, however, while we are able to pinpoint an exact level of processing for our inner monitoring contrast thanks to the error priming manipulation, for the errors this was not possible because overt nonlexical errors are so rare that not all participants have observations for these. For the same reason, the errors included in the external monitoring contrast are also diverse in nature (i.e., all errors were pooled together and contrasted with correct responses). Finally, given that in the external monitoring contrast the cerebellar activation was accompanied by pSTG/SPT activation, a parsimonious assumption is that the less predictable auditory response associated with an error led to a lowered reafference cancelation.

Turning to the whole-brain analyses of the BOLD response, unexpectedly, a region in the left posterior cerebellum was differentially activated in the contrast targeting internal monitoring. With the aim of guiding future hypotheses concerning language processing and monitoring in the cerebellum we visualized the peak coordinates of the cluster in an atlas viewer of the cerebellum, SUIT (e.g., Diedrichsen 2006), allowing to overlay different task contrast maps onto an anatomical template. Nine contrasts overlapping with the observed region could be more or less directly linked to the current task contrast through the notion of (verbal) working memory (object 2 back, object 2 back+, verbal 2 back, and verbal 2 back+); prediction outcome (true, violated, and scrambled predictions); and response difficulty (easy and medium responses). Broadly, all 3 groupings are consistent with the notion of increases in monitoring load engaging processes of forward modeling (e.g., Runnqvist et al. 2016). More generally, this result shows that the left cerebellum should not be neglected in studies of language where it is often assumed that cerebellar contributions to language processing are right lateralized. Furthermore, the results of the whole-brain analysis highlight the fact that cerebellar activity is elusive and may go undetected without an appropriate task (sufficiently demanding), analysis of different task stages (early vs. late stages, e.g., Imamizu et al. 2000), or statistical approach (such as a ROI approach, see Johnson et al. 2019, for an extended argumentation).

Medial Frontal Cortex for External Monitoring

For the contrast targeting external monitoring, we observed a differential percent signal change bilaterally for ACC as well as for left pre-SMA in our ROI analyses. Previous studies contrasting errors and correct trials have reported a similar pattern and this has been interpreted in terms of conflict-based monitoring (Gauvin et al. 2016; Riès et al. 2011). However, no such differential percent signal change in medial frontal cortex was observed for inner monitoring, and when comparing both contrasts directly the difference was significant (i.e., more ACC and pre-SMA percent signal change in external compared with internal monitoring). Consistent with this, the whole-brain analyses of the BOLD response revealed 2 very broadly extended clusters in the left superior medial gyrus and 1 in the right ACC only for the external monitoring contrast. As for temporal cortex, given their size, they are likely to comprise the voxels targeted by our ROI but again show that the activation is much more extended than these. While the current study cannot add much anatomical specificity to the debate, the dissociation of medial frontal activity for the internal and external monitoring contrasts hints that conflict is not the mechanism behind the differential percent signal change and BOLD response. One possibility is that the involvement of the medial frontal cortex observed here is related to a vocal cognitive control network shared across primates as proposed recently by Loh et al. (2020). These authors argue that, across primates, area 44 is in charge of cognitive control of orofacial and nonspeech vocal responses, and the midcingulate cortex is in charge of analyzing vocal nonspeech feedback driving response adaptation. Furthermore, the cognitive control of human-specific speech vocal information would require the additional recruitment of area 45 and pre-SMA. In this framework, it would not be the conflict that generates the ACC and pre-SMA percent signal change and BOLD response observed here but rather the feedback provided through the articulated error. An advantage of this feedback-based network account of vocal cognitive control is that it also predicts the BOLD response clusters in the left inferior frontal gyrus that we observed (while a conflict account would rather predict activation in dorsolateral prefrontal cortex, e.g., MacDonald et al. 2000).

In summary, monitoring for errors during speech production seems to rely on a broad network of brain regions that can be linked to different monitoring mechanisms (e.g., modeling of self-generated actions, cognitive control, and sensorial perception) in accordance with previous findings reported in the literature. Importantly, however, this is the first time that multiple monitoring mechanisms are investigated simultaneously in the context of both speech planning and articulation, allowing us to show that certain regions (pSTG, SPT, ACC, and pre-SMA) seem to be implicated preferentially in the context of overt errors and thus seemingly more dependent on the sensorial feedback. However, perhaps the most striking result is that the same posterior part of the right cerebellum is involved both in inner and external monitoring, a finding that is challenging for all current brain models of language production. The results reported here show the importance of adopting a broad approach when addressing complex cognitive processes like error monitoring of multidimensional representations (language) at the service of a combined mental and motor action (speaking). Previous studies may have failed to detect the involvement of certain monitoring

regions because only 1 ROI or only 1 manipulation of monitoring demands were examined at the same time. To be addressed in future research is whether these different functional regions are competitively or collaboratively interconnected or whether they are instances of partially redundant cognitive mechanisms that, in an analogous way to redundant input in the environment, could serve to increase the likelihood of detecting and correcting errors in noisy neural communication channels (Barlow 2001).

Supplementary Material

Supplementary material can be found at *Cerebral Cortex Communications* online.

Funding

This work was performed in the Centre IRM-INT (UMR7289, AMU-CNRS), platform member of France Life Imaging network (grant ANR-11-INBS-0006). The study received funding from "Investissements d'Avenir" French government program managed by the French National Research Agency (reference: ANR-16-CONV-0002) and from Excellence Initiative of Aix-Marseille University—"A*MIDEX" through the Institute of Language, Communication and the Brain; French National Agency for Research, French government (ANR grant ANR-18-CE28-0013 to E.R., ANR-16-CE28-0007-01 to K.S., ANR-16-CE37-011-01 to P.B.); European Research Council (788240-COVOPRIM to P.B.).

Notes

Conflict of Interest: None declared.

References

Abel S, Dressel K, Weiller C, Huber W. 2012. Enhancement and suppression in a lexical interference fMRI-paradigm. *Brain Behav*. **2**(2):109–127.

Ackermann H, Vogel M, Petersen D, Poremba M. 1992. Speech deficits in ischaemic cerebellar lesions. *J Neurol*. **239**(4):223–227.

Argyropoulos GP. 2016. The cerebellum, internal models and prediction in 'non-motor' aspects of language: a critical review. *Brain Lang*. **161**:4–17.

Baars BJ, Motley MT, MacKay DG. 1975. Output editing for lexical status in artificially elicited slips of the tongue. *J Verbal Learn Verbal Behav*. **14**(4):382–391.

Barlow H. 2001. Redundancy reduction revisited. *Network*. **12**(3):241–253.

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Scheipl FG, Grothendieck G, Green P et al. 2015. Package 'lme4. *Convergence*. **12**:1.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B*. **57**(1):289–300.

Blakemore SJ, Frith CD, Wolpert DM. 2001. The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*. **12**(9):1879–1884.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH , White JSS. et al. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol*. **24**(3):127–135.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. *Psychol Rev*. **108**(3):624.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of Interest Analysis Using an SPM Toolbox. In: *8th International conference on functional mapping of the human brain*. June 2–6, Sendai, Japan **16**(2):497.

Christoffels IK, Formisano E, Schiller NO. 2007. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum Brain Mapp*. **28**(9):868–879.

Dell GS. 1986. A spreading-activation theory of retrieval in sentence production. *Psychol Rev*. **93**(3):283–321.

Desmond JE, Fiez JA. 1998. Neuroimaging studies of the cerebellum: language, learning and memory. *Trends Cognit Sci*. **2**(9):355e362.

De Zubicaray GI, Wilson SJ, McMahon KL, Muthiah M. 2001. The semantic interference effect in the picture-word paradigm: an event-related fMRI study employing overt responses. *Hum Brain Mapp*. **14**(4):218–227.

Diedrichsen J. 2006. A spatially unbiased atlas template of the human cerebellum. *Neuroimage*. **33**(1):127–138.

D'Mello AM, Gabrieli JD, Nee DE. 2020. Evidence for hierarchical cognitive control in the human cerebellum. *Curr Biol*. **30**(10):1881–1892.

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*. https://doi.org/10.1093/texcom/tgab040.

Fisher RA. 1925. Theory of statistical estimation. *Math Proc Cambridge Philos Soc*. **22**(5):700–725.

Gauvin HS, De Baene W, Brass M, Hartsuiker RJ. 2016. Conflict monitoring in speech processing: an fMRI study of error detection in speech production and perception. *NeuroImage*. **126**:96–105.

Gauvin HS, Hartsuiker RJ. 2020. Towards a new model of verbal monitoring. *J Cognit*. **3**(1):1–37.

Ghosh SS, Tourville JA, Guenther FH. 2008. A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *J Speech Lang Hearing Res*. **51**(5):1183–1202.

Golfinopoulos E, Tourville JA, Guenther FH. 2010. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*. **52**(3):862–874.

Guenther FH, Ghosh SS, Tourville JA. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*. **96**(3):280–301.

Hartsuiker RJ, Kolk HH. 2001. Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognit Psychol*. **42**(2):113–157.

Hartsuiker RJ, Corley M, Martensen H. 2005. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: related reply to Baars, Motley, and MacKay (1975). *J Memory Lang*. **52**:58–70.

Henson R. 2007. Efficient experimental design for fMRI. In: Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, editors. *Statistical parametric mapping: the analysis of functional brain images*. San Diego, Elsevier. 193–210.

Hester R, Fassbender C, Garavan H. 2004. Individual differences in error processing: a review and reanalysis of three event-related fMRI studies using the GO/NOGO task. *Cereb Cortex*. **14**(9):986–994.

Hickok G. 2012. Computational neuroanatomy of speech production. *Nat Rev Neurosci*. **13**(2):135–145.

Hickok G. 2014. The architecture of speech production and the role of the phoneme in speech processing. *Lang Cognit Neurosci*. **29**(1):2–20.

Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H et al. 1997. Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport*. **8**:2379–2382.

Imamizu H, Miyauchi S, Tamada T, Sasaki Y, Takino R, Puetz B, Kawato M. 2000. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*. **403**(6766):192–195.

Indefrey P, Levelt WJ. 2004. The spatial and temporal signatures of word production components. *Cognition*. **92**(1):101–144.

Ito M. 2008. Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci*. **9**(4):304–313.

Jaeger TF. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J Memory Lang*. **59**(4):434–446.

Jeannerod M. 1988. *The neural and behavioural organization of goal-directed movements*. Oxford: Clarendon Press/Oxford University Press.

Johnson JF, Belyk M, Schwartze M, Pinheiro AP, Kotz SA. 2019. The role of the cerebellum in adaptation: ALE meta-analyses on sensory feedback error. *Hum Brain Mapp*. **40**(13):3966–3981.

Leggio MG, Silveri MC, Petrosini L, Molinari M. 2000. Phonological grouping is specifically affected in cerebellar patients: a verbal fluency study. *J Neurol Neurosurg Psychiatry*. **69**(1):102–106.

Lesage E, Hansen PC, Miall RC. 2017. Right lateral cerebellum represents linguistic predictability. *J Neurosci*. **37**(26):6231–6241.

Levelt WJ. 1983. Monitoring and self-repair in speech. *Cognition*. **14**(1):41–104.

Levelt WJ, Roelofs A, Meyer AS. 1999. A theory of lexical access in speech production. *Behav Brain Sci*. **22**(1):1–38.

Lind A, Hartsuiker RJ. 2020. Self-monitoring in speech production: comprehending the conflict between conflict-and comprehension-based accounts. *J Cognit*. **3**(1):1–4.

Loh KK, Procyk E, Neveu R, Lamberton F, Hopkins WD, Petrides M, Amiez C. 2020. Cognitive control of orofacial motor and vocal responses in the ventrolateral and dorsomedial human frontal cortex. *Proc Natl Acad Sci*. **117**(9):4994–5005.

MacDonald AW, Cohen JD, Stenger VA, Carter CS. 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*. **288**(5472):1835–1838.

McCloskey DI. 1981. Corollary discharges: motor commands and perception. *Compr Physiol*. 1415–1447.

Meekings S, Scott SK. 2021. Error in the superior temporal gyrus? A systematic review and activation likelihood estimation meta-analysis of speech production studies. *J Cognit Neurosci*. **33**(3):422–444.

Miall RC, King D. 2008. State estimation in the cerebellum. *Cerebellum*. **7**(4):572–576.

Möller J, Jansma BM, Rodriguez-Fornells A, Münte TF. 2007. What the brain does before the tongue slips. *Cereb Cortex*. **17**(5):1173–1178.

Nachev P, Rees G, Parton A, Kennard C, Husain M. 2005. Volition and conflict in human medial frontal cortex. *Curr Biol*. **15**(2):122–128.

Nooteboom S, Quené H. 2008. Self-monitoring and feedback: a new attempt to find the main cause of lexical bias in phonological speech errors. *J Memory Lang*. **58**(3):837–861.

Nozari N. 2020. A comprehension-or a production-based monitor? Response to Roelofs (2020). *J Cognit*. **3**(1):1–21.

Nozari N, Dell GS, Schwartz MF. 2011. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognit Psychol*. **63**(1):1–33.

Okada K, Hickok G. 2006. Left posterior auditory-related cortices participate both in speech perception and speech production: neural overlap revealed by fMRI. *Brain Lang*. **98**(1):112–117.

Oppenheim GM, Dell GS. 2008. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*. **106**(1):528–537.

Postma A, Noordanus C. 1996. Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Lang Speech*. **39**(4):375–392.

Protopapas A. 2007. Check vocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav Res Methods*. **39**(4):859–862.

Pickering MJ, Garrod S. 2013. An integrated theory of language production and comprehension. *Behav Brain Sci*. **36**(04):329–347.

Riès S, Janssen N, Dufau S, Alario FX, Burle B. 2011. General-purpose monitoring during speech production. *J Cognit Neurosci*. **23**(6):1419–1436.

Roelofs A. 2020. Self-monitoring in speaking: In defense of a comprehension-based account. *J Cognit*. **3**(1):1–13.

Runnqvist E, Bonnard M, Gauvin HS, Attarian S, Trébuchon A, Hartsuiker RJ, Alario FX. 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.

Severens E, Kühn S, Hartsuiker RJ, Brass M. 2012. Functional mechanisms involved in the internal inhibition of taboo words. *Soc Cognit Affective Neurosci*. **7**(4):431–435.

Shergill SS, Brammer MJ, Williams SCR, Murray RW, McGuire PK. 2000. Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Arch General Psychiatry*. **57**:1033–1038.

Stoodley CJ, Schmahmann JD. 2009. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*. **44**(2):489–501.

Strick PL, Dum RP, Fiez JA. 2009. Cerebellum and nonmotor function. *Annu Rev Neurosci*. **32**:413–434.

Strijkers K. 2016. A neural assembly–based view on word production: the bilingual test case. *Lang Learn*. **66**(S2):92–131.

Tian X, Poeppel D. 2010. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front Psychol*. **1**:166.

Tourville JA, Reilly KJ, Guenther FH. 2008. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. **39**(3):1429–1443.

Tourville JA, Guenther FH. 2011. The DIVA model: a neural theory of speech acquisition and production. *Lang Cognit Process*. **26**(7):952–981.

Wolpert DM, Ghahramani Z, Jordan MI. 1995. An internal model for sensorimotor integration. *Science*. **269**(5232):1880.

2 A robust temporal map of speech monitoring: from planning to articulation

Dorokhova Lydia
Morillon Benjamin
Baus Cristina
Belin Pascal
Dubarry Anne-Sophie
Alario François-Xavier
Runnqvist Elin

Outline

Abstract

Speakers continuously monitor their own speech to optimize fluent production. However, the precise timing and underlying variables influencing speech monitoring remain insufficiently understood. This study aimed to provide a comprehensive temporal map of monitoring processes ranging from speech planning to articulation. Two similar experiments were conducted, focusing on effects that consistently emerged across both. Participants engaged in a speeded language production task designed to elicit speech errors of either a lexical or articulatory-phonetic origin, while their EEG activity was recorded. On correctly produced utterances, we explored error probability at different levels of processing (lexical and articulatory-phonetic) and we also compared errors with correct trials to capture the potential diversity of response conflict and monitoring processes. Concerning the effects of error probability on correct trials, differences driven by the lexical status of a competing response were observed during initial stages of speech planning, while differences related to articulatory phonetically driven response competition emerged during speech motor preparation. In contrast, errors showed differences with correct utterances in both early and late speech motor preparation and during articulation. Taken together, these findings suggest that (a) response conflict on ultimately correct trials does not persist during articulation; (b) the timecourse of response conflict is restricted to the time window during which a given linguistic level is task relevant (early on for response appropriateness related variables and later for articulation relevant variables); and (c) monitoring during the response seems to be primarily triggered by pre-response monitoring failure.

Key words: error monitoring; response-conflict; language production; EEG, MVPA

2.1 Introduction

Speaking is a complex process that engages both cognitive and motor components, encompassing semantic and lexical retrieval as well as articulatory programming and execution. Extensive research has provided evidence that both cognitive and motor aspects of speech are continuously monitored to optimize fluent production. For instance, naturally occurring and laboratory induced speech errors show patterns suggesting the anticipation of potential undesired outcomes during speech planning. In particular, contextually inappropriate responses such as taboo words or non-lexical speech errors occur below chance even in controlled error protocols (Severens et al. 2011; Hartsuiker et al. 2005). Monitoring is also evident in speakers' behavior in response to their own speech errors, including accurate self-reporting of errors in various environments (Postma and Noordanus 1996; Gauvin et al. 2016); post-error increases in response latencies (Ganushchak and Schiller 2006); and self-repairs (Levelt 1983) . It has been observed that certain speech error repairs occur too rapidly to be attributed to the interception and planning of corrections after the error is produced, suggesting that errors are intercepted before becoming overt (Levelt 1983; Hartsuiker and Kolk 2001). Furthermore, studies involving modulated speech feedback have demonstrated error monitoring during articulation, as participants adapt their speech production (pitch or formants) to compensate for perceived distortions in feedback (e.g. Savariaux et al. 1995, Niziolek and Guenther 2013). Somatosensory feedback has also been altered to the same effect (e.g., Tremblay et al. 2003). Overall, speech error patterns reveal the coexistence of both cognitive and motor dimensions in speech monitoring. Concerning the temporal dynamics of monitoring, it has been shown that error-to-cutoff times display a bimodal distribution, with an interruption of an erroneous segment occurring either shortly after the error or around 500 ms later (Nooteboom and Quené 2017). This implies the existence of at least two distinct time points during which monitoring processes occur or interact with the production process. Employing time-sensitive electroencephalograpic recordings (EEG), prior literature has revealed three relevant time points to observe EEG activity in speech-monitoring tasks: speech planning in it's initial stages, speech motor planning, and speech articulation. In the context of *initial stage speech planning*, several studies have examined the EEG signal following stimulus presentation and preceding motor response preparation. In a speech production task designed to prime errors, trials resulting in errors showed an increased negativity between 350 ms and 600 ms after the appearance of a written word pair to be pronounced aloud (Moller et al. 2006). Additionally, semantic response conflict on correct trials in a phoneme detection task elicited a negativity around 450 ms after stimulus presentation (Ganushchak and Schiller 2008a). Concerning *speech motor planning*, previous EEG studies have analyzed the signal following the presentation

of a speech-cue, immediately before the response. In one study, within the 50-150 ms and 230-300 ms time windows erroneous trials exhibited more negative potentials compared to correct productions (Moller et al. 2006). In another study, correct trials primed to result in taboo word errors resulted in an increased negativity in the 550-625 ms time window compared to correct trials primed to result in neutral errors (e.g. Severens et al. 2011). Lastly, concerning *speech articulation*, previous studies analyzed the EEG signal following a response (button press or verbal). An error-related negativity (ERN) was observed following a button press in a phoneme detection task, where false alarms generated a larger ERN compared to correct hits (Ganushchak and Schiller 2008a). Furthermore, the magnitude of this ERN was modulated by the semantic relatedness of the auditory distractor, being greater for semantically related distractors. Additionally, during picture naming, incorrectly named pictures resulted in a larger ERN compared to correctly named pictures, with the ERN also being influenced by semantic naming context (Ganushchak and Schiller 2008b; Masaki et al. 2001; Riès et al. 2011; Baus et al. 2020).

Despite these valuable insights, a comprehensive understanding of the temporal dynamics of monitoring is hindered by the focus on specific variables and time frames of production in each study. The majority of previous studies have targeted monitoring through comparisons involving overt errors, lacking insight into how monitoring operates in contexts where errors are probable but ultimately avoided (but see Severens et al. 2011). As such, it remains unclear whether monitoring occurs at multiple time points for overt errors only or also for correct trials where errors are likely. Additionally, while some previous research has explored the impact of meaning-related variables (e.g., semantic relatedness, taboo status) on speech monitoring, the influence of other linguistic variables susceptible of producing response conflict, such as lexical or articulatory-phonetic variables, remains underexplored, and to our knowledge no study has explored more than one linguistic variable in the same study. Thus, the extent to which the time course of monitoring is similar for all aspects of speech or varies based on the level of representation remains an open question. Some hints to the answer to these open questions can be obtained through the results of two recent fMRI studies that examined correct utterances produced in contexts of high lexically or articualtory phonetic driven error probability respectively, and that also examined erroneous as opposed to correct utterances (Runnqvist et al. 2021 and Todorović et al. 2023). It was observed that cerebellar structures (Crus I) related to predictive internal modeling were active for both monitoring of correct but error prone utterances and for overt errors compared to correct utterances across lexical and articulatory phonetic variables. In contrast, additional frontal and medial frontal structures were recruited for error prone utterances at the articulatory phonetic level and for overt errors, indicating that distinct mechanisms are at play in function of distance from articulation. Finally, across both

studies, the anterior cingulate cortex was only differentially activated for the contrasts involving overt errors, supporting the involvment of a different feedback control related mechanism for overt errors. While the observed differential brain activity in these studies provides evidence supporting dissociations in monitoring mechanisms depending on whether errors are probable or overt and depending on linguistic variables (see also Runnqvist 2023; Teghipco et al. 2023; Okada et al. 2018; Hansen et al. 2019a; Hansen et al. 2019b; Volfart et al. 2022), it does not offer information about the specific timing of when these mechanisms are engaged during speech production. Doing so was the objective of the current study.

To this end, we conducted two similar resembling experiments, focusing on robust effects that replicated across both. Participants engaged in a speeded language production task designed to elicit speech errors of either a lexical or articulatory phonetic origin, while their EEG activity was recorded. The EEG signal was segmented into three distinct epochs (stimulus-locked, speech-cue-locked, and response-locked, see Figure 2.1) allowing us to cover the entire speech production process as reflected both by externally triggered events (e.g., stimulus and speech cue appearance, see Figure 2.1) and internally initiated events (e.g., the response). This design allowed us to explore monitoring processes related to correct but error prone production at both the lexical and articulatory - phonetic levels, as well as to explore monitoring related to overt speech production errors. Finally, we performed both event-related potential (ERP) analyses and multivariate pattern (MVPA) analyses on the data. The reason behind this analysis approach was that MVPA can be more sensitive to subtle variations in neural activity. It can detect distributed patterns of neural activation that ERPs might miss, making it especially useful when studying cognitive processes that involve complex and distributed neural networks (e.g., King and Dehaene 2014).

2.2 Methods and Materials

2.2.1 Participants

The study received appropriate ethical approval (filed under id “RCB: 2011-A00562-39") at the regional ethical committee “Comité de Protection des Personnes Sud Méditerranée I".

Experiment I

Twenty-nine right-handed native speakers of French (22 women) with normal or corrected-to-normal vision took part in the experiment in exchange for monetary compensation (mean age = 21, min = 19, max = 23). No participant reported any history of language or neurological disorders. One participant was excluded from the analyses

because they had participated in another study using the same task only a few weeks before the experimental session. Four participants were excluded from further analyses because of excessive noise or EEG data loss. Thus, 24 participants were included in the final analysis.

Experiment II

Fourty-four right-handed native speakers of French with normal or corrected-to-normal vision took part in the experiment in exchange for monetary compensation. No participant reported any history of language or neurological disorders. Thirteen participants were excluded from the analyses due to different issues: behavioral (5 participants had an error rate outside of the criteria: 1%<n<50%), problems during EEG recordings (8 participants, excessive noise or EEG data loss). Thus, 31 participants (29 women, mean age = 23, min = 19, max = 31) were included in the final analysis.

2.2.2 Stimuli

Experiment I

Target stimuli consisted of 160 printed French nouns (those used in Runnqvist et al. 2016) to be presented in pairs. For illustrative purposes, the examples in the text are given in English. To control for differences due to auditory stimulation, motor activity, or articulator specific modulations of the signal (e.g., Szirtes and Vaughan 1977), the same words were to be produced across conditions (albeit combined differently to prime **lexical** and **non-lexical errors**). Thus, across participants, each word was used twice in combination with another word (e.g., *mole sail*, *mole fence*). Exchanging the first letters of these combinations would result in a new word pair in one case (*sole mail*, lexical error outcome) and in a non-word pair in the other case (*fole mence*, non-lexical error outcome). An orthographic criterion was used for selecting stimuli, but even when applying a phonological criterion post-hoc only 7/160 non-lexical items resulted in real words for one of the words in a pair when changing orthography (which sometimes also entailed a change in wordclass , e.g., for the pair *caverne bouton* the primed noun error *couton* does not exist but *coûtons* is a conjugated form of the verb *couter*). All combinations were used in both possible orders across participants (e.g., *mole sail* and *sail mole*). Further, all combinations for which the exchange of initial phonemes resulted in new word-pairs (*mole sail*) were used also in reversed order (*sole mail*). The words in the target pairs were selected with the criterion that they should be unrelated. Despite this effort, for 9/320 word pairs (4 lexical and 5 non-lexical) there was some form of semantic relationship between the two words (e.g., *sick sinus. chicken tavern*). A given participant was only presented with one combination for each word (lexical

or non-lexical outcome), and was only presented with one of the words differing in only the first sound (*mole* or *sole*). This resulted in the creation of eight experimental lists with 80 word pairs (40 lexical and 40 non-lexical error outcome) counterbalanced across participants. Finally, all word pairs were coded for the degree of shared phonetic features (place, manner of articulation and voicing) of initial consonants of words in a pair being assigned a number from 0 (**AP distant words**, e.g., [m] is labial, nasal and voiced and [s] is dental, fricative and voiceless) to 2 (**AP close words**, e.g., [m] and [b] both being labial and voiced). Of the 80 word pairs, on average across the 8 lists 25.5 word pairs did not share any features, 40.25 word pairs shared one feature and 14.25 shared 2 features. As a first step, to control for a possible confound between our lexical and articulatory phonetic variables, we controlled whether the stimuli across the lexical and non-lexical conditions differed in the average amount of shared features (SF) and this was not the case (lexical 0.9 shared features vs. non-lexical 0.8 shared features).The two-tailed independent samples t-test comparing average SF values between the lexical and non-lexical conditions (n = 8 for both groups) yielded a non-significant result (p = 0.11). As a subsequent step, we introduced Articulatory-Phonetic Proximity (AP) as a three-level factor (No SF, 1 SF and 2 SF) during the analysis of errors and reaction times. This allowed us to investigate potential interactions between this factor and the lexical status, as detailed in the Analysis 2.2.5 and Results 2.3 sections. More importantly for the current purposes, coding the phonetic proximity between our word pairs also allowed us to assess the impact of this articulatory - phonetic variable known to modulate speech error rates (e.g., Nooteboom and Quené 2008; Oppenheim and Dell 2008) on participants' electrophysiological recordings.

During the experiment, three priming word pairs preceded each target word pair. The first two shared the initial consonants, and the third pair had further phonological overlap with the error being primed (*sun mall – sand mouth – soap mate – mole sail*). Note that, to induce errors, the order of the two initial consonants (/s/ and /m/) is different for the primes and the target. Participants were also presented with 153 filler pairs that had no specific relationship to their corresponding target pairs. One to three such filler pairs were presented to participants before each prime and target sequence. Thus, each participant was presented with 473 unique word combinations (80 targets, 240 primes and 153 fillers). Each list contained three blocks in which these 473 words were repeated three times in different orders. Participants were instructed to read all target word pairs aloud, all prime pairs silently, 35% of the filler pairs aloud and 65% of the filler pairs silently. Prior to the commencement of the experimental trials, participants underwent a task familiarization phase. This entailed exposure to a concise task sample, comprising 10 instances, under the direct supervision of the experimenter.

Experiment II

240 French monosyllabic (120) and bisyllabic (120) nouns were presented in pairs. Pairs were constant; there was no cross-combination of words as in Experiment I. Just as in Experiment I, exchanging the initial consonants of words in a pair gave a **lexical** or **non-lexical** outcome. We applied a phonological criterion for selecting the stimuli. In this stimuli set we manipulated the number of shared phonetic features of the onsets of words in pairs: half of the stimuli were **AP close** (2 shared feature among 3 possible: place, manner and voicing) and half were **AP distant** (no shared features), AP close and AP distant pairs were distributed equally across the lexical status conditions. Thus, here were no intermediate values (1 feature in common) as in Experiment I, because we aimed to maximize the effect by using the extremes. Words in pairs were always presented in the same order. Thus, the list was composed of 120 words pairs, where 60 were bisyllabic, 60 - monosyllabic, each syllabic condition contained 30 lexical and 30 non-lexical outcome pairs, and each lexical condition contained 15 AP close and 15 AP distant pairs. As in Experiment I, each target pair was preceded by 3 primes with the inverted order of onsets of words in pairs. Participants also saw 234 filler pairs, as in Experiment I, 1 to 3 of such pairs were presented before each sequence of primes and targets. The list was repeated 2 times with different order of sequences and primes. Participants were instructed to read all target word pairs aloud, all prime pairs silently, 49% of the filler pairs aloud and 51% of the filler pairs silently. Participants underwent pre-experiment task familiarization, involving a supervised exposure to a 10-instance task sample.

2.2.3 Procedure

Experiment I & Experiment II

The experiment was controlled by Eprime 2.0 software (Schneider et al. 2002). Each word pair remained on the screen for 700 ms and words presented for silent reading were followed by a blank screen for 200 ms. All targets and 35% of the filler items in the Experiment I and 50% of the filler items in Experiment II were followed by (a) a question mark remaining on the screen for 500 ms. (b) an exclamation mark presented 500 ms after the presentation of the question mark and remaining on the screen for 1000 ms, (c) a blank screen for 500 ms before the next trial started (see Schematic representaton of the task in Figure 2.1). Participants wore a microphone attached to the head in Experiment I, the microphone was placed on the table in front of them in Experiment II. They were instructed to silently read the word pairs as they appeared, but to name aloud the last word pair they had seen whenever a question mark was presented, and before the appearance of an exclamation mark. Productions were recorded both through

E-prime and the software Audacity®to be processed off-line.

2.2.4 Electrophysiological Recordings

Experiment I & Experiment II

The EEG was recorded from 64 Ag/AgCl Active-Two pre- amplified electrodes (BIOSEMI, Amsterdam; 10−20 system positions). The sampling rate was 1024 Hz for Experiment I (online filters: DC to 208 Hz, 3 db/octave) and 2048 for Experiment II. Two additional electrodes placed close to Cz, the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode, were used to form a feedback loop that maintains the average potential of the participant as close as possible to the AD-box reference potential. Two additional electrodes placed over the left and right mastoid were used to re-reference the signals (average mastoids reference). The vertical EOG was obtained by subtracting the signal of C29 (corresponding to FP2) from the signal of an external electrode placed underneath the left eye. The horizontal EOG was recorded with two external electrodes positioned over the two outer canthi.

2.2.5 Analyses

2.2.5.1 Behavior

Experiment I & Experiment II

Annotation. A person naive to the purpose of the experiment transcribed all spoken productions, then inspected and coded vocal response onsets of all individual recordings using Check-vocal (Protopapas 2007). Check-vocal is a software that allows for semi-automatic codification of the response accuracy and timing based on two sources of information: the speech waveform and the spectrogram. The transcriptions were scored as correct, disfluencies, partial responses (e.g., only one word produced), full omissions, and erroneous productions. The latter were classified as **priming related errors** or **other errors**. **Priming related errors** included full exchanges (*mill pad* ⇒ *pill mad*), partial exchanges (anticipations, e.g., *mill pad* ⇒ *pill pad*, perseverations, e.g., *mill pad* ⇒ *mill mad*, other partial exchanges, e.g., *mill pad* ⇒ *mill pack*). repaired and interrupted exchanges (*mill pad* ⇒ *pi..mill pad*), full and partial competing errors (*mill pad* ⇒ *pant milk/pant pad*), and other related errors (*mill pad* ⇒ *mad pill*), **Other errors**' included diverse phonological substitutions that were unrelated to the priming manipulation (e.g., *mill pad* ⇒ *chill pant/gri..mill pad/..pant*).

Experiment I

Data overview. Data of 24 participants initially presented 5760 trials, where each of the 80 pairs was repeated three times, resulting in 240 trials per participant. The lexicality condition was equally distributed in halves of the total number of trials, while the Articulatory-Phonetic (AP) condition introduced three levels (as detailed in Stimuli, Section 2.2.2). This configuration yielded 1834 pairs with no shared features, 2898 pairs with one shared feature, and 1028 pairs with two shared features between the initial consonants of the word pairs.

To ensure data quality, an initial filtering step excluded trials featuring full omissions, leading to the removal of 327 trials (5.68%). Subsequently, instances with RTs less than 100 ms or exceeding 1000 ms were identified as outliers and eliminated, accounting for 99 trials (1.72%). The final data set consisted of 5334 trials, distributed across conditions as follows: lexical (2663 trials), non-lexical (2671 trials), 0 shared features (1701 trials), 1 shared feature (2687 trials), and 2 shared features (946 trials).

Prior to statistical analysis, orthogonal contrasts were implemented for the AP condition using Helmert coding via the R built-in function `contr.helmert`. For the lexicality factor and priming_related_errors factor, sum coding (`contr.sum`) was applied (Chambers et al. 1990 through R documentation).

Experiment II

Data overview. Data of 31 participants initially presented 7440 trials, wherein each of the 120 pairs was repeated twice, resulting in 240 trials per participant. The lexicality and AP conditions were evenly distributed in halves of the total trial count (3720 per condition).

Following the same data filtering process as in Experiment I, the exclusion of full omissions led to a reduction in the number of trials to 7344 (1.29% excluded). The removal of RT outliers (those falling outside the range of 100 ms < RT < 1000 ms) accounted for 144 trials (1.44%). The resultant dataset comprised 7200 trials, distributed across conditions as follows: lexical (3579 trials), non-lexical (3621 trials), AP close (3588 trials), and AP distant (3621 trials).

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (Chambers et al. 1990 via R documentation) prior to conducting the statistical analysis in R Core Team 2022.

Experiment I & Experiment II

Statistical models. The data analysis was conducted using the RStudio R Core Team 2022 and key analytical tasks were performed utilizing specific packages including `'Matrix'` (version 1.5-1, Bates et al. 2021), `'lme4'` (version 1.1-34, Bates et al. 2015),

'lmerTest' (version 3.1-3, Kuznetsova et al. 2021), 'multcomp' (version 1.4-25, Hothorn et al. 2021), dplyr (version 1.1.2, Wickham et al. 2021), ggplot2 (version 3.4.3, Wickham 2016), each of which facilitated critical statistical and visualization procedures. For the analysis of errors, we employed a mixed-effects logistic regression model using the glmer function of lme4 package in R Core Team 2022. The initially proposed model:

```
glmer(errors ~ lexicality + AP + lexicality * AP +
      (1 +Lexicality| Subject) + (1 +AP| Subject) + (1 | WordPair),
            family = binomial)
```

involved two primary predictor variables, specifically Lexicality and AP, along with their interaction. Furthermore, the model accounted for a diverse set of random effects, encompassing random slopes and intercepts for both Lexicality and AP based on each individual subject in conjunction with a random intercept for the Word pair. However, due to convergence issues encountered during the modeling process, the complexity of the initial model needed to be streamlined. As a result, the model, shown below, aimed to scrutinize the effects of lexicality, AP condition, and their interaction on priming-related errors:

```
glmer(errors ~ lexicality + AP + lexicality * AP +
      (1 | Subject) + (1 | WordPair), family = binomial)
```

This model featured fixed effects for lexicality and AP condition as well as their interaction, while random intercepts for subjects and Word Pairs captured both individual differences and item-specific effects.

Additionally, two separate lmer functions were employed. The first aimed to uncover Reaction Time (RT) discrepancies between overt errors and correct productions:

```
lmer(RT ~ priming_related_errors +
     (1 + priming_related_errors | Subject) + (1 | WordPair))
```

This model included random intercepts for both subjects and items, along with a random slope for priming_related_errors within subjects, addressing subject-specific variations in how these errors influenced Reaction Time.

The second lmer model was exclusively applied to correct trials and focused on examining the interplay between Reaction Time (RT) and the variables of Lexicality and Articulatory-Phonetic (AP), along with their interaction.

```
lmer(RT ~ lexicality + AP + lexicality * AP +
     (1 | Subject) + (1 | WordPair))
```

This model integrated random intercepts for subjects and items to accommodate individual differences and item-specific effects.The final model is an outcome of simplifying

a more complex model that originally included random slopes and intercepts for both `Lexicality` and `AP` with respect to each individual subject. The decision to simplify was prompted by issues with convergence that were encountered during the modeling process.

2.2.5.2 EEG signal processing

Experiment I & Experiment II

Preprocessing. The EEG data was processed using the EEGLAB toolbox (Delorme and Makeig 2004) in MATLAB Inc. 2020. Continuous EEG data were filtered offline through a 0.1 Hz to 30 Hz band-pass filter. Activity from the left and right mastoid electrodes were used off-line to re-reference scalp recordings. For the analyses we defined three epochs of interest: ERPs were either (1) time locked to the stimulus and segmented into 800 ms epochs (-100 to 700 ms), (2) locked to the speech cue into 500 ms epochs (-100 to 400 ms) or (3) locked to the response into 1500 ms epochs (-1000 to 500 ms) (Figure 2.1). Only segments without artifacts (activity ± 75 μV) were included. The epochs were then averaged and referenced to a 100 ms pre-stimulus, pre-speech-cue and pre-response baseline, respectively.

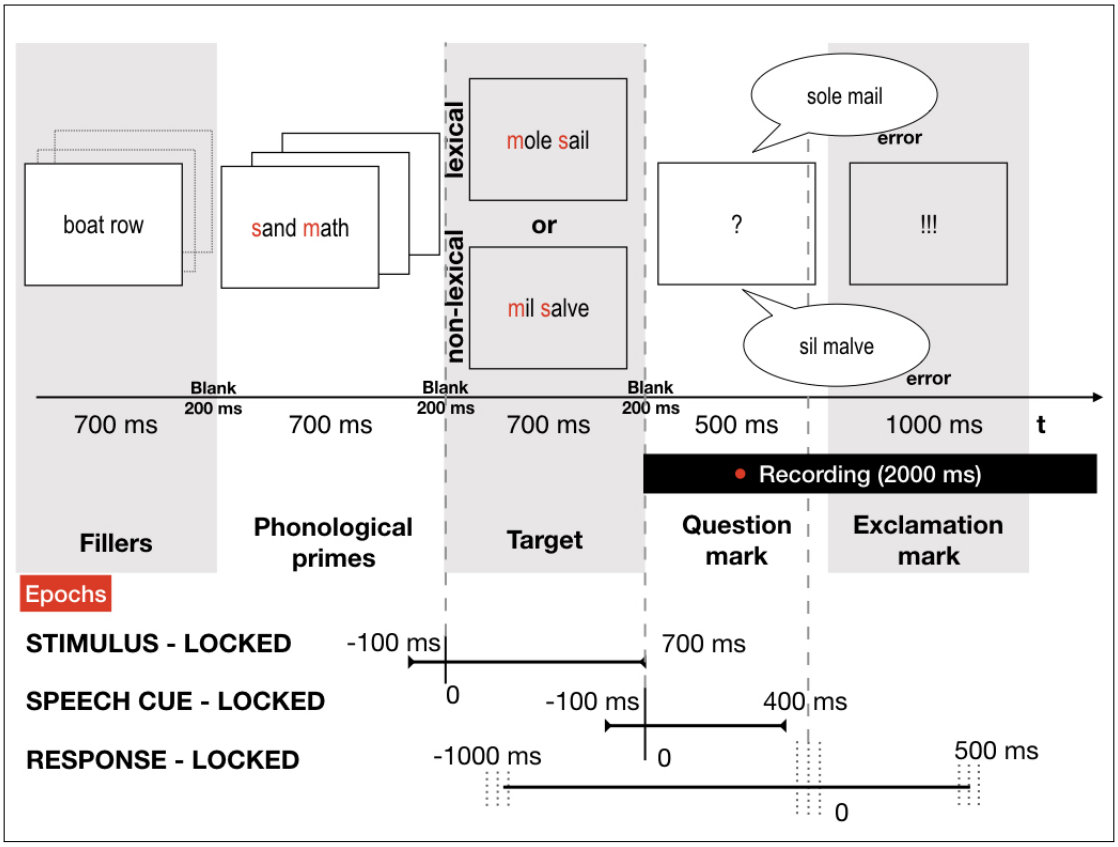


Figure 2.1 – Schematic representation of the procedure and epoching of EEG signal.

ERP analyses. As a next step we conducted a peak search within the epochs using ERP lab Lopez-Calderon and Luck 2014. The different conditions (overt (priming related) errors/correct; AP close/distant and lexical/non-lexical outcome) were averaged prior to this analysis, and we considered only the response peaks that were observed in both Experiments for subsequent analyses. For this, grand average waves of each epoch of both Experiments were inspected visually for the appearance of positive or negative peaks. Then the approximate time interval was given to the peak latency research function in ERP Measurement tool to obtain the exact peak latency value. These values were compared through two Experiments : when the difference in peak latency was less than 15 ms in between two Experiments, the mean value between the two peaks was used, when the difference was bigger, the peaks were not retained for further analyses. Subsequently, these peaks were used as centers of 100 ms time-windows. When it was impossible to use 100 ms time-windows, the largest possible symmetrical window was defined. The sum up of all the considered time-windows is reported in Table 2.1.

Statistical models. We utilized the same software and packages mentioned in the 2.2.5.1 Behavior "Statistical models" section to analyze EEG signal data. Each window of each Experiment was analyzed with Linear Mixed-Effects Models on 9 fronto-central electrodes of interest (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) same as in Grisoni et al. 2019 and on all electrodes (57 after excluding frontal electrodes F7, AF7, Fp1, Fpz, Fp2, and AF8). Separate regressions were applied to each of three conditions: (1) overt errors vs. correct:

```
lmer (Mean amplitude ~ priming_related_errors*Electrode +(1|subject)
```

(2) lexical vs. non-lexical primed error outcome on correct trials :

```
lmer (Mean amplitude ~ lexicality*Electrode +(1|subject)
```

and (3) Articulatory phonetic proximity (AP) (close (2 shared features) vs. distant (no shared features) on correct trials) -only the conditions of 0 and 2 shared features were contrasted in the analysis of AP proximity to ensure comparability across the two experiments:

```
lmer (Mean amplitude ~ AP*Electrode +(1|subject)
```

In summary, each time-window was analysed with three separate models to investigate the effects of errors, lexicality and AP proximity respectively (see below). Even though we conducted two separate experiments and focused on the cross-validated findings, we decided to apply a Bonferroni correction to handle the issue of multiple comparisons. Summarized p-values for both corrected and uncorrected regressions are available in the Supplementary materials 2.6, specifically in Table S.1 for 9 fronto-central electrodes of interest, and Table S.2 for all 57 electrodes.

MVPA. Multivariate Pattern Analysis (MVPA) was also conducted on both Experiments on the same time-windows as in the ERP analysis, with the sklearn software (Pedregosa et al. 2011). We fitted segmented data into a 2D space-time Riemannian manifold to then run a logistic regression to classify across trials each of our three binary variables: we performed binary classification of (1) overt errors vs. correct, (2) lexical vs. non-lexical status and (3) AP close vs. AP distant. We performed nested cross-validations (5-fold) to optimize the regularization strength while preventing overfitting, and to explore generalizability. Splitting of the data was performed using a stratified folding approach, to prevent models from biasing toward the most numerous class. The performance of the selected model was calculated with area under the receiver operating characteristic curve (ROC_AUC). The outer loop of the nested cross-validation was carried out 10 times and averaged per subject. This analysis was repeated for each of the three contrasts and on each time-window. Analyses were performed at the single-subject level and followed by standard parametric one-tail paired t-tests at the group level (distribution of ROC_AUC values across subject compared to chance level (50%)).

2.3 Results

2.3.1 Behavior

Experiment I

Errors. Participants made errors in 230 trials (3.99% of all data). After the filtering described in 2.2.5, this number was reduced to 222 (4.16% of filtered data) of which 124 errors were priming related (2.32% of filtered data). More priming related errors were made in the lexical condition (114, 2.14%) than in the non-lexical one (10, 0.18%). The dispatch of errors in articulatory − phonetic proximity groups was as follows : 26 errors out of 1701 trials without shared features AP (1.53%), 75 errors out of 2687 trials with 1 SF (2.79%), and 23 errors out of 946 trials with 2 SF (2.43%). The effect of lexicality was significant ($z = 6.31$; $p<.001$) on priming-related errors according to the Generalized Linear Mixed-Effects Model (see Analyses 2.2.5 for description). There was no effect of the articulatory-phonetic condition (1SF: $z = .02$; $p = .9$; 2SF: $z = .49$; $p = .62$) and no interaction with the lexicality condition (1SF: $z = 1.45$; $p = .14$; 2SF: $z = .18$; $p = .85$).

RT. Participants were slower in overall error trials (mean $RT_{222} = 612$ ms) that included priming related errors (mean $RT_{124} = 570$ ms)) than in correct trials (mean $RT_{5112} = 515$ ms). The effect of priming related errors on the RT was significant ($df = 20.59$; $t = -3.03$; $p<.01$). On correct trials, no significant difference in RT was observed between lexical (mean $RT_{2489} = 515$ ms and non-lexical condition (mean $RT_{2623} = 515$ ms), ($df = 598.8$; $t = .41$, $p = .68$). Participants were gradually slower through the AP condition

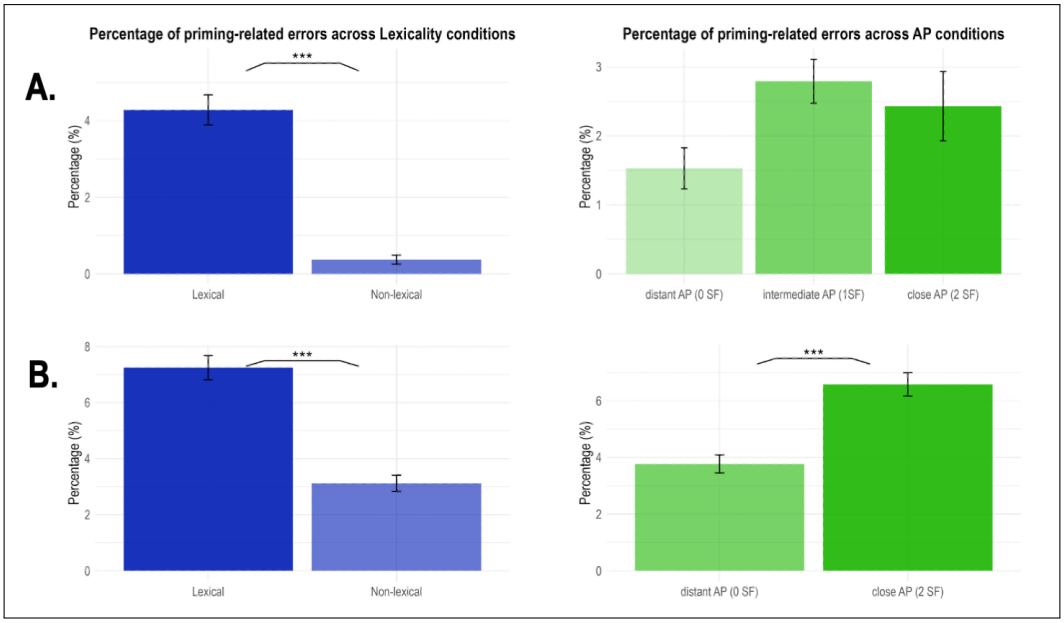


Figure 2.2 – Percentages of priming related errors by Lexicality and AP conditions. Panel A for Experiment I, panel B for Experiment II. Each portion is displayed with respect to the number of trials within the condition (i.e., the potential maximum number of errors).

ranging from mean RT_{1640} = 508 ms for 0 SF, RT_{2568} = 518 ms for 1 SF to RT_{904} = 522 ms for 2 SF. The AP effect was significant (1SF: df = 580.48; t = 2.73; p = .006; 2SF: df = 586.79; t = 2.3; p = .02), but no interaction with lexicality was observed (1SF: df = 587.19; t = −1.35; p = .17; 2SF: df = 606.73; t = .03; p = .97).

Experiment II

Errors. Participants made errors in (1100) trials (14.78% of all data), after the filtering described in 2.2.5, this number was reduced to 912 with 372 of priming related errors (5.16% of filtered data). More errors were committed in the lexical condition (259, 3.59%) compared to the non-lexical(113, 1.56%) and in the close articulatory-phonetic condition (236, 3.27%) compared to the distant articulatory-phonetic condition (136, 1.88%). The effects of lexicality and articulatory-phonetic condition were significant (z = 4.12; p<.001; z=-3.35; p<.001 respectively) on priming-related errors without interaction of the variables (z=1.19; p=.23).

RT. Participants were slower in overall error trials (mean RT_{912} = 628 ms) that included priming related errors (mean RT_{372} = 633 ms)) than in correct trials (mean RT_{6278} = 531 ms). The effect of priming related errors on the RT was significant (df = 31.99; t = −8.6; p<.001). On correct trials, no significant difference in RT was observed between lexical (mean RT_{3057} = 532 ms and non-lexical condition (mean RT_{3221} = 530 ms), (df = 113.53; t = .97; p = .33). Participants were slower in the close AP condition (mean RT_{3028} = 538 ms) than in the distant AP condition (RT_{3250} = 525 ms). The AP effect

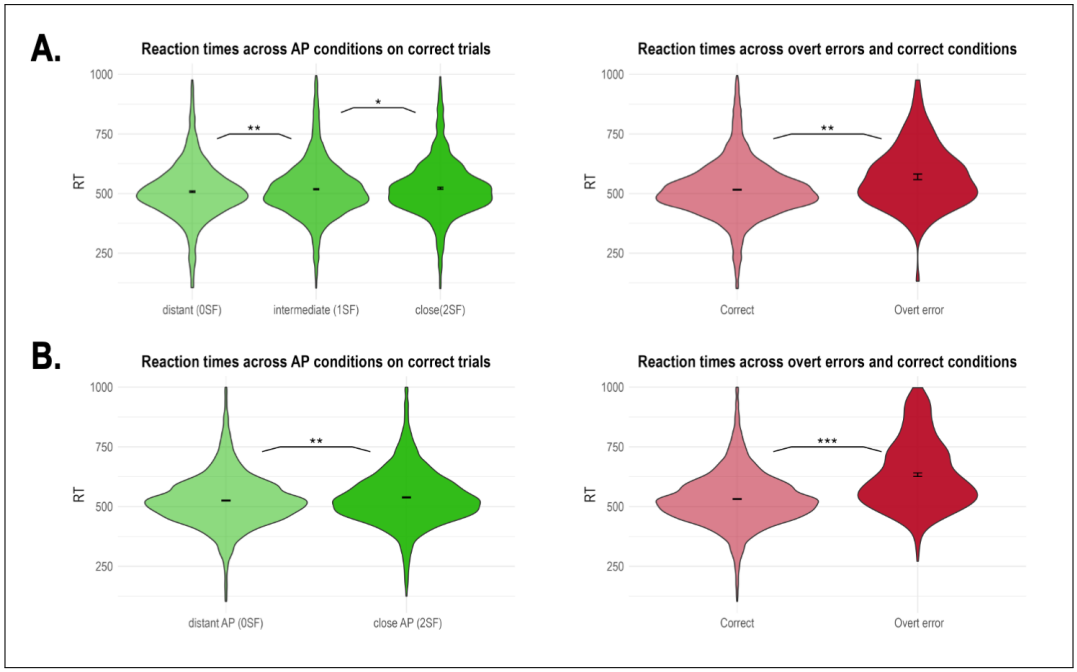


Figure 2.3 – Reaction Time (RT) Distribution by Articulatory phonetic (AP) Condition in Correct Trials (Left) and RT Distribution for Correct Responses and Overt Priming-Related Errors (Right). Panel A corresponds to Experiment I, while Panel B corresponds to Experiment II.

was significant ($df = 113.51; t = -3.1; p < .01$), but no interaction with lexicality was observed (1SF: $df = 113.52; t = -1; p = .28$).

2.3.2 EEG signal results

At the neural level, we investigated significant effects of lexicality (lexical vs. non-lexical error priming), phonetic articulatory (AP close vs. distant onsets) and error (overt errors vs. correct). We investigated them with both a univariate (ERP) and a multivariate (MVPA) method. Each analysis was performed on the two experiments, the three distinct types of epochs (stimulus-locked, speech-cue-locked, and response-locked) and either all electrodes or a ROI analysis including 9 fronto-central electrodes of interest (see Methods; Grisoni et al. 2019). Below we especially focus on the effects that were significant across both experiments, but the effects that did not replicate through Experiments are visualized and marked with asterisk in Figure 2.4 for response-locked epochs, in Figure 2.5 for speech cue-locked epochs.

ERP results. Three time windows in the response-locked epoch elicited significant differences for overt errors vs. correct condition in both datasets in the ROI analysis of 9 fronto-central electrodes. The first two time windows occurred before production onset. Firstly, the waveform of correct trials differed significantly from errors in the time window between [-483 to -383 ms] (Figure 2.6). This effect was followed by a significant difference between errors and correct trials during the pre-response positive drop [-115

ms, -15 ms] (Figure 2.4), ([-483 ms, -383 ms] and [-115 ms, -15 ms]). A third significant difference between correct trials and errors occurred after the onset of articulation ([62 ms, 162 ms]) (see Figure 2.6, Figure 2.4). For the stimulus-locked epochs, no significant effects were observed consistently across both experiments. In the 'all electrodes' analysis, the initial window of the speech cue-locked epoch [164 to 264 ms] exhibited a significant effect of overt errors, albeit not reaching significance after applying Bonferroni correction. The ERP analysis did not reveal any significant effect -cross-validated across Experiments- of the Articulatory-Phonetic (AP) and Lexicality effects. Supplementary tables of p-values can be found in the Supplementary Materials 2.6.

MVPA results. Both datasets showed significant decoding for the first time window (164 − 264 ms) of the speech cue-locked epoch for AP close vs. distant (Experiment 1: t =3.10; [ROC_AUC] =0.53; p<0.01; Experiment 2: t =4.70; [ROC_AUC] =0.57; p<0.01) and overt errors vs. correct (Experiment 1: t =2.42; [ROC_AUC] =0.52; p=0.01; Experiment 2: t =3.29; [ROC_AUC] =0.56; p<0.01). Furthermore, both datasets showed significant decoding for the response-locked epochs: the first window (-793, -693 ms) revealed a significant lexicality effect (Experiment 1: t =2.85; [ROC_AUC] =0.53; p<0.01; Experiment 2: t =5.7; [ROC_AUC] =0.61; p<0.01) and the second window (-483; -383 ms) revealed a significant overt errors vs. correct effect (Experiment 1: t =3.52; [ROC_AUC] =0.53; p<0.01; Experiment 2: t = 4.35; [ROC_AUC] =0.59; p<0.01). No significant effects consistent across both experiments were observed for the stimulus-locked epochs. Additional significant decoding results, not cross-validated between the two experiments, are depicted in Figures 2.4, 2.5, and also in Table S.3 and S.2 in the Supplementary Materials 2.6.

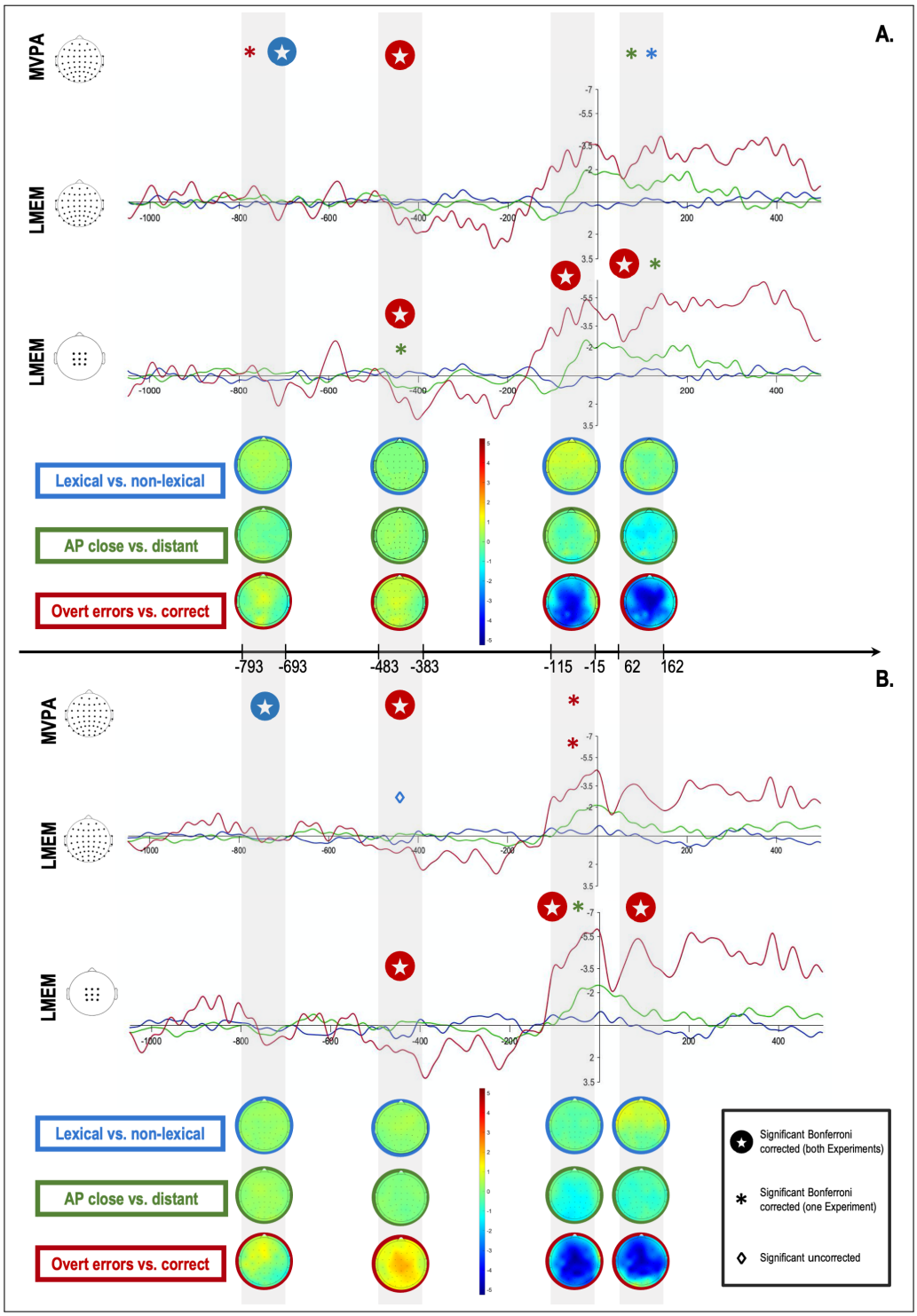


Figure 2.4 – Summary of results of all performed analyses: (top) MVPA, (middle) mean difference waves of all electrodes and (bottom) of the 9 fronto-central electrodes and their topographic maps across Experiment I (panel A.) and Experiment II (panel B.) in response-locked epochs. Color code is used to differentiate the conditions: blue for lexicality, green for articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant bonferroni corrected p-values if observed in one Experiment, empty diamonds mark significant p-values without correction while bold stars mark significant p-values if observed in both Experiments.

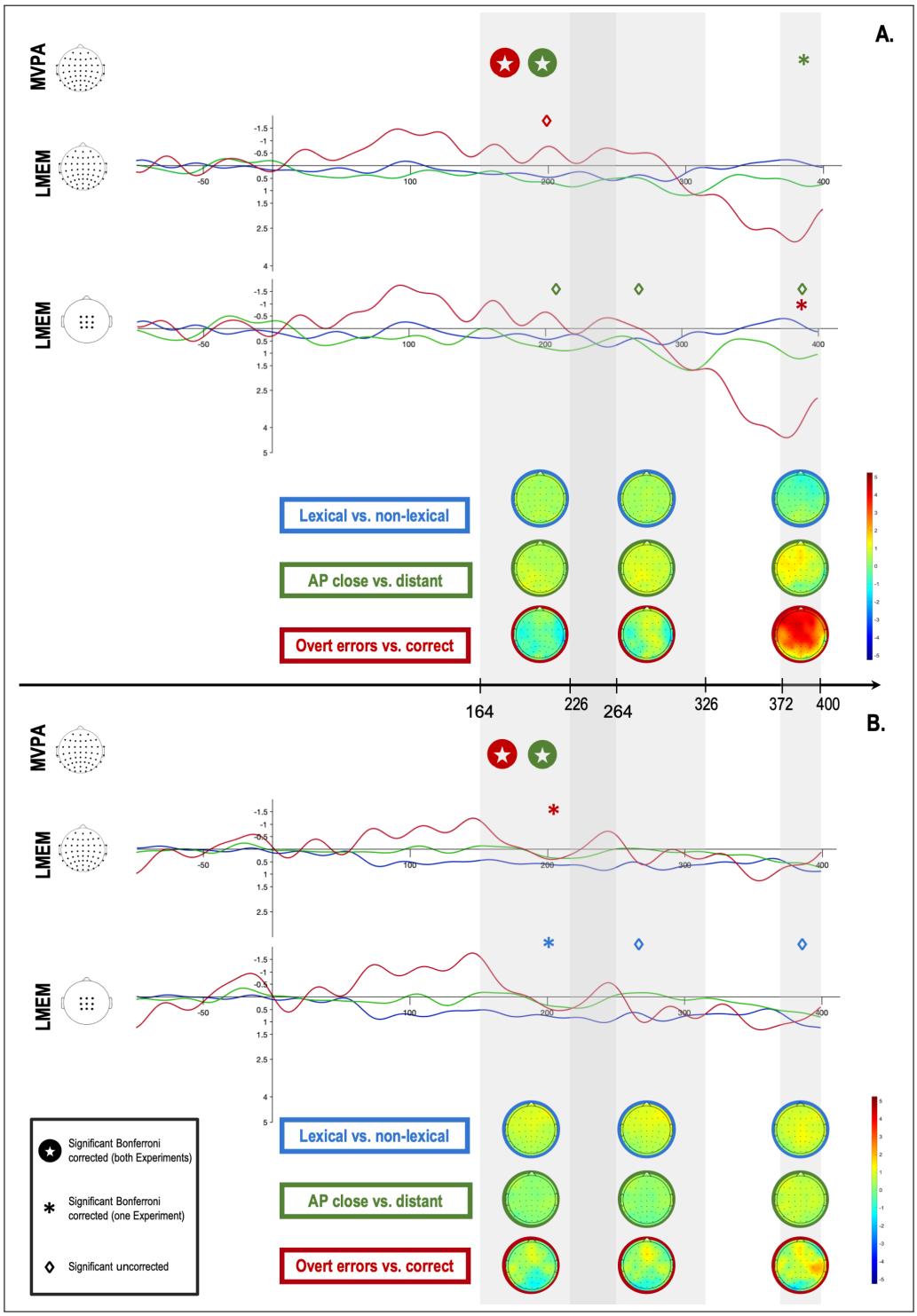


Figure 2.5 – Summary of results of all performed analyses: (top) MVPA, (middle) mean difference waves of all electrodes and (bottom) of the 9 fronto-central electrodes and their topographic maps across Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Color code is used to differentiate the conditions: blue for lexicality, green for phonetic articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant bonferroni corrected p-values if observed in one Experiment, empty diamonds mark significant p-values without correction while bold stars mark significant p-values if observed in both Experiments.

		latency (ms)		difference (ms)	window		
		Exp.1	Exp.2		mean	start	end
Stimulus - locked	negative	93	97	-4	95	45	145
	positive	204	213	-9	209	159	259
	negative	249	270	-21			
	negative	434	445	-11	439	389	489
Speech-cue - locked	negative	28	59	-31			
	positive	212	216	-4	214	164	264
	negative	269	272	-3	270	220	320
	positive	386	386	0	386	336	436
Response - locked	negative	-737	-740	3	-739	-789	-689
	positive	-667	-628	-39			
	negative	-437	-430	-7	-433	-483	-383
	negative	-56	-68	12	-62	-112	-12
	negative		51				
	positive		69				
	negative		119				
	positive		201				

Table 2.1 – Summary table of common peaks found in 2 Experiments

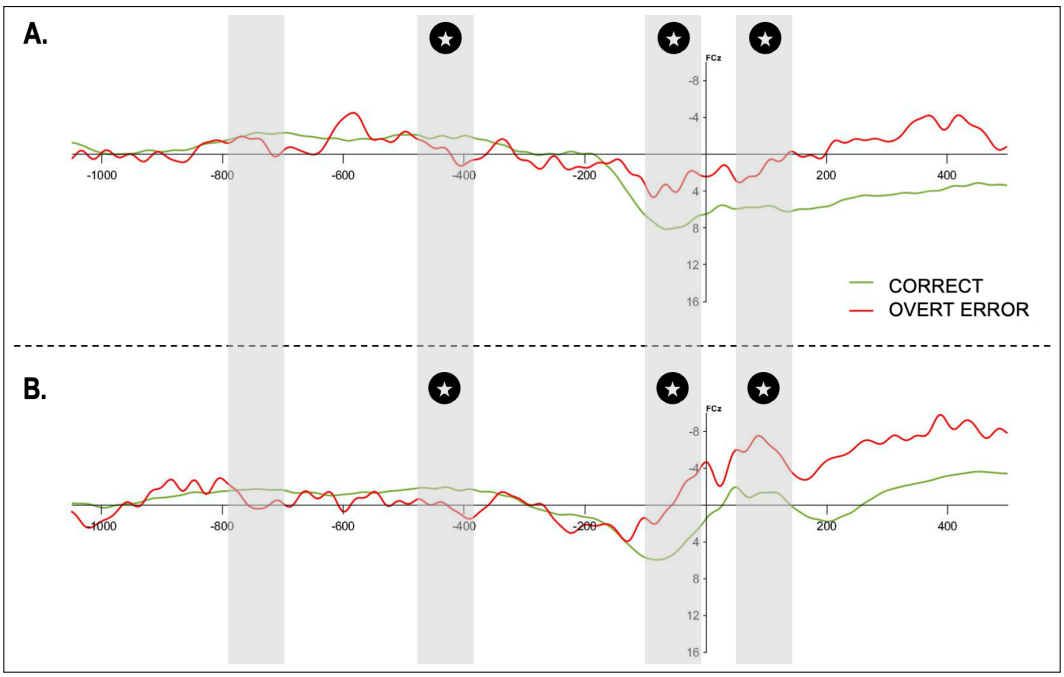


Figure 2.6 – Grand average wave of FCz electrode for correct (green) and overt error (red) trials in Experiment I (panel A.) and Experiment II (panel B.) in response-locked epochs. Stars indicate significant effect in GLM on 9 fronto-central electrodes (FC1 FCz FC2 C1 Cz C2 CP1 CPz CP2).

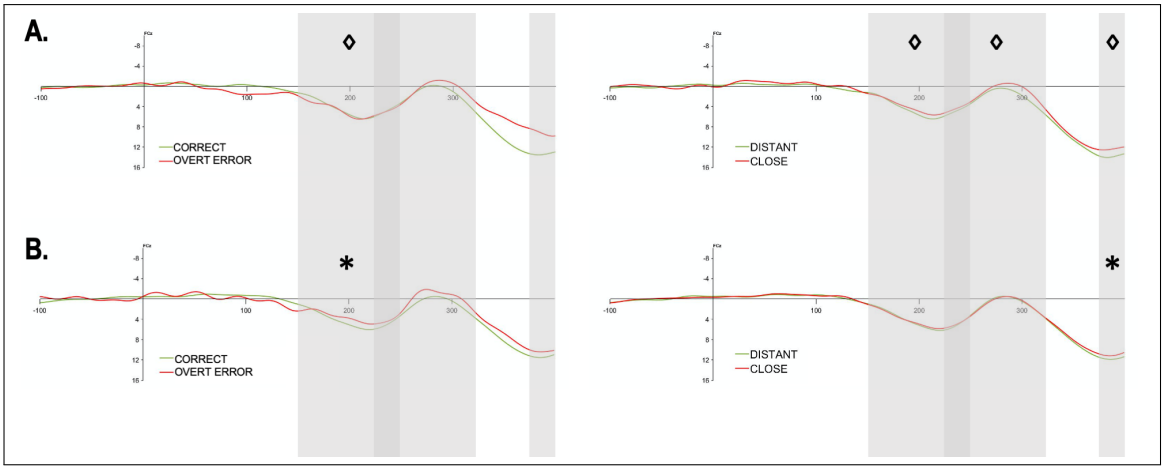


Figure 2.7 – Left : Grand average wave of FCz electrode for correct (green) and overt error (red) trials in Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Right : Grand average wave of FCz electrode for AP close (red) and AP distant (green) correct trials in Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Asterisks and diamonds indicate significant effect in one or the other GLM (9 fronto-central electrodes or all electrode) of the ERP.

2.4 Discussion

The principal aim of this study was to thoroughly investigate the temporal dynamics of the monitoring process, encompassing speech planning, speech-motor preparation, and articulation. Our specific focus was to explore potential variations in the temporal dynamics of monitoring of (a) correctly produced utterances with high error probability inflicted by either lexical or articulatory-phonetic related variables; and (b) incorrect utterances. To accomplish this, we conducted two EEG experiments. To have a comprehensive temporal map of the entire speech production process, the three contrasts of interest (high vs low lexically driven error probability, high vs low articulatory phonetically driven error probability, and errors vs correct trials) were examined on three distinct epochs that allowed us to asses speech planning, speech motor preparation and articulation. Finally, we employed two types of analyses (ERP and MVPA). It is worth noting that our emphasis was on robust effects that consistently emerged across both experiments, ensuring the reliability and validation of the results. In what follows we will discuss the results we obtained for initial stages of speech planning, speech motor planning and articulation respectively.

Initial stages of speech planning

The initial stages of speech planning were examined through the stimulus-locked epochs and through the early part of the response-locked epochs (see Figure 2.1). There was an effect of lexicality on the early pre-response part of the response-locked time window [-793 -693]. That is, based on the electrophysiological reponse across all electrodes, the MVPA distinguished above chance those correct trials that were more error prone due to the lexical response competition from those that were less error

prone due to an absence of lexical competition. Interestingly, this effect falls almost in the same time-window where previous studies had observed ERP effects of semantic response conflict (Ganushchak and Schiller 2008a) and of response conflict (Moller et al. 2006).Taken together, these findings seem to indicate that the kind of response competition affecting early stages of speech planning is related to the meaning or appropriateness of a potential response (i.e., competing responses that are semantically related as opposed to unrelated, or that are real words as opposed to non-words are all more appropriate).Additionally, this time window resulted in an overt errors vs. correct effect in Experiment I, similar to Moller et al. 2006, but the absence of this effect in Experiment II, despite a larger number of observations, indicates its lesser robustness. One possibility is that the effect was more robust in their study because all their critical trials were primed to result in lexical errors and were thus always response appropriate. To gain further insights on the nature of the effect of lexical error probability that we observed, we will briefly consider the results of the fMRI study by Runnqvist et al. 2021, using the exact same task as here and the same stimuli as in our Experiment 1. In that study, lexically driven error probability engaged the Crus I of the cerebellum, which was linked to internal modeling of upcoming speech as a means of error monitoring. Interestingly, and consistent with this interpretation of predictive internal modeling, another study using EEG found that the readiness-potential (RP), was modulated by predictability (Grisoni et al. 2019). Although we did not observe a distinguishable RP in our data, the timing of our MVPA lexicality effect preceding the verbal response is consistent with this component that is usually observed preceding a motor response,. Thus, linking our findings with a previously found cerebellar origin of the effect and with modulations related to predictability occurring in similar time windows in previous studies, a plausible interpretation is that monitoring during the initial stages of speech planning is carried out through predictive internal modeling (e.g., Pickering and Garrod 2013, Runnqvist et al. 2016, Runnqvist et al. 2021).

Speech motor preparation

The speech motor preparation period was investigated through the speech-cue locked epochs and the late pre-response part of response-locked epochs. Leveraging multivariate pattern analysis (MVPA), we observed significant decoding rates for both overt errors as compared to correct trials and for high as compared to low articulatory-phonetic error probability on correct trials within a [164 264 ms] time-window after the speech cue. This time window is similar to the one where previously more negative event-related potential (ERP) for errors were reported (Moller et al. 2006). Our MVPA findings seem to mirror the EEG findings of Moller et al. 2006 in what concerns overt errors, but also extend their findings by showing that articulatory phonetic conflict also impacts the same time window. In that study, the spatial source of the EEG effect was localized to the medial frontal cortex (SMA), with a potential involvement of the

anterior cingulate region. Additionally, the SMA activation during speech planning (post - speech cue) was previously linked by Moller et al. 2006 to conflicts in articulatory gestures planning, which is in line with recent findings by Todorović et al. 2023 and the appearance of the pre-SMA region activation in situations involving increased articulatory-motor complexity (e.g., Alario et al. 2006). In summary, our findings are consistent with the SMAs known implication in phonetic encoding and articulation complexity, and provides support for the hypothesis that also response conflict involving articulatory phonetic representations may lead to greater SMA activation during speech motor planning. Connecting both sets of findings (fMRI and EEG), the SMA might work in concert with frontal and parietal structures and the superior cerebellum in a forward modeling loop preparing for motor execution (e.g., Riecker et al. 2005, Todorović et al. 2023). Regarding the late pre-response window of the response locked epochs, we observed a significant difference between errors ans correct trials in the ERPs [-115 to -15 ms]. We propose two plausible interpretations for this effect: The first interpretation is consistent with the findings discussed in the speech-planning section regarding the readiness potential (RP) from Grisoni et al. 2017. In fact, this window corresponds to the greatest negativity of the RP, preceding the positive drop. The effect on this window supports the idea of prediction mechanisms as a monitoring component and implies that errors are inherently more unpredictable than correct trials. Alternatively, the effect may also be attributed to proprioceptive error detection involving somatosensory speech targets and stemming from the preparation of articulatory muscles before the onset of speaking (Riès et al. 2020; Guenther and Hickok 2016).

Articulation

The articulation-related effects were examined by analyzing the post-response part of the response-locked epochs. The contrast of overt errors vs. correct trials yielded a significant effect in the time window [62-162 ms] that is consistent with the error-related negativity (ERN) (Ganushchak and Schiller 2008a). We thus interpret the effect in the [62-162 ms] time window as the ERN, which is typically associated with conflict monitoring in the Anterior Cingulate Cortex (ACC) (Dehaene et al. 1994, Falkenstein et al. 1991,Nozari et al. 2011). However, the ACC was found to be active only for overt errors (and not for internal monitoring) in previous studies (Runnqvist et al. 2021; Todorović et al. 2023), supporting the idea of the ACC having a feedback-related function especially for the time-window when the error is articulated. Surprisingly, multivariate pattern analysis (MVPA) did not show significant decoding rates for overt errors vs. correct trials, while the event-related potential (ERP) effects were robustly significant even after applying Bonferroni correction. This discrepancy raises questions about the underlying factors reflected by these two types of analyses and will be discussed further.

To summarize the global picture from both experiments: Internal monitoring of

correct trials reveals early differences, discernible only during speech planning, suggesting that conflict on correct trials related to response appropriateness or meaning is resolved or substantially diminished by the time motor planning and articulation take place. On the other hand, articulatory-phonetic conflict on correct trials triggers differences only during speech motor preparation, but this conflict is also resolved or greatly diminished before the actual response. In contrast, overt fluent errors trigger differences during both early and late speech motor preparation, as well as during articulation. Thus, when examining correct trials, there seems to be a temporal coincidence between the moment of conflict emergence and the moment when the object of this conflict becomes task-relevant (i.e., response appropriateness or meaning related conflict arises when participants read and process meaning, and articulatory-phonetic conflict arises during speech motor planning). Note that while these results indeed suggest that processes operating on the already activated linguistic representations to be produced unfold in a sequential manner according to their task relevance, they may be compatible with both sequential and parallel processing in language production (e.g., Fairs et al. 2021). For instance, Fairs et al. 2021 proposed that while all dimensions of words are subject to a first pass of parallel activation (ignition) due to their holistic nature, selection and checking processes are likely to proceed sequentially during later reverberation processes. Concerning the mechanism underlying these internal monitoring effects, the combined evidence of this and previous studies suggest that this monitoring is carried out through predictive internal modeling. If correct, such internal modeling seems to generate error signals in a temporally distributed and task-relevant fashion (as opposed to only, for instance, upon phonological encoding). Conversely, overt errors show both pre-response and response differences in processing, suggesting that when the error signal of the predictive internal modeling is not enough to stop an error, additional, presumably more feedback dependent, processes are triggered during articulation. That is, the persistence of the effect related to overt errors suggests that errors may be detected multiple times, possibly through different processes. Concretely, error detection may occur through internal modeling before the response, followed by proprioceptive feedback, and finally, feedback-related mismatch. This interpretation aligns with fMRI studies that have identified the cerebellum, housing the internal models, as playing a significant role for monitoring of both error probability and overt errors (Runnqvist et al. 2021; Todorović et al. 2023), while other structures seemed to be exclusively triggered by overt errors (e.g., the ACC).

In a more comprehensive discussion of the present study, two intriguing questions remain unanswered. The first question pertains to determining the most pertinent event during speech production for precisely accounting for the underlying monitoring processes. Specifically, we consider two types of segmentation: one based on external events, such as stimulus presentation and speech cue appearance, and the other based

on the participant's initiated response. While the response-locked epochs provided dynamic windows across participants, the stimulus-locked epochs remained stable. To effectively analyze groups of individuals, alignment in processes is essential to observe and statistically quantify the effects. Remarkably, the response-locked segmentation yielded a greater number of significant effects in both experiments and revealed the lexical effect during the speech planning stage, which coincided with the window of the stimulus-locked segmentation where this effect was not observed. This finding suggests that the initiation of the response may serve as the departure point that allows for alignment among individuals, as external events may introduce perceptual differences at multiple levels (e.g., reading speed, encoding period, lexical retrieval speed), potentially disaligning the group from a process-oriented perspective.

The second question emerges from the diverging results obtained from the two types of analyses conducted on the data. For example, while both the ERP analysis and MVPA yielded consistent cross-validated results in the early pre-response window [-483 to -383 ms], the strong error-related negativity (ERN) effect observed in the post-response window with the ERP analysis was not observed with MVPA. This discrepancy raises the question of what each type of analysis reflects and how to interpret the differences between them.

Finally, one potential limitation of our study should be pointed out. While our approach of cross validation at multiple levels (time window selection, consideration of significnt effects) ensures that the effects observed are robust and generalizable, it is possible that this rather stringent approach made certain true but more subtle effects go undetected. As we focused on the discussion of the effects that we did observe here it does not compromise our conclusions, and can be addressed in future studies by, for instance, including the time windows reported as significant in one of our two experiments in more focused an hypothesis driven analyses.

Conclusion

Cross-validated results from two experiments revealed robust electrophysiological effects of high versus low lexically and articulatory phonetically driven error probability on correct trials, and of overt speech errors versus correct trials. Temporal dissociations were observed across these contrasts with lexicality affecting the early stages of speech planning, articulatory phonetic proximity the early stages of speech motor preparation, and overt errors affecting both early and late stages of speech motor preparation as well as articulation. These results suggest the presence of temporally distributed predictive internal modeling in charge of monitoring before articulation, and of an additional mechanism relying on somatosensory and auditory feedback recruited successively in the case of the occurrence of a speech error.

Funding

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). E.R. has benefited from support from the French government, managed by the French National Agency for Research (ANR) through a research grant (ANR-18-CE28-0013).

2.5 References

Alario, F., Chainay, H., Lehéricy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production [Publisher: Elsevier]. Brain Research, 1076(1), 129–143. https://doi.org/10.1016/j.brainres.2005.11.104

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2021). Matrix: sparse and dense matrix classes and methods [R package version 1.3-4]. https://CRAN.R-project.org/package=Matrix

Baus, C., Santesteban, M., Runnqvist, E., Strijkers, K., & Costa, A. (2020). Characterizing lexicalization and self-monitoring processes in bilingual speech production. Journal of Neurolinguistics, 56, 100934. https://doi.org/10.1016/j.jneuroling.2020.100934

Chambers, J., Hastie, T., & Pregibon, D. (1990). Statistical Models in S. In K. Momirović & V. Mildner (Eds.), Compstat (pp. 317–321). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-50096-1_48

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a Neural System for Error Detection and Compensation [_eprint: https://doi.org/10.1111/j.1467-9280.1994.tb00630.x]. Psychological Science, 5(5), 303–305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

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), tgab040. https://doi.org/10.1093/texcom/tgab040

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalography and Clinical Neurophysiology, 78(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Ganushchak, L. Y., & Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring: An ERP study. Brain Research, 1125(1), 104–115. https://doi.org/10.1016/j.brainres.2006.09.096

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

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

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

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. The Journal of Neuroscience, 37(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Mohr, B., & Pulvermüller, F. (2019). Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. NeuroImage, 199, 206–216. https://doi.org/10.1016/j.neuroimage.2019.05.071

Guenther, F. H., & Hickok, G. (2016). Neural Models of Motor Speech Control. In Neurobiology of Language (pp. 725–740). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00058-4

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

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

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). Journal of Memory and Language, 52(1), 58–70. https://doi.org/10.1016/j.jml.2004.07.006

Hartsuiker, R. J., & Kolk, H. H. J. (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

Hothorn, T., Bretz, F., & Westfall, P. (2021). multcomp: simultaneous inference in general parametric models [R package version 1.4-25]. https://CRAN.R-project.org/package=multcomp

Inc., T. M. (2020). Matlab version: 9.8.0 (r2020a). Natick, Massachusetts, United States. https://www.mathworks.com

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends in Cognitive Sciences, 18(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2021). lmerTest: tests in linear mixed effects models [R package version 3.1-4]. https://CRAN.R-project.org/package=lmerTest

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. Cognition, 14(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. Frontiers in Human Neuroscience, 8. Retrieved May 8, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00213

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. Neuroreport, 12(9), 1851–1855. https://doi.org/10.1097/00001756-200107030-00018

Moller, J., Jansma, B. M., Rodriguez-Fornells, A., & Munte, T. F. (2006). What the Brain Does before the Tongue Slips. Cerebral Cortex, 17(5), 1173–1178. https://doi.org/10.1093/cercor/bhl028

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. The Journal of Neuroscience, 33(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Nooteboom, S. G., & Quené, H. (2008). Self-monitoring and feedback : A new attempt to find the main cause of lexical bias in phonological speech errors [Accepted: 2017-06-09T19:35:25Z]. Journal of Memory and Language, 58, 837–861. Retrieved September 8, 2021, from http://localhost/handle/1874/350522

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

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

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

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

Postma, A., & Noordanus, C. (1996). Production and Detection of Speech Errors in Silent, Mouthed, Noise-Masked, and Normal Auditory Feedback Speech [Publisher: SAGE Publications Ltd]. Language and Speech, 39(4), 375–392. https://doi.org/10.1177/002383099603900403

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behavior Research Methods, 39(4), 859–862. https://doi.org/10.3758/BF03192979

R Core Team. (2022). R: a language and environment for statistical computing [Version 4.2.2, released on 2022-10-31]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. Neurology, 64(4), 700–706. https://doi.org/10.1212/01.WNL.0000152156.90779.89

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

Riès, S., Nadalet, L., Mickelsen, S., Mott, M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2020). Pre-output Language Monitoring in Sign Production. Journal of Cognitive Neuroscience, 32(6), 1079–1091. https://doi.org/10.1162/jocn_a_01542

Runnqvist, E. (2023). Self-Monitoring: The Neurocognitive Basis of Error Monitoring in Language Production [Num Pages: 23]. In Language Production. Routledge.

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

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), tgab038. https://doi.org/10.1093/texcom/tgab038

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. The Journal of the Acoustical Society of America, 98(5), 2428–2442. https://doi.org/10.1121/1.413277

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime User's Guide.

Severens, E., Janssens, I., Kühn, S., Brass, M., & Hartsuiker, R. J. (2011). When the brain tames the tongue: Covert editing of inappropriate language [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2011.01190.x]. Psychophysiology, 48(9), 1252–1257. https://doi.org/10.1111/j.1469-8986.2011.01190.x

Szirtes, J., & Vaughan, H. (1977). Characteristics of cranial and facial potentials associated with speech production. Electroencephalography and Clinical Neurophysiology, 43(3), 386–396. https://doi.org/10.1016/0013-4694(77)90261-9

Teghipco, A., Okada, K., Murphy, E., & Hickok, G. (2023). Predictive Coding and Internal Error Correction in Speech Production. Neurobiology of Language (Cambridge, Mass.) 4(1), 81–119. https://doi.org/10.1162/nol_a_00088

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S., & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. Neurobiology of Language, 1–47. https://doi.org/10.1162/nol_a_00113

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. https://doi.org/10.1038/nature01710

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

Wickham, H. (2016). ggplot2: elegant graphics for data analysis [R package version 3.4.3]. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: a grammar of data manipulation [R package version 1.0.7]. https://CRAN.R-project.org/package=dplyr

2.6 Supplementary Materials

			Overt error		Lexicality		AP proximity	
		Time-window	p-values					
				(uncorrected)		(uncorrected)		(uncorrected)
Exp.1	Stimulus	62 162	0.132	0.066	0.904	0.452	0.460	0.230
		142 242	0.212	0.106	0.643	0.322	0.505	0.253
	Speech cue	164 264	2.141	0.714	0.353	0.118	0.141	**0.047**
		226 326	1.989	0.663	0.754	0.251	0.120	**0.040**
		372 400	**0.010**	**0.003**	2.545	0.848	0.098	**0.033**
	Response	-793 -693	0.806	0.269	0.483	0.161	1.739	0.580
		-483 -383	**0.001**	**<0.001**	1.083	0.361	**0.044**	**0.015**
		-115 - 15	**<0.001**	**<0.001**	0.783	0.261	0.349	0.116
		62 162	**0.003**	**0.001**	2.099	0.700	**0.036**	**0.012**
Exp.2	Stimulus	62 162	0.490	0.245	0.890	0.445	0.634	0.317
		142 242	1.560	0.780	0.578	0.289	1.745	0.872
	Speech cue	164 264	1.884	0.628	**0.048**	**0.016**	1.919	0.640
		226 326	1.475	0.492	0.132	**0.044**	2.850	0.950
		372 400	1.566	0.522	0.139	**0.046**	0.764	0.255
	Response	-793 -693	0.669	0.223	1.560	0.520	0.324	0.108
		-483 -383	**0.007**	**0.002**	0.215	0.072	1.856	0.619
		-115 - 15	**<0.001**	**<0.001**	0.855	0.285	**0.015**	**0.005**
		62 162	**0.002**	**0.001**	1.646	0.549	0.574	0.191

Table S.1 – Summary of p-values of 9 fronto-central electrode GLM. Significant values in one experiment are in bold, significant in both experiments are on grey background

			Overt error		Lexicality		AP proximity	
		Time-window	p-values					
				(uncorrected)		(uncorrected)		(uncorrected)
Exp.1	Stimulus	62 162	1.100	0.367	1.716	0.572	0.853	0.284
		142 242	0.243	0.081	2.480	0.827	2.855	0.952
	Speech cue	164 264	0.097	**0.032**	0.905	0.302	0.729	0.243
		226 326	0.161	0.054	2.834	0.945	1.670	0.557
		372 400	0.129	0.064	1.842	0.921	1.944	0.972
	Response	-793 -693	0.124	0.062	1.876	0.938	1.689	0.844
		-483 -383	2.734	0.911	1.035	0.345	1.637	0.546
		-115 - 15	2.351	0.784	1.351	0.450	1.537	0.512
		62 162	**0.014**	**0.005**	2.175	0.725	1.272	0.424
Exp.2	Stimulus	62 162	0.500	0.167	1.414	0.471	0.853	0.284
		142 242	0.283	0.094	0.780	0.260	2.855	0.952
	Speech cue	164 264	**0.028**	**0.009**	1.950	0.650	0.729	0.243
		226 326	0.070	**0.023**	1.391	0.464	1.670	0.557
		372 400	0.315	0.157	1.941	0.970	1.944	0.972
	Response	-793 -693	1.085	0.542	1.939	0.969	1.689	0.844
		-483 -383	2.265	0.755	0.148	**0.049**	1.637	0.546
		-115 - 15	1.962	0.654	0.217	0.072	1.537	0.512
		62 162	1.400	0.467	1.203	0.401	1.272	0.424

Table S.2 – Summary of p-values of all electrode GLM. Significant values in one experiment are in bold, significant in both experiments are on grey background

			Overt error	Lexicality	AP proximity
		Time-window	p-values		
Exp.1	Stimulus	62 162	0,350	0,431	0,178
		142 242	**0,027**	0,271	0,809
	Speech cue	164 264	**0,002**	0,095	**<0.001**
		226 326	0,750	0,374	0,359
		372 400	0,099	0,073	**0,005**
	Response	-793 -693	**0,001**	**<0.001**	0,072
		-483 -383	**<0.001**	0,345	0,992
		-115 - 15	0,101	0,130	0,089
		62 162	0,081	**0,003**	**0,007**
Exp.2	Stimulus	62 162	0,085	0,692	0,111
		142 242	0,644	**0,028**	0,494
	Speech cue	164 264	**0,011**	0,060	**0,002**
		226 326	0,697	0,278	0,520
		372 400	0,578	0,214	0,629
	Response	-793 -693	0,097	**0,004**	0,055
		-483 -383	**0,001**	0,863	0,590
		-115 - 15	**0,012**	0,569	0,420
		62 162	0,583	0,151	0,201

Table S.3 – Summary of p-values of standard parametric one-tail paired t-test of the distribution of MVPA ROC_AUC values compared to chance level. Significant values are in bold, significant in both experiments are on grey background

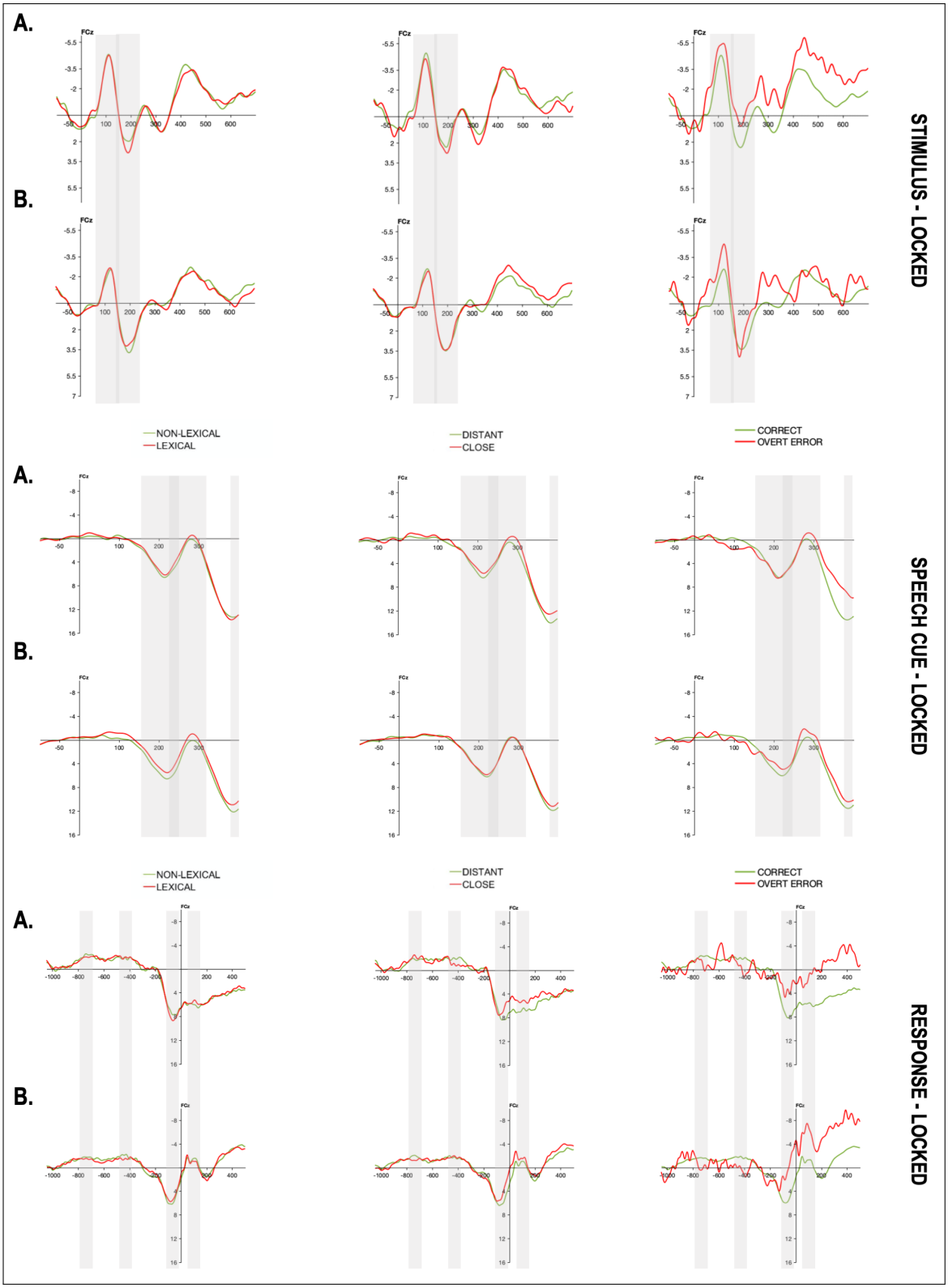


Figure S.1 – Grand average wave of FCz electrode for non-lexical (green) and lexical (red) trials (left column); for AP close (red) and AP distant (green) correct trials (middle column); for correct (green) and overt error (red) trials (right column) in Experiment I (panel A) and Experiment II (panel B) in stimulus - locked (top panel), speech cue-locked (central panel) and response-locked (down panel) epochs.

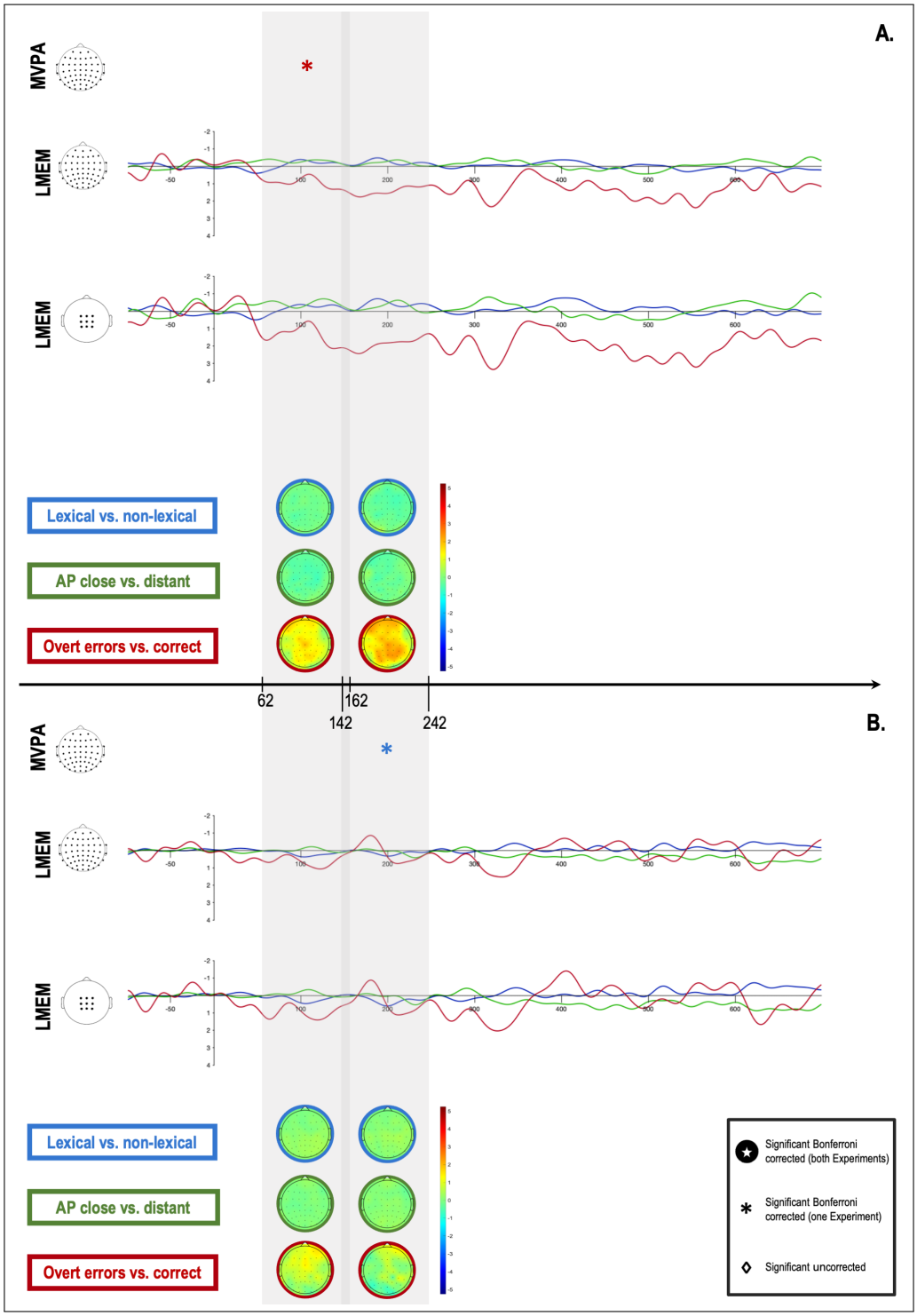


Figure S.2 – Summary of results of all performed analyses: MVPA, mean difference waves of all and 9 fronto-central electrodes and their topographic maps across Experiment I (panel A) and Experiment II (panel B.) in stimulus-locked epochs. Color code is used to differentiate the conditions: blue for lexicality, green for articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant Bonferroni corrected p-values if observed in one Experiment.

Experimental lists of Experiment I

A				B			
lexical		nonlexical		lexical		nonlexical	
0SF	cadeau rocher	0SF	belote ciment	0SF	cage rap	0SF	carreau roulette
0SF	douche salle	0SF	bordée fournée	0SF	fagot rumeur	0SF	caverne mouton
0SF	lierre poupe	0SF	cÅŞur robe	0SF	farine mission	0SF	ceinture région
0SF	malade sinus	0SF	cote lueur	0SF	patte nièce	0SF	doc geste
0SF	marine fission	0SF	dague four	0SF	pierre loupe	0SF	fête lobe
0SF	ministre seringue	0SF	disque suite	0SF	pote nuits	0SF	glace fuite
0SF	natte pièce	0SF	façon gardon	0SF	radeau cocher	0SF	lanière fêtard
0SF	note puits	0SF	filleul monteur	0SF	salade minus	0SF	lapin fusée
0SF	rage cap	0SF	fosse masque	0SF	sinistre meringue	0SF	lecteur joker
0SF	ragot fumeur	0SF	jointure boulette	0SF	souche dalle	0SF	maçon journée
0SF	tenue voiture	0SF	jonction loto	0SF	venue toiture	0SF	pelle risque
1SF	titre voile	0SF	lamelle têtard	0SF	vitre toile	0SF	rampe soeur
1SF	butte lave	0SF	manière cuisson	1SF	berger vison	0SF	verbe tour
1SF	clé bol	0SF	panne roc	1SF	blé col	1SF	boisson caresse
1SF	crochet briquet	0SF	pion vase	1SF	bonus toucan	1SF	case sueur
1SF	dentier répit	0SF	recteur tracas	1SF	brochet criquet	1SF	casque fraise
1SF	dune lieu	0SF	rouleau tonton	1SF	butin local	1SF	cordée frimeur
1SF	durée pédale	0SF	tête liège	1SF	clic foin	1SF	coupe frange
1SF	flic coin	1SF	barreau légion	1SF	gosier râteau	1SF	fonction troupier
1SF	foire prime	1SF	brique vieux	1SF	loir sac	1SF	gamelle ponton
1SF	gag troupe	1SF	contre braise	1SF	lune dieu	1SF	linge bourse
1SF	garage palette	1SF	coussin bouton	1SF	lutte bave	1SF	lion vote
1SF	gaule tare	1SF	fable place	1SF	maison raquette	1SF	molosse lardon
1SF	lutin bocal	1SF	filon croupier	1SF	parage galette	1SF	pause mouche
1SF	matin passage	1SF	forge course	1SF	patin massage	1SF	pelote fracas
1SF	nature ration	1SF	fraction paresse	1SF	poire frime	1SF	poussin savoir
1SF	pillage sommier	1SF	gerbe plaque	1SF	purée dédale	1SF	sable poudre
1SF	raison maquette	1SF	lampe boule	1SF	rame dose	1SF	tilleul buisson
1SF	rosier gâteau	1SF	moisson lavoir	1SF	rature nation	1SF	toison boulet
1SF	serre valve	1SF	peinture musée	1SF	rentier dépit	2SF	biche dieux
1SF	soir lac	1SF	pomme mâche	1SF	sillage pommier	2SF	bosse montre
1SF	tonus boucan	1SF	sapin poker	1SF	tag groupe	2SF	bouleau moto
1SF	verger bison	1SF	singe veste	1SF	taule gare	2SF	crique pousse
2SF	banque marque	1SF	touche grange	1SF	verre salve	2SF	flaque somme
2SF	casse tube	2SF	colosse primeur	2SF	bec mise	2SF	gorge canne
2SF	ciel fil	2SF	fiche selle	2SF	fiel cil	2SF	pilon bouture
2SF	coteau poupon	2SF	foudre soupe	2SF	manque barque	2SF	pointure conteur
2SF	dame rose	2SF	gousse cause	2SF	poteau coupon	2SF	tâche piège
2SF	faveur semelle	2SF	poison couture	2SF	saveur femelle	2SF	traction piment
2SF	mec bise	2SF	taverne poulet	2SF	tasse cube	2SF	vague foule

C				D			
lexical		nonlexical		lexical		nonlexical	
0SF	boule fosse	0SF	cadeau voiture	0SF	biche fraise	0SF	cil nièce
0SF	braise fiche	0SF	casse voile	0SF	bosse foule	0SF	farine galette
0SF	colosse monteur	0SF	faveur maquette	0SF	gorge frange	0SF	foin dalle
0SF	façon musée	0SF	lutin fission	0SF	lion piège	0SF	gare sac
0SF	grange forge	0SF	marine semelle	0SF	maçon fusée	0SF	loir toile
0SF	liège pion	0SF	natte coin	0SF	molosse conteur	0SF	lune cage
0SF	mâche touche	0SF	ragot palette	0SF	montre casque	0SF	lutte cube
0SF	masque contre	0SF	raison pédale	0SF	mouche tâche	0SF	maison cocher
0SF	vase cote	0SF	rosier sommier	0SF	verbe geste	0SF	poteau vison
1SF	veste gerbe	0SF	serre rose	0SF	vote case	0SF	purée rumeur
1SF	barreau cuisson	0SF	verger poupon	1SF	bouleau roulette	0SF	rentier pommier
1SF	bordée couture	1SF	butte cap	1SF	carreau buisson	0SF	salade dépit
1SF	course brique	1SF	ciel tare	1SF	ceinture piment	0SF	saveur massage
1SF	filleul tracas	1SF	coteau bocal	1SF	cordée bouture	0SF	sillage râteau
1SF	filon primeur	1SF	crochet fumeur	1SF	crique bourse	0SF	sinistre nation
1SF	four tête	1SF	dentier passage	1SF	fête tour	0SF	tag loupe
1SF	fraction têtard	1SF	douche puits	1SF	fonction journée	1SF	barque clic
1SF	jointure poker	1SF	dune prime	1SF	gamelle lardon	1SF	blé taule
1SF	jonction fournée	1SF	foire troupe	1SF	lanière moto	1SF	butin coupon
1SF	lamelle gardon	1SF	gag marque	1SF	lapin savoir	1SF	fagot toucan
1SF	lueur singe	1SF	garage ration	1SF	lecteur région	1SF	gosier toiture
1SF	manière loto	1SF	gaule lac	1SF	linge sueur	1SF	patin femelle
1SF	peinture ciment	1SF	lierre valve	1SF	pilon frimeur	1SF	pierre manque
1SF	place gousse	1SF	malade répit	1SF	pointure joker	1SF	poire groupe
1SF	plaque foudre	1SF	matin rocher	1SF	poudre flaque	1SF	pote fiel
1SF	recteur légion	1SF	mec lave	1SF	pousse glace	1SF	radeau mission
1SF	robe lampe	1SF	nature bison	1SF	rampe lobe	1SF	rame bec
1SF	roc disque	1SF	note lieu	1SF	risque doc	1SF	rap dieu
1SF	rouleau boulette	1SF	rage bise	1SF	soeur coupe	1SF	rature local
1SF	sapin lavoir	1SF	soir tube	1SF	somme pelle	1SF	tasse salve
1SF	selle pomme	1SF	tenue sinus	1SF	tilleul fracas	1SF	venue raquette
1SF	soupe coeur	1SF	titre fil	1SF	traction fêtard	1SF	verre bave
1SF	vieux dague	1SF	tonus seringue	1SF	vague dieux	1SF	vitre dose
2SF	belote poulet	2SF	banque poupe	2SF	boisson mouton	2SF	berger minus
2SF	cause panne	2SF	clé pièce	2SF	canne pause	2SF	bonus meringue
2SF	coussin paresse	2SF	dame bol	2SF	caverne troupier	2SF	brochet dédale
2SF	moisson bouton	2SF	durée gâteau	2SF	pelote boulet	2SF	mise nuits
2SF	poison tonton	2SF	flic salle	2SF	poussin caresse	2SF	parage criquet
2SF	suite fable	2SF	ministre boucan	2SF	sable fuite	2SF	patte col
2SF	taverne croupier	2SF	pillage briquet	2SF	toison ponton	2SF	souche frime

E				**F**			
lexical		nonlexical		lexical		nonlexical	
0SF	cap rage	0SF	boulette jointure	0SF	cocher radeau	0SF	cité biche
0SF	fission marine	0SF	ciment belote	0SF	dalle souche	0SF	fêtard lanière
0SF	fumeur ragot	0SF	cuisson manière	0SF	loupe pierre	0SF	fuite glace
0SF	pièce natte	0SF	four dague	0SF	meringue sinistre	0SF	fusée lapin
0SF	poupe lierre	0SF	fournée bordée	0SF	minus salade	0SF	geste doc
0SF	puits note	0SF	gardon façon	0SF	mission farine	0SF	journée maçon
0SF	rocher cadeau	0SF	liège tête	0SF	nièce patte	0SF	lobe fête
0SF	salle douche	0SF	loto jonction	0SF	nuits pote	0SF	mouton caverne
0SF	seringue ministre	0SF	lueur cote	0SF	rap cage	0SF	région ceinture
0SF	sinus malade	0SF	masque fosse	0SF	rumeur fagot	0SF	risque pelle
0SF	voile titre	0SF	monteur filleul	0SF	toile vitre	0SF	roulette carreau
1SF	voiture tenue	0SF	robe coeur	0SF	toiture venue	0SF	soeur rampe
1SF	bison verger	0SF	roc panne	1SF	bave lutte	0SF	tour verbe
1SF	bocal lutin	0SF	suite disque	1SF	col blé	1SF	boulet toison
1SF	bol clé	0SF	têtard lamelle	1SF	criquet brochet	1SF	bourse linge
1SF	boucan tonus	0SF	tonton rouleau	1SF	dédale purée	1SF	buisson tilleul
1SF	briquet crochet	0SF	tracas recteur	1SF	dépit rentier	1SF	caresse boisson
1SF	coin flic	0SF	vase pion	1SF	dieu lune	1SF	fracas pelote
1SF	gâteau rosier	1SF	boule lampe	1SF	dose rame	1SF	fraise casque
1SF	lac soir	1SF	bouton coussin	1SF	foin clic	1SF	frange coupe
1SF	lave butte	1SF	braise contre	1SF	frime poire	1SF	frimeur cordée
1SF	lieu dune	1SF	course forge	1SF	galette parage	1SF	joker lecteur
1SF	maquette raison	1SF	croupier filon	1SF	gare taule	1SF	lardon molosse
1SF	palette garage	1SF	grange touche	1SF	groupe tag	1SF	mouche pause
1SF	passage matin	1SF	lavoir moisson	1SF	local butin	1SF	ponton gamelle
1SF	pédale durée	1SF	légion barreau	1SF	massage patin	1SF	poudre sable
1SF	prime foire	1SF	mâche pomme	1SF	nation rature	1SF	savoir poussin
1SF	ration nature	1SF	musée peinture	1SF	pommier sillage	1SF	sueur case
1SF	répit dentier	1SF	paresse fraction	1SF	raquette maison	1SF	troupier fonction
1SF	rose dame	1SF	place fable	1SF	râteau gosier	1SF	vote lion
1SF	sommier pillage	1SF	plaque gerbe	1SF	sac loir	2SF	bouture pilon
1SF	tare gaule	1SF	poker sapin	1SF	salve verre	2SF	canne gorge
1SF	troupe gag	1SF	veste singe	1SF	toucan bonus	2SF	conteur pointure
1SF	valve serre	1SF	vieux brique	1SF	vison berger	2SF	foule vague
2SF	bise mec	2SF	cause gousse	2SF	barque manque	2SF	montre bosse
2SF	fil ciel	2SF	couture poison	2SF	cil fiel	2SF	moto bouleau
2SF	marque banque	2SF	poulet taverne	2SF	coupon poteau	2SF	piège tâche
2SF	poupon coteau	2SF	primeur colosse	2SF	cube tasse	2SF	piment traction
2SF	semelle faveur	2SF	selle fiche	2SF	femelle saveur	2SF	pousse crique
2SF	tube casse	2SF	soupe foudre	2SF	mise bec	2SF	somme flaque

G				H			
lexical		nonlexical		lexical		nonlexical	
0SF	contre masque	0SF	coin natte	0SF	case vote	0SF	cage lune
0SF	cote vase	0SF	fission lutin	0SF	casque montre	0SF	cocher maison
0SF	fiche braise	0SF	maquette faveur	0SF	conteur molosse	0SF	cube lutte
0SF	forge grange	0SF	palette ragot	0SF	foule bosse	0SF	dalle foin
0SF	fosse boule	0SF	pédale raison	0SF	fraise biche	0SF	dépit salade
0SF	gerbe veste	0SF	poupon verger	0SF	frange gorge	0SF	galette farine
0SF	monteur colosse	0SF	rose serre	0SF	fusée maçon	0SF	loupe tag
0SF	musée façon	0SF	semelle marine	0SF	geste verbe	0SF	massage saveur
0SF	pion liège	0SF	sommier rosier	0SF	piège lion	0SF	nation sinistre
0SF	touche mâche	0SF	voile casse	0SF	tâche mouche	0SF	nièce cil
1SF	boulette rouleau	0SF	voiture cadeau	1SF	bourse crique	0SF	pommier rentier
1SF	brique course	1SF	bise rage	1SF	bouture cordée	0SF	râteau sillage
1SF	ciment peinture	1SF	bison nature	1SF	buisson carreau	0SF	rumeur purée
1SF	coeur soupe	1SF	bocal coteau	1SF	coupe soeur	0SF	sac gare
1SF	couture bordée	1SF	cap butte	1SF	dieux vague	0SF	toile loir
1SF	cuisson barreau	1SF	fil titre	1SF	doc risque	0SF	vison poteau
1SF	dague vieux	1SF	fumeur crochet	1SF	fêtard traction	1SF	bave verre
1SF	disque roc	1SF	lac gaule	1SF	flaque poudre	1SF	bec rame
1SF	foudre plaque	1SF	lave mec	1SF	fracas tilleul	1SF	clic barque
1SF	fournée jonction	1SF	lieu note	1SF	frimeur pilon	1SF	coupon butin
1SF	gardon lamelle	1SF	marque gag	1SF	glace pousse	1SF	dieu rap
1SF	gousse place	1SF	passage dentier	1SF	joker pointure	1SF	dose vitre
1SF	lampe robe	1SF	prime dune	1SF	journée fonction	1SF	femelle patin
1SF	lavoir sapin	1SF	puits douche	1SF	lardon gamelle	1SF	fiel pote
1SF	légion recteur	1SF	ration garage	1SF	lobe rampe	1SF	groupe poire
1SF	loto manière	1SF	répit malade	1SF	moto lanière	1SF	local rature
1SF	poker jointure	1SF	rocher matin	1SF	pelle somme	1SF	manque pierre
1SF	pomme selle	1SF	seringue tonus	1SF	piment ceinture	1SF	mission radeau
1SF	primeur filon	1SF	sinus tenue	1SF	région lecteur	1SF	raquette venue
1SF	singe lueur	1SF	tare ciel	1SF	roulette bouleau	1SF	salve tasse
1SF	têtard fraction	1SF	troupe foire	1SF	savoir lapin	1SF	taule blé
1SF	tête four	1SF	tube soir	1SF	sueur linge	1SF	toiture gosier
1SF	tracas filleul	1SF	valve lierre	1SF	tour fête	1SF	toucan fagot
2SF	bouton moisson	2SF	bol dame	2SF	boulet pelote	2SF	col patte
2SF	croupier taverne	2SF	boucan ministre	2SF	caresse poussin	2SF	criquet parage
2SF	fable suite	2SF	briquet pillage	2SF	fuite sable	2SF	dédale brochet
2SF	panne cause	2SF	gâteau durée	2SF	mouton boisson	2SF	frime souche
2SF	paresse coussin	2SF	pièce clé	2SF	pause canne	2SF	meringue bonus
2SF	poulet belote	2SF	poupe banque	2SF	ponton toison	2SF	minus berger
2SF	tonton poison	2SF	salle flic	2SF	troupier caverne	2SF	nuits mise

Experimental list of Experiment II

lexical		nonlexical	
0SF	2SF	0SF	2SF
balade sonnet	bague digue	ballet sapeur	bande membre
boule fosse	banque marque	bourse foin	baroque patin
braise fiche	bêche peau	bride farce	beige poste
cadeau rocher	bêcheur piquet	courant record	beurrier passion
durée fiction	belote poulet	dépit fourmis	bible daube
festin dada	carte torse	façade dépense	cerne forces
fête bac	casse tube	ferme brute	coin tact
gaule sang	cause panne	glace faune	compte taux
grange forge	ciel fil	gué signe	confit puma
lavoir sapin	coteau poupon	lamelle sofa	coup puce
liqueur pétale	croupier taverne	légume pédale	croyante tampon
lueur secteur	daron bélier	levier souris	dopage barbu
mâche touche	derme ton	mari talent	double tombe
masque contre	faveur semelle	mer cape	fiston sondage
matin taxi	gage coût	miracle sérum	galion bascule
ministre seringue	gallon bourde	modèle famille	gamin bonheur
musée façon	garage bateau	muette top	guide corne
natte pièce	malais barquette	nombre pont	maillet baleine
note pain	mec bise	nymphe peuple	mot bave
poker jointure	mouture boisson	potion jumelle	muret bacille
rage pÃt'le	nature marine	rabat ficus	niveau moquette
ragot fondeur	paresse coussin	roque pelle	palier cuiller
râpe paille	poing coupe	rythme poivre	plumet tournée
soupe liège	primeur tension	sucre loutre	pouce corde
tacle rÃt't	prune baume	tapis visage	proue berge
tenue voiture	suite fable	tireuse vanille	sÅŞur feutre
terrine vernis	têtard ponton	toge rêne	tabou pavot
vase cote	torche prÃt'ne	veau cure	tofu cobra
verre toile	tracas cantine	vieux taupe	tonne ponce
village panneau	veste gerbe	visite pétrin	vers gifle

3 Linking movement and cognition through the hierarchical buildup of internal modeling

Dorokhova Lydia
Loh Kep Kee
Anton Jean-Luc
Nazarian Bruno
Sein Julien
Belin Pascal
Runnqvist Elin

Outline

Abstract

Movement and cognition have traditionally been investigated separately, hypothesized to rely on different mechanisms. More recently, networks of simpler motor reflexes have been proposed as precursors of more complex cognitive functions (Friston, 2011; Pezzulo et al., 2021). In particular, there are theories that propose recycling motor control processes for cognitive processing, such as predictive coding or internal modeling (Ito, 2008). While the brain network associated with the control of movement is known to include the cerebellum and primary sensory and motor cortex, the hierarchical buildup of internal modeling would suppose a shift towards more prefrontal regions such as inferior frontal gyrus (IFG) and parieto-temporal regions. In this study we were interested in assessing this hypothesized gradual shift within individuals. We acquired the fMRI data of 20 participants performing recurrent movements of different body parts with two conditions: movement alone and monitored movement. In the movement alone condition participants were instructed to perform specific and paced movements with either the hand or the tongue. In the monitored movement condition, they received the additional instruction not to touch the bed of the fMRI or the walls of their mouth. The monitored movement condition was used to maximize the cognitive involvement in the task. The results averaged across body parts revealed a hierarchical shift in cortex with increased cognitive demands: movement alone activated primary motor, primary sensory regions, pre-SMA, basal ganglia and cerebellar regions. Crucially, additional activations in more prefrontal (BA10, BA9, BA8 and BA45) as well as in associative parietal (supramarginal gyrus) regions were observed for monitored movement. These findings provide direct evidence for a hierarchical buildup of internal modeling and support the hypothesis that this mechanism may have emerged gradually during evolution.

Key words: motor control; executive control; hierarchical processing; internal modeling; fMRI

3.1 Introduction

A powerful framework accounting for how humans monitor and optimize their actions is that of internal modeling (also known as active inference or predictive coding, e.g. Wolpert et al. 1995 Friston 2011; Ito 2008). Having its origin in the domain of motor control, this process is thought to involve a controlled object (CO, a body part); a controller (CT, a center sending commands to the body part) and an internal model (a copy of the dynamics of the body part, e.g., Ito 2008, see Figure 3.1). With repeated practice entailing refinement of the internal model, the latter can ultimately be used to guide our actions by predicting their sensory consequences, alleviating the need to rely on external sensory feedback. Many authors have proposed that similar mechanisms of control may apply to cognition, connecting motor processing with cognitive processing through prediction of the neural consequences of actions (Wolpert and Flanagan 2001) and suggesting an evolutionary link from simple motor reflexes to high-level cognitive processing (Friston 2011; Pezzulo et al. 2022). The shift from modeling movement to modeling mental representations has been proposed to be mirrored in a hierarchical shift in the involved brain structures. For example, Ito 2008 proposed the use of similar internal modeling mechanisms and neural networks to control both movement and mental actions but with a slightly different brain basis. While the cerebellum is the hypothesized center of internal modeling for both types of actions, the controller is thought to be located in primary motor cortex for voluntary movements (Roland 1984 and in more prefrontal structures for mental actions (Sanes and Donoghue 2000; Ito 2008; Sokolov et al. 2017). Moreover, in the case of mental actions, the body part (CO) and the feedback loop in between CT and CO would be replaced by a Mental Model (MM) located in temporo-parietal cortex (see Figure 3.1). Put differently, there is a proposed shift from primary towards associative cortical regions. Empirical support for the hypothesis that mental actions also involve the use of internal modeling comes from an increasing amount of studies that have observed fronto-temporo-parietal or cortico-cerebellar brain networks sensitive to manipulations of predictability in more complex cognitive conditions that do not contain or partial out the motor dimension (e.g., Imamizu and Kawato 2009; Caucheteux et al. 2023; Runnqvist et al. 2016; Runnqvist et al. 2021). On the other hand, patterns of hierarchically organized motor and executive control have been observed in frontal and parietal cortices (e.g., Amiez and Petrides 2018; Badre 2008; Choi et al. 2018; Grafton and Hamilton 2007) as well as in the cerebellum (D'Mello et al. 2020). Concretely, these regions have been shown to contain subdivisions in which somatomotor proximal regions represent more concrete goal-oriented action and somatomotor distal regions represent more abstract and future oriented processing. Relatedly, Grafton and Hamilton 2007 established a motor hierarchy (from lowest to highest level of complexity) that they tested in the

domain of action observation (grasping movements). They identified a distributed set of brain regions that were differentially activated as a function of behavioral motor complexity in relationship to the final outcome of a movement. In summary, previous studies have provided valuable insights concerning both motor and mental actions separately in what concerns internal modeling. Previous studies have also made a compelling case that the brain is organized along functional hierarchical gradients (e.g., Huntenburg et al. 2018 Guell et al. 2018), and that such gradients or similar hierarchical patterns are present for action control and action observation. Here we wanted to link internal modeling and hierarchical processing theories by testing whether variations in cognitive demands result in hierarchical shifts in frontal, parietal and cerebellar regions for self-produced actions. We took Ito's model as our point of departure, thus assuming that the controller would shift from BA4 towards IFG and prefrontal structures, and that mental models would engage posterior parietal cortex. Based on the observation of hierarchical processing gradients in parietal cortex, we added the assumption that internal modeling of motor actions would recruit regions more proximate to primary somatosensory cortex (BA1). Feedback from the body part to the motor cortex is mediated by primary sensory areas (e.g., Kwon et al. 2016). We thus assumed that as the feedback becomes internal through the predicted sensory consequences it would become progressively posterior along the motor to cognitive continuum. Finally, while Ito mentions the cerebellum as the locus of the internal model, he does not specify a hierarchical shift for this structure. However, recent literature has shown that the cerebellum, just as cortex, is organized along hierarchical processing gradients. Concretely, a primary gradient has been identified progressing from primary (motor) to transmodal (default-mode network (DMN), task-unfocused) regions. Moreover, cerebellar activation in a task manipulating the contextual and temporal cognitive control also followed this processing gradient (D'Mello et al. 2020). Thus, we added the assumption that also cerebellar structures would display a hierarchical shift from unimodal towards transmodal regions with increased cognitive demands. To test this, we conducted a simple motor task with two conditions that differed in cognitive demands. Specifically, we asked participants to perform hand or tongue movements with or without a supplementary cognitive rule entailing a constant monitoring of the outcome of their action. While all movements were assumed to contain a cognitive component because they were performed in an instructed and paced fashion, the supplementary rule was intended to maximize this component.

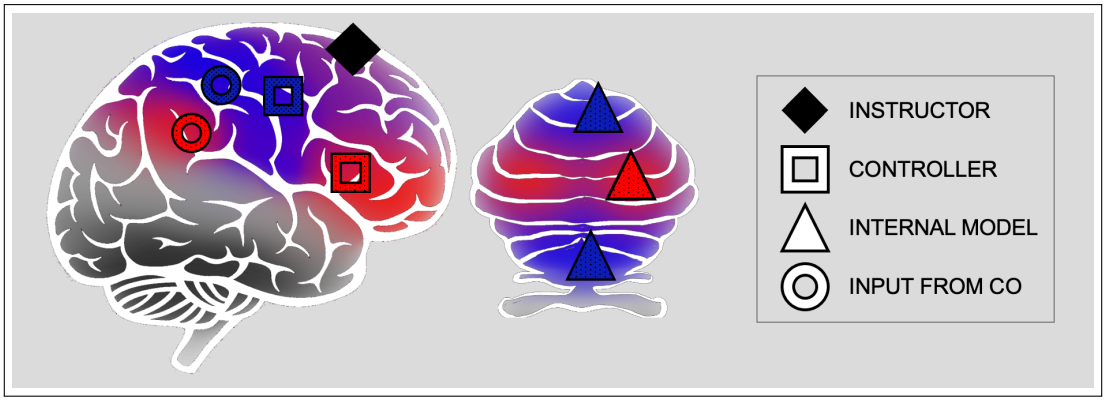


Figure 3.1 – Integrating Functional Gradients (Huntenburg et al. 2018; Guell et al. 2018) and Ito's Internal Model (Ito 2008) for Simple Motor (blue) and More Cognitively Complex Motor (red) Acts. The gradients from blue to red represents the continuum from primary or unimodal (motor/sensory) to more associative or transmodal regions.

3.2 Methods

3.2.1 Participants

The study obtained ethical approval from the regional ethical committee Comité de Protection des Personnes Sud Méditerranée I, with the approval filed under identification number 2017-A03614-49. Twenty right-handed young adults (min = 20, max = 33, mean = 25,5 y.o.) took part in the study (12 women) in exchange for monetary compensation.

3.2.2 Task

Participants were instructed to produce recurrent movements using different body parts: the hand and the tongue (Amiez and Petrides 2014). For hand movements, participants made up and down movements with their wrist and hand (fingers straightened) using their working (right) hand. For tongue movements, participants made vertical circular movements with their tongue, keeping their lips closed but loosening their jaws to create more space in the oral cavity. For both types of movements, participants were instructed to make one movement cycle per second: up and down or a circle. The schematic representation of the movements can be found in Figure 3.2. For both body parts there were two movement conditions: the movement alone (described above, MV) and the monitored movement (MN). For the monitored movement condition (MN), participants were given additional instructions not to touch the bed of the fMRI scanner for hand movements, and not to touch the walls of the oral cavity or teeth for tongue movements. There was also a baseline condition during which participants did not perform any movement.

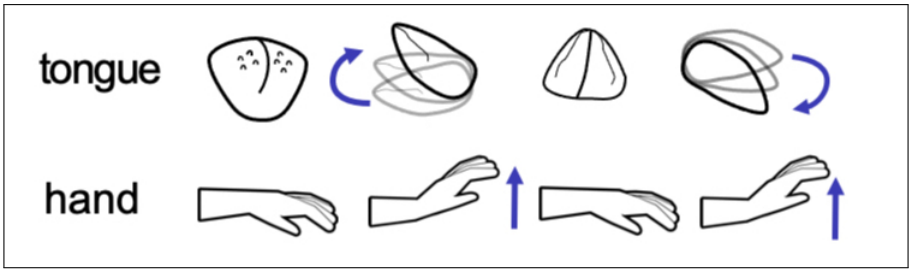


Figure 3.2 – Schematic representation of movements performed during the task.

3.2.3 Procedure

A blocked design fMRI protocol was used. All trials (as shown in Figure 3.3) began with an instruction displayed on the screen ("tongue", "hand" or "fixate"). To differentiate between the movement alone condition (MV) and the monitored movement (MN) condition, the former was displayed in white and the latter in red. For the movement conditions, participants were instructed to perform the movement repeatedly while looking at a fixation cross. For the baseline condition participants were instructed not to move and only look at the fixation cross. One trial (MV, MN or FX) lasted 15 seconds. One run contained 2 repetitions of each of the 5 conditions (10 trials per run: 2 hand MV, 2 hand MN, 2 tongue MV, 2 tongue MN, 2 FX). Participants completed 4 runs of 10 trials (120 sec per condition overall).

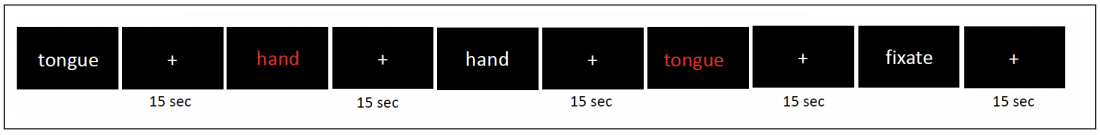


Figure 3.3 – Procedure

One trial (MV, MN or FX) lasted 15 seconds. One run contained 2 repetitions of each of the 5 condition trials (10 trials per run: 2 hand MV, 2 hand MN, 2 tongue MV, 2 tongue MN, 2 FX). Participants completed 4 runs of 10 trials (120 sec per condition overall). Participants were trained to perform the task before entering the scanner, and they also had a short training session in the scanner before the task in order to monitor their headmovements.

3.2.4 MRI Data Acquisition

A 3-Tesla Siemens Prisma Scanner (Siemens, Erlangen, Germany) was used to collect data using a 64-channel head coil at the Marseille MRI Center (Centre IRM-INT@CERIMED, UMR7289 CNRS & AMU). Functional BOLD images (EPI sequence, 72 slices per volume, multi-band = 4, repetition time = 1.386 s, spatial resolution = $2 \times 2 \times 2\ mm^3$, echo time = 33.4 ms, flip angle = 56°) covering the whole brain were acquired during the task

performance. Four runs were acquired using a BOLD protocol containing 155 images. Whole-brain anatomical MRI data were acquired using high-resolution structural T1 - weighted image (MPRAGE sequence, repetition time = 2.3 s, spatial resolution = 0.8 × 0.8 × 0.8 mm^3, echo time = 3.1 ms, inversion time = 0.9 s flip angle = 9°) in the sagittal plane. Prior to functional imaging, whole-brain Fieldmap images were acquired with a spin-echo EPI sequence with the same spatial parameters as the BOLD images and acquired twice with opposite phase encode directions along the anterior-posterior axis with the following parameters: TR/TE = 7220/59 ms, voxel size = 2 x 2 x 2 mm^3, slices = 72, flip angle = 90/180°.

3.2.5 Image processing and analyses

The data was preprocessed using fMRIPrep 20.2.6 (Esteban et al. 2018b; Esteban et al. 2018a; RRID:SCR_016216), which is based on Nipype 1.7.0 (K. Gorgolewski et al. 2011; K. J. Gorgolewski et al. 2018; RRID:SCR_002502).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al. 2010, distributed with ANTs 2.3.3 Avants et al. 2008, RRID:SCR_004757), 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, RRID:SCR_002823, Zhang et al. 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al. 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) 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, RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym)

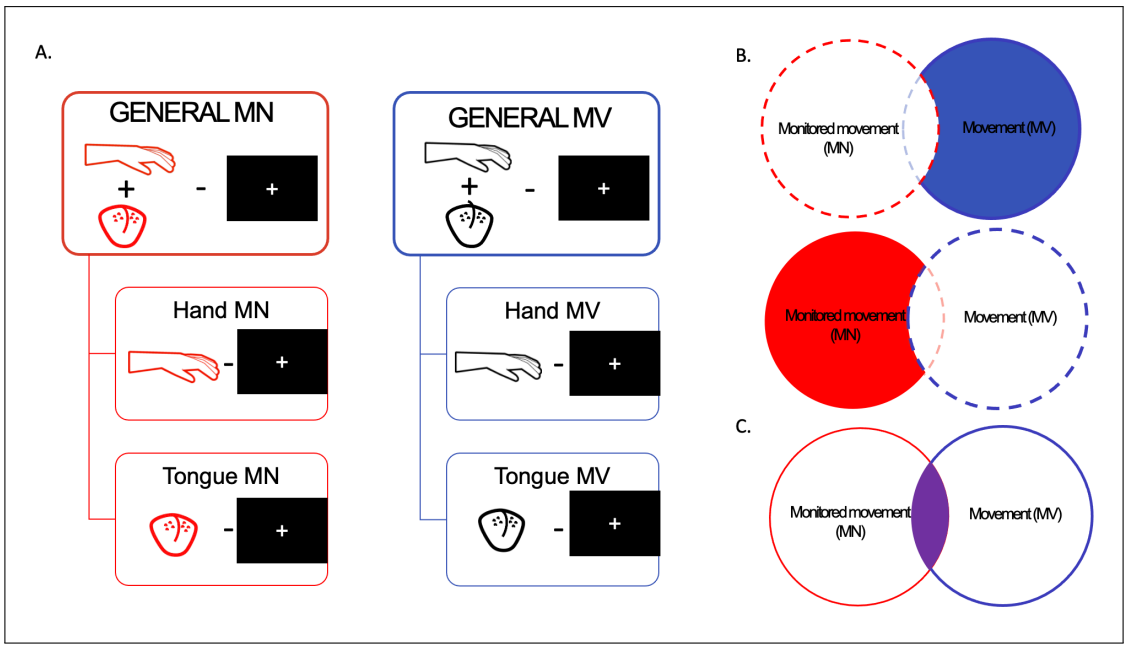


Figure 3.4 – Panel A: Schematic representation of contrasts in the model, Monitored movement (MN) is presented in red, Movement (MV) in blue. Fixation was subtracted from all movement (MV and MN) conditions. Panel B: Schematic representation of exclusive masking where clusters that were active for MV and MV and MN together are not considered for statistical analyses. Panel C: Schematic representation of exclusive masking where clusters that were active for MV and MV and MN together were considered.

Functional data preprocessing

For each of the 4 BOLD runs found per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band references (SBRefs). A B0-nonuniformity map (or fieldmap) was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` Cox and Hyde 1997 (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) 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 `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with nine degrees of freedom 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) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al. 2002). First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. 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. Two global signals are extracted within the CSF and the WM. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the anatomical denoising approach (aCompCor: two probabilistic masks (CSF and WM) are generated in anatomical space. Components are calculated separately within the WM and CSF masks. For each CompCor decomposition, the 12 first components with the largest singular values are retained. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. 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` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Subsequently, the data was analyzed using the Statistical Parametric Mapping software (SPM Development Team 1991) on MATLAB R2018b (Inc. 2020).The preprocessed BOLD data were smoothed with an isotropic three-dimensional Gaussian kernel (FWHM = 5 mm).For each run and each subject, a general linear model (GLM) was designed with 6 behavioral conditions: fixation (FX), hand movement (HMV), tongue movement (TMV), hand monitored movement (HMN), tongue monitored movement (TMN), instruction. The regressors were convolved by canonical hemodynamic response. The default SPM autoregressive model AR(1) and high-pass filter (128s) were applied. 48 nuisance regressors were also included in the model : 24 head movement derived and 24 aCompCor (12 CSF and 12 WM) and the two global signals within CSF and WM masks. To obtain contrasts targeting movement alone, we subtracted the fixation condition from all movement conditions (HMV and TMV). To obtain contrasts targeting monitored movement we subtracted fixation from all monitored movement conditions (HMN and TMN). These first level contrast images were then entered in a second level factorial 2x2 design (factor body part with the levels hand and mouth; factor movement with the levels alone and monitored). This allowed us to assess the main effect of movement alone regardeless of the body part (MV), as well as the main effect of monitored movement, regardless of the body part (MN) in addition to examining the body parts separately (A schematic representation of this procedure is given in Figure 3.4, panel A). To achieve a FDR corrected extent threshold of $p < 0.05$ at the cluster level (voxel level $p < 0.001$, uncorrected), a minimum cluster size of 25 voxels was used. To investigate commonalities across the contrasts, we subsequently masked the MV contrast inclusively with the MN contrast

(Figure 3.4, panel C.). Similarly, to investigate specificities of each contrast, we masked MN exclusively with MV, and MV exclusively with MN (Figure 3.4, Panel B.). All masked contrasts used a threshold of p=.001 for the second contrast. The same procedure was also applied to examine each of the body parts, hand and tongue, separately.

3.3 Results

The results of the analysis masking movement inclusively with monitored movement are summarized in Table 3.1 and on Figure 3.5). Significantly overlapping activation was observed in frontal (bilaterlly in primary motor cortex (BA4) and cerebellar regions (motor areas (lobules V, VI, and VIIIb). Furthermore, a cluster of thalamic activation was commonly engaged across the two conditions.

The results of the analysis in which movement was exclusively masked by monitored movement are summarized in Table 3.2 and on Figure 3.5). Movement alone revealed significant clusters of activation in frontal (primary motor (BA4) and dorsal premotor regions (BA6), parietal (primary somatosensory cortex (B1)), and cerebellar (motor (lobule VI) and non-motor (CrusI)) regions. Heightened activation for movement alone was also observed in the basal ganglia (caudate, globus pallidus) and in the insula (BA13).

				MNI coord.		
BA	*q*FDR-corr	*k*E	T	x	y	z
BA4	<.001	7570	13.15	-38	-22	54
CB(V)	<.001	2282	11.72	24	-46	-22
CB(VI)	<.001	614	11.67	-18	-64	-8
BA4	<.001	1590	10.52	56	-4	32
CB(VIIIb)	<.001	841	10.8	12	-64	-50
Thalamus	<.001	575	9.28	-14	-18	8

Table 3.1 – General Movement and General Monitored movement (for both body parts) commonly activated clusters

GMV exclusively masked by GMN								GMN exclusively masked by GMV						
				MNI coord.								MNI coord.		
BA	*q*FDR-corr	*k*E	T	x	y	z		BA	*q*FDR-corr	*k*E	T	x	y	z
BA6(PMd)	.001	153	5.52	42	-6	36	frontal	BA9	0.009	101	4.9	46	36	30
BA4	.001	110	5.23	68	-8	22		BA6(PMv)	<.001	138	4.71	62	10	20
								BA10	0.001	114	5.23	28	56	-16
								BA6(pre-SMA)	<.001	249	6.99	-16	-12	70
								BA6/BA44	<.001	184	5.73	-54	6	4
								BA8	0.014	62	5.49	-6	40	44
								BA10	0.009	95	5.48	-34	52	-8
								BA45	0.009	63	4.52	-34	16	12
								BA24	<.001	370	6.11	0	-2	48
BA1	.001	476	7.54	-50	-30	62	parietal	BA40	<.001	180	4.73	54	-28	30
BA1	.001	266	6.31	-56	-12	14		BA1	<.001	227	5.13	-30	-38	54
								BA40	0.009	74	5.01	-52	-28	26
CB(VI)	.021	60	5.04	28	-72	-20	cerebellar	CB(I-IV)		73	4.82	2	-48	-2
CB(VI)	.001	189	4.97	16	-78	-16								
CB(CrusI)	.032	49	4.64	50	-52	-32								
GlobPal	.001	118	5.57	-20	2	-4	other	Thalamus	<.001	175	4.84	-10	-22	4
Caudate	.021	80	6.5	-18	20	4								
	.021	60	7.01	-20	12	6								
BA13(Insula)	.021	77	4.49	-42	0	0								

Table 3.2 – General Movement and General Monitored Movement significant clusters of activity exclusively masked by General Monitored Movement and General Movement respectively, at uncorrected p<.001. BioImage Suite Web 2023 atlas is used to define BAs and Diedrichsen Lab 2023 altas for cerebellar areas. For BA6, labels from Mayka et al. 2006 are enclosed in parentheses, and labels from Heim et al. 2007 are presented after the slash.

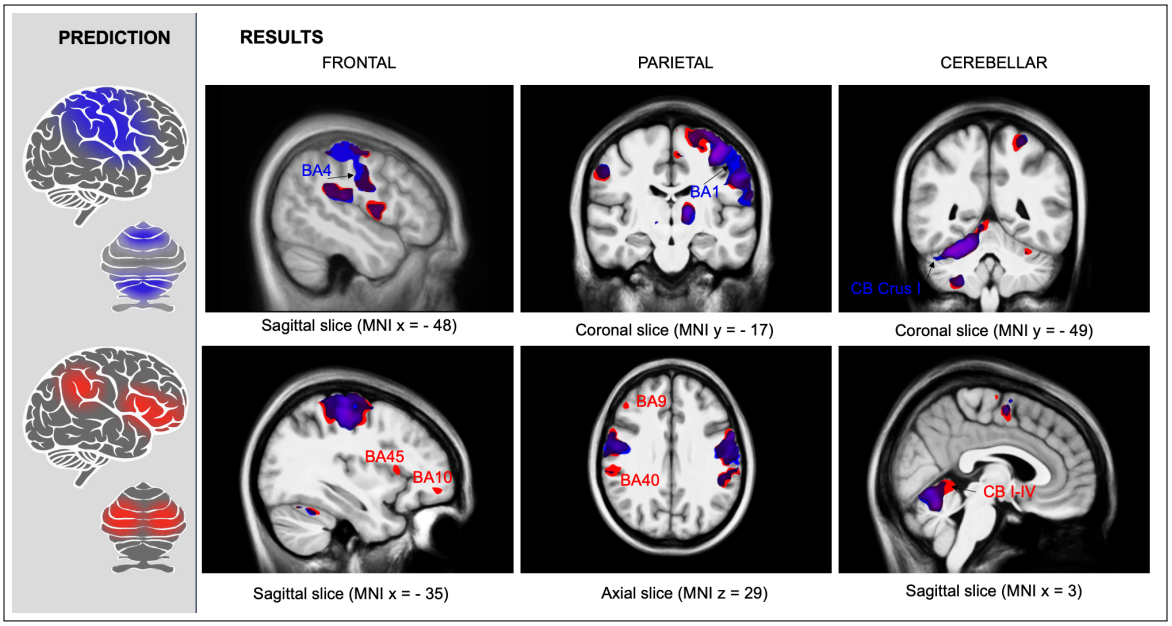


Figure 3.5 – Aggregate anatomical scan depicting the mean representation across all participants, juxtaposed with functional statistical maps corresponding to GMV (blue) and GMN (red) conditions. Overlapping regions (violet) denote areas commonly activated, while distinct colors denote exclusive masking outcomes.

The results of the analysis in which monitored movement was exclusively masked by movement are summarized in Table 3.2 and on Figure 3.5). Monitored movement resulted in significantly active clusters in frontal (laterally in ventral premotor cortexBA6 extending to the ventrolateral prefrontal cortex (BA44, BA45), and to orbitofrontal and dorsolateral prefrontal cortices (BA10, BA9); medially in the anterior cingulate cortex (BA24), pre-SMA (BA6) and BA8), parietal (primary somatosensory cortex (BA1) and

supramarginal gyrus (BA40), and cerebellar (motor regions (I-IV)) areas. A significant cluster was also observed in the Thalamus.

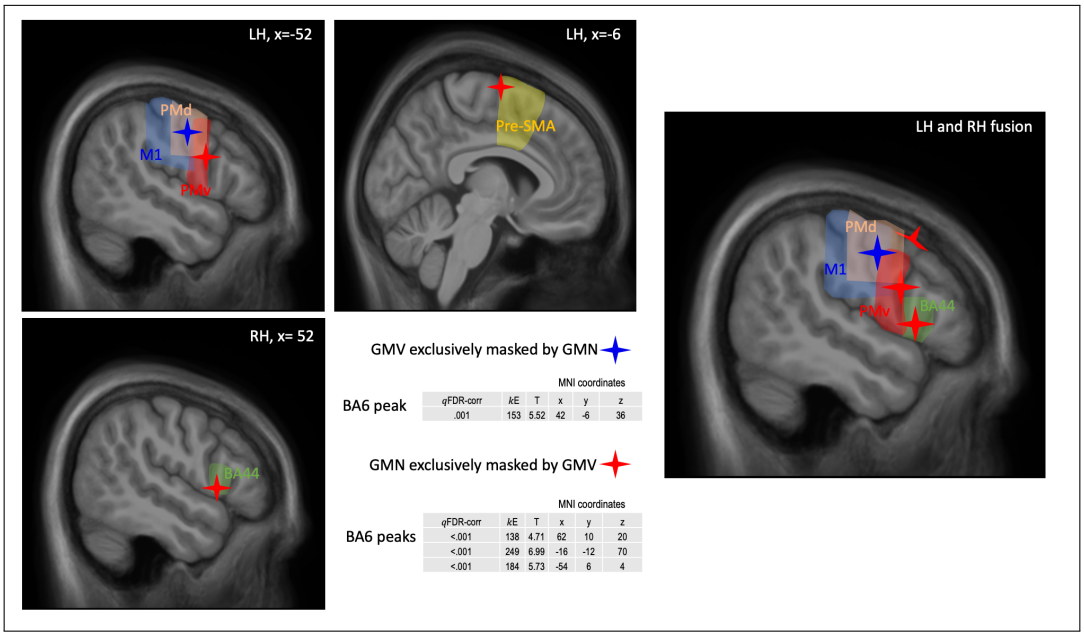


Figure 3.6 – An amalgamation of all BA6 peaks in conditions uniquely masked by their respective opposites (stars). Various transparent colors demarcate the functional areas (adapted from Mayka et al. 2006 and Heim et al. 2007.

Significant clusters of TMV, TMN, HMV, and HMN exclusively masked by the opposite condition (TMN, TMV, HMN, and HMV respectively), can be found in Tables SM.1 and SM.2. A comprehensive discussion of these findings can be found in the Supplementary Discussion 3.6.2 of the Supplementary Materials 3.6 .

3.4 Discussion

Our study aimed to investigate the hierarchical buildup of internal modeling within individuals during the performance of motor actions with low or high cognitive demands. Previous research has emphasized the role of internal modeling in facilitating smooth motor execution by anticipating sensory consequences without heavy reliance on external feedback. Such internal modeling is thought to involve a fronto-parieto-cerebellar brain network. By conducting a motor task with two conditions differing in cognitive demands, we sought to explore the potential adaptation of the brain network sustaining internal modeling in response to this increase in cognitive demands. Following Ito's model on cerebellar internal modeling for both motor and mental actions, as well as recent studies showing the existence of hierarchical functional gradients related to motor and executive control (Badre and D'Esposito 2009; D'Mello et al. 2020; Guell et al. 2018; Huntenburg et al. 2018) we expected a shift from primary towards associative areas in frontal, parietal and cerebellar regions when comparing movement alone and monitored movement. In what follows we will zoom in more closely on the findings observed in frontal, parietal and cerebellar regions respectively.

Frontal hierarchical shift along a rostro-caudal axis

Two clusters located bilaterally in primary motor cortex were commonly activated by both movement and monitored movement. Given that the exact same movement was performed in both conditions this is not surprising. One cluster was exclusive to movement alone. Again, because the exact same movement was performed across the two conditions, this cluster is likely not related to the movement itself but to a difference in control across the two conditions. This supports the hypothesis that BA4 serves a more important role as a controller in the movement alone condition but not in the monitored movement condition.

Moving further along the rostral direction, both movement alone and monitored movement resulted in significant clusters in BA6. BA6 is a region of considerable size and functional heterogeneity (e.g., Picard and Strick 1996; Rizzolatti et al. 2002), comprising a mesial premotor part (MPMC) which can be subdivided into pre-SMA and SMA, a lateral premotor part which can be subdivided into a dorsal and a ventral (PMd and PMv), at the border of BA4 and BA44 respectively (e.g., Mayka et al. 2006). A closer inspection of the local maxima in BA6 related to movement and monitored movement revealed an image largely consistent with the idea of a hierarchical subdivision. Movement revealed a cluster of activation in PMd, very close to BA4 (see Figure 3.6). Monitored movement resulted in three clusters, two of which peaked bilaterally in PMv right at the border of BA44 (or in BA 44 following Heim et al. 2007), and the third was located in the pre-SMA region (see dicussion of medial frontal clusters). Previous studies have shown that the PMv is involved in the visuomotor transformation required to

elaborate motor commands for goal oriented actions (such as grasping or manipulating objects) while PMd was shown to have an earlier role related to cued motor preparation or sequencing, (e.g., Davare et al. 2006). Other authors have put forward the hypothesis that both PMd and PMv are functionally related as they would both be involved in mental simulation of either natural and familar (PMv) or associatively learned (PMd) events (e.g., Cisek and Kalaska 2004). Finally, a number of studies have identified the PMv and BA 44 as the locus of mirror neurons (e.g., Binkofski and Buccino 2006; Rizzolatti et al. 2002). Regardless of the details of these different functions that have been ascribed to the dorsal and ventral parts of the premotor cortex, all of them align well with the findings of the present study that seem to suggest a privileged role of PMv in observing and monitoring one's own actions (be it through visuomotor transformation, mental rehearsal or mirror neurons).

Even further prefrontally, BA45, BA9 and BA10 all resulted in significant clusters for monitored movement. These prefrontal regions are thought to be responsible for more abstract, schematic cognitive control (e.g., Badre and Nee 2018). Thus, in frontal cortex the clusters revealed across our conditions aligned perfectly along a hierarchical rostrocaudal gradient going from BA4 and dorsal premotor cortex to ventral premotor cortex, IFG and prefrontal regions. Furthermore, several medial frontal regions outside of this processing gradient were also preferentially activated by monitored movement (Pre-SMA, BA24 and BA8), consistent with a role of medial frontal cortex in self-monitoring of different types of acions (Botvinick et al. 2001). These findings are consistent with a synergy between the lateral and medial parts of the frontal cortex as discussed by Badre and Nee 2018. For instance, BA9, or dorsolateral prefrontal cortex, is known to work in concert with the ACC (BA24) to monitor conflict and adapt behavior (e.g., Alexander and Brown 2015; Boschin et al. 2017; Botvinick et al. 2004).These authors further point to a similar caudal to rostral hierarchical organization of the medial frontal cortex, equally consistent with our findings as several mid and rostral medial clusters showed heightened activation for monitored movement.

Parietal hierarchical shift along a posterior-going axis

No shared parietal clusters were observed across movement and monitored movement, but both contrasts revealed significant clusters along the postcentral gyrus peaking in the left hemisphere of the somatosensory cortex. In the case of movement alone, these clusters were located respectively at the superior (S1) and inferior (S2) boundaries of somatosensory cortex. In the case of monitored movement there was one cluster located at a slightly posterior site closer to the border of the intraparietal sulcus. Following the previous literature, this distribution suggests an emphasis on sensory motor integration in the case of movement alone and on perceptual-motor coordination in the case of monitored movement (e.g., Borich et al. 2015; Tunik et al. 2007). Moving posteriorly, the monitored movement contrast resulted in enhanced

bilateral activation in the supramarginal gyrus (BA40). This is consistent with the proposed shift of the controlled object from a body part in the case of movement to a parietal mental model in the case of mental activities (e.g., Ito 2008). More generally, our results are consistent with previously reported hierarchical gradients along the parietal direction, from the central sulcus through BA1 to the posterior parietal cortex, as indexed by decreases in selectivity and laterality (e.g., Saadon-Grosman et al. 2020).

Activation patterns in the cerebellum do not mirror the cortical hierarchy

Both movement and monitored movement commonly activated lobules V and VIII. These regions are known to contain somatotopic representations of the body including the tongue and hand that were used in our task, and form part of the double motor representation in the cerebellum (e.g., Guell et al. 2018). Moreover, the posterolateral lobule VI was also commonly activated across both contrasts. This region, adjacent to the anterior lobe of the cerebellum is similarly known to contain somatotopic representations of motor effectors and has been shown to be involved in conditions requiring awareness over specific cues to guide current actions (concrete and current cognitive control, e.g., D'Mello et al. 2020). As both contrasts of interest of the current study involved keeping track of the pace of movement, a sort of concrete and current cognitive control, this result is not surprising. Contrary to our expectations, movement alone engaged additional portions of lobule VI as well as Crus I, while monitored movement engaged the anterior lobules I-IV. Crus I is the cognitive region *par excellence* in the cerebellum, and has previously been observed to be involved in abstract, schematic cognitive control (e.g., D'Mello et al. 2020). Lobules I-IV on the other hand are typically categorized as part of the somatomotor network (e.g., Diedrichsen 2006). In the context of the presence of hierarchical effects in the expected direction in frontal and parietal cortices, but not in the cerebellum, these results suggest an independent organization across cerebral and cerebellar regions supporting movement and more complex cognitive behavior. This stands in contrast to a previous report supporting a parallel organizational processing structure across prefrontal cortex and the posterolateral cerebellum. A main difference between that study and ours is the type of task that was used. D'Mello et al. 2020 focused on variations in contextual and temporal control, ranging from concrete and current to abstract and distant, and very little attention was drawn to the execution of the movement required to perform the task (a key-press with the index finger). On the contrary, here we focused participants' attention on the execution of a movement that was in itself more complex and differed across tongue and hand trials, while varying the need for applying and the amount of current and concrete control. While this difference does not provide any explanation of the unexpected cerebellar pattern observed here, it underscores the need of considering a wide range of tasks and variables in order to understand the principles that govern hierarchical processing in both cortical and cerebellar structures. An alternative possibility that might account

for the pattern we observed is that Crus I is in fact activated in preparation for the future demands of the more cognitiviely demanding monitored movement contrast as a means of planning. D'Mello et al. 2020 observed a correlation between future reaction times (i.e., occurring right after fMRI activations) and Crus I. We don't have any possibility of assessing this hypothesis as our task did not generate reaction times, but note that this merits testing in future research. Concerning the engagement of lobules I-IV in the monitored movement contrast, this anterior part of the cerebellum has been linked to the ventral attention network in addition to the somatotmotor network (e.g., Diedrichsen 2006). In the context of the current task, the increased activation for monitored movement would be consistent with a role of pivoting attention as new information is introduced.

The hierarchical buildup of internal modeling

Ito's 2008 model largely fits with the observations of the present study as for both our movement and monitored movement contrasts we observed a network of frontal, parietal and cerebellar regions, and the increase in cognitive demands was accompanied by a shift in both frontal and parietal structures from primary towards associative regions. A clear strength of this model is its capacity to predict the coordinated patterns across several distant brain regions and how their interaction accounts for the optimizing of a range of human actions, from movement to cognition. Other models have described a hierarchy in the prefrontal cortex providing a processing hierarchy of executive control (e.g., Badre and Nee 2018). While these models have no means of predicting synergy across distant brain regions, their rich detail concerning within-region interactions might provide an explanatory framework for the relationship between the many clusters observed in frontal cortex, and possibly shed light on the principles applying to coordinated hierarchical processes as here. Some attempts in this direction were made of Badre and Nee 2018 to account for synergies between prefrontal and medial frontal cortices. Similarly, D'Mello et al. 2020 discuss synergies across prefrontal and cerebellar regions. They point to the possibility that hierarchies might occur at the level of networks rather than at the level of individual brain regions, much in line with Ito's internal modeling network. Moving to a different domain of research, there could be much knowledge to integrate from the large body of research on mirror neurons that is related to the present study via the notion of self-monitoring as a sort of action observation. A distributed fronto parietal hierarchy of action representation has been proposed within the framework of a mirror neuron system (e.g., Grafton and Hamilton 2007).

Conclusion

In summary, our study has provided valuable insights into the hierarchical organization of brain regions involved in internal modeling during self-produced actions. While our findings support the notion of hierarchical shifts in frontal and parietal regions in response to cognitive demands, the cerebellum exhibited a distinct pattern, emphasiz-

ing the complexity of neural processing across these structures. This research enhances our understanding of how internal modeling adapts to varying cognitive demands and provides a foundation for further investigations into the complex interplay between motor control and cognitive processing. Ultimately, it underscores the importance of considering a wide range of tasks and variables to unravel the principles governing hierarchical processing in both cortical and cerebellar structures.

3.5 References

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex [Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . .]. Neural Computation, 27(11), 2354–2410. Retrieved September 29, 2023, from https://direct.mit.edu/neco/article-abstract/27/11/2354/8116

Amiez, C., & Petrides, M. (2014). Neuroimaging Evidence of the Anatomo-Functional Organization of the Human Cingulate Motor Areas. Cerebral Cortex, 24(3), 563–578. https://doi.org/10.1093/cercor/bhs329

Amiez, C., & Petrides, M. (2018). Functional rostro-caudal gradient in the human posterior lateral frontal cortex. Brain Structure & Function, 223(3), 1487–1499. https://doi.org/10.1007/s00429-017-1567-z

Avants, B., Epstein, C., Grossman, M., & Gee, J. (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

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends in Cognitive Sciences, 12(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? Nature Reviews Neuroscience, 10(9), 659–669. https://doi.org/10.1038/nrn2667

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. Trends in Cognitive Sciences, 22(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Behzadi, Y., Restom, K., Liau, 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

Binkofski, F., & Buccino, G. (2006). The role of ventral premotor cortex in action execution and action understanding. Journal of Physiology, Paris, 99(4-6), 396–405. https://doi.org/10.1016/j.jphysparis.2006.03.005

BioImage Suite Web. (2023). BioImage Suite Web Atlas. https://bioimagesuiteweb.github.io/webapp/mni2tal.html

Borich, M. R., Brodie, S. M., Gray, W. A., Ionta, S., & Boyd, L. A. (2015). Understanding the role of the primary somatosensory cortex: Opportunities for rehabilitation. Neuropsychologia, 79(Pt B), 246–255. https://doi.org/10.1016/j.neuropsychologia.2015.07.007

Boschin, E. A., Mars, R. B., & Buckley, M. J. (2017). Transcranial magnetic stimulation to dorsolateral prefrontal cortex affects conflict-induced behavioural adaptation in a Wisconsin Card Sorting Test analogue. Neuropsychologia, 94, 36–43. https://doi.org/10.1016/j.neuropsychologia.2016.11.015

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Botvinick, M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. Nature Human Behaviour. https://doi.org/10.1038/s41562-022-01516-2

Choi, E. Y., Drayna, G. K., & Badre, D. (2018). Evidence for a Functional Hierarchy of Association Networks. Journal of Cognitive Neuroscience, 30(5), 722–736. https://doi.org/10.1162/jocn_a_01229

Cisek, P., & Kalaska, J. F. (2004). Neural correlates of mental rehearsal in dorsal premotor cortex. Nature, 431(7011), 993–996. https://doi.org/10.1038/nature03005

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

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: i. segmentation and surface reconstruction. NeuroImage, 9(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Davare, M., Andres, M., Cosnard, G., Thonnard, J.-L., & Olivier, E. (2006). Dissociating the role of ventral and dorsal premotor cortex in precision grasping. The Journal of Neuroscience, 26(8), 2260–2268. https://doi.org/10.1523/JNEUROSCI.3386-05.2006

Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. NeuroImage, 33(1), 127–138. https://doi.org/10.1016/j.neuroimage.2006.05.056

Diedrichsen Lab. (2023). Diedrichsen Lab Atlas Viewer. https://www.diedrichsenlab.org/imaging/AtlasViewer/viewer.html

D'Mello, A. M., Gabrieli, J. D. E., & Nee, D. E. (2020). Evidence for Hierarchical Cognitive Control in the Human Cerebellum. Current biology: CB, 30(10), 1881–1892.e3. https://doi.org/10.1016/j.cub.2020.03.028

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R.,

Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018a). Fmriprep. Software. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018b). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, Supplement 1, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Friston, K. (2011). What Is Optimal about Motor Control? Neuron, 72(3), 488–498. https://doi.org/10.1016/j.neuron.2011.10.018

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, 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

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. Software. https://doi.org/10.5281/zenodo.596855

Grafton, S., & Hamilton, A. (2007). Evidence for a Distributed Hierarchy of Action Representation in the Brain. Human movement science, 26, 590–616. https://doi.org/10.1016/j.humov.2007.05.009

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

Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum [Publisher: eLife Sciences Publications Limited]. Elife, 7, e36652.

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., & Amunts, K. (2007). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. Human Brain Mapping, 30(2), 392–402. https://doi.org/10.1002/hbm.20512

Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-Scale Gradients in Human Cortical Organization. Trends in Cognitive Sciences, 22(1), 21–31. https://doi.org/10.1016/j.tics.2017.11.002

Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: Implications for higher-order cognitive functions [Place: Germany Publisher: Springer]. Psychological Research, 73(4), 527–544. https://doi.org/10.1007/s00426-009-0235-1

Inc., T. M. (2020). Matlab version: 9.8.0 (r2020a). Natick, Massachusetts, United States. https://www.mathworks.com

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

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

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. PLOS Computational Biology, 13(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kwon, S. E., Yang, H., Minamisawa, G., & O'Connor, D. H. (2016). Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. Nature Neuroscience, 19(9), 1243–1249. https://doi.org/10.1038/nn.4356

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

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging. NeuroImage, 31(4), 1453–1474. https://doi.org/10.1016/j.neuroimage.2006.02.004

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1844), 20200531. https://doi.org/10.1098/rstb.2020.0531

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. Cerebral Cortex (New York, N.Y.: 1991), 6(3), 342–353. https://doi.org/10.1093/cercor/6.3.342

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. Current Opinion in Neurobiology, 12(2), 149–154. https://doi.org/10.1016/s0959-4388(02)00308-2

Roland, P. E. (1984). Organization of motor control by the normal human brain. Human Neurobiology, 2(4), 205–216.

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

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), tgab038. https://doi.org/10.1093/texcom/tgab038

Saadon-Grosman, N., Arzy, S., & Loewenstein, Y. (2020). Hierarchical cortical gradients in somatosensory processing. NeuroImage, 222, 117257. https://doi.org/10.1016/j.neuroimage.2020.117257

Sanes, J. N., & Donoghue, J. P. (2000). Plasticity and primary motor cortex. Annual Review of Neuroscience, 23, 393–415. https://doi.org/10.1146/annurev.neuro.23.1.393

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

SPM Development Team. (1991). Statistical Parametric Mapping (SPM) (Version Version 2) [Software package for fMRI data analysis.]. http://www.fil.ion.ucl.ac.uk/spm/

Tunik, E., Rice, N., Hamilton, A., & Grafton, S. (2007). Beyond grasping: Representation of action in human anterior intraparietal sulcus. NeuroImage, 36 Suppl 2, T77–86. https://doi.org/10.1016/j.neuroimage.2007.03.026

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

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. Current Biology, 11(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration [Publisher: American Association for the Advancement of Science]. Science, 269(5232), 1880–1882. Retrieved September 8, 2021, from http://www.jstor.org/stable/2889276

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

3.6 Supplementary Materials

3.6.1 Supplementary Results

TMV exclusively masked by TMN									TMN exclusively masked by TMV							
					MNI coord.									MNI coord.		
	BA	*q*FDR-corr	*k*E	T	x	y	z			BA	*q*FDR-corr	*k*E	T	x	y	z
LH	BA41	.005	136	6,35	-62	-4	4	cortical	LH	BA6	<.001	301	6,24	-48	-4	14
	BA6	.005	104	4,9	-38	-14	36			BA6	<.001	898	5,69	-22	-12	62
	BA40	.024	48	4,6	-60	-28	50			BA40	<.001	245	5,53	-56	-28	22
	BA9	.024	65	4,56	-44	32	40			BA10	.004	71	5,14	-24	52	-8
	BA40	.024	54	4,52	-66	-30	20			BA45	.004	70	4,95	-34	22	8
										BA7	.004	74	4,04	-40	-42	50
RH	BA4	<.001	668	6,42	36	-6	16		RH	BA6	<.001	180	4,99	22	-10	72
	BA40	.024	51	4,24	58	-22	32									
LH	CB (VI)	<.001	174	4,39	-8	-72	-12	cerebellar	LH	CB (VIIb)	.018	51	5,13	-18	-70	-44
									RH	CB (VIIIa)	<.001	151	4,97	22	-68	-56
	Thalamus	.005	101	6,93	16	-16	8	other		GlobPal	.03	42	4,97	26	-8	-4
	Insula	<.001	174	6,36	-42	0	-2									
	Putamen	.024	66	4,82	-26	-8	-8									

Table SM.1 – Tongue Movement and Tongue Monitored Movement significant clusters of activity exclusively masked by Tongue Monitored Movement and Tongue Movement respectively, at uncorrected p<.001

HMV exclusively masked by HMN									HMN exclusively masked by HMV							
					MNI coord.									MNI coord.		
	BA	*q*FDR-corr	*k*E	T	x	y	z			BA	*q*FDR-corr	*k*E	T	x	y	z
LH	BA1	<.001	402	8,22	-50	-30	58	cortical	LH	BA40	<.001	139	5,63	-60	-30	24
	BA40	<.001	185	5,53	-44	-32	18			BA7	<.001	127	4,98	-20	-46	66
	CB V	.006	74	4,4	-2	-66	-6									
	BA18	.006	80	3,95	-22	-76	-14		medial	BA6	<.001	400	5,89	0	-2	50
RH	BA22	.006	57	6,24	68	-32	18		RH	BA40	<.001	206	4,95	58	-28	28
RH	CB (VIIIa)	.001	152	5,58	12	-64	-40	cerebellar								
	CB (VI)	<.001	281	5,29	22	-70	-20									
	Putamen	<.001	335	5,84	-22	12	4	other								

Table SM.2 – Hand Movement and hand Monitored Movement significant clusters of activity exclusively masked by Hand Monitored Movement and Hand Movement respectively, at uncorrected p<.001

3.6.2 Supplementary Discussion

The findings regarding the varying body parts provided intriguing insights into the lateralization and differences in brain activation patterns during motor actions. Specifically, when comparing tongue movement (TMV) with exclusive masking by tongue monitored movement (TMN), bilateral activation in primary motor and somatosensory cortex was observed. However, the inversed exclusive masking did not confirm this bilateral activation and instead positioned the tongue monitored movement (TMN) to the left. A different pattern was observed for hand movements, where the movement of the right hand was found to be left-lateralized, and the monitored movement, exclusively masked by movement, showed above all bilateral activation. These intriguing observations raise several possible interpretations.

One interpretation is related to the functional specialization of the tongue as an articulator involved in language processing. The lateralization of tongue movement control

could be linked to the evolutionary development of the left hemisphere's specialization for language processing. This raises the question of whether proficient signers of sign language would exhibit left-lateralized monitored movements in the hand, serving as a marker of language anchoring to the left hemisphere.

An alternative explanation could stem from the differences in the nature of the movement. The tongue movement was solenoidal, characterized by maintained difficulty throughout the process due to constant monitoring challenges imposed by the obstacles (walls of the mouth and teeth). In contrast, the up and down movement of the hand had only one obstacle (downward segment), making monitoring apparent only during the downward phase of the movement. According to the Assymetrical sampling theory Poeppel 2003, such conditions involving solenoidal movements might entail different brain wiring. Tongue monitored movement, which involves fast-integrating changes and constant monitoring, may be associated with left hemisphere involvement. On the other hand, hand monitored movement might be slower and segmented since one movement segment does not require external signal monitoring. This raises the question of how performing a solenoidal movement with the hand might affect the localization of the monitored movement.

4 Speech monitoring activations in the anterior cingulate cortex depend on cingulate sulcal morphology.

Dorokhova Lydia
Anton Jean-Luc
Nazarian Bruno
Sein Julien
Belin Pascal
Chanoine Valérie
Loh Kep Kee
Runnqvist Elin

Outline

Abstract

The self-monitoring of speech recruits the anterior cingulate cortex (ACC) region in the human brain (Runnqvist et al. 2021). This region shows substantial in- terindividual variations in sulcal morphology : while everyone has a cingulate sulcus (CgS) in both hemispheres, some individuals may have an additional paracingulate sulcus (PCgS) that runs dorsally and parallel to the CgS in one or both hemispheres. Critically, cingulate sulcal morphology has been demonstrated to influence both anatomical and functional organisation in the ACC region. Here, we wanted to investigate whether functional activations associated with speech error monitoring are influenced by whether or not a PCgS is present. To this end, we performed an analysis focusing on the ACC region from the data of 24 participants who performed a speech production task designed to induce errors (Runnqvist et al. 2021) in an event-related fMRI protocol. BOLD activation in trials where participants committed a speech error was contrasted with trials where participants made a correct production. Participants' hemispheres were classified into three groups based on their cingulate sulcal morphology : 1) Prominent âĹŠ when prominent paracingulate sulcus is present that runs dorsal and parallel to the cingulate sulcus, 2) Absent âĹŠ when a paracingulate is absent , and 3) Emerging âĹŠ when there are short sulcal segments above the cingulate sulcus that do no qualify as a prominent paracingulate sulcus. We averaged the BOLD contrast corresponding to error trials minus correct trials across hemispheres in each of the three cingulate morphology groups. Our results revealed distinct patterns of activation within the ACC based on cingulate sulcal morphology: In individuals with a prominent PCgS, error-related activations were observed primarily along the PCgS. In contrast, individuals lacking a PCgS exhibited activations predominantly on the cingulate sulcus (CgS). Interestingly, in emerging cases, the ACC activation is observed on both the PCgS and CgS. These results underscore the importance of considering individual sulcal morphology when localising functional activations in the human anterior cingulate region. Importantly, we revealed potential anatomical landmarks for localising speech monitoring cortical regions in individual brains which are highly valuable for neurosurgical interventions and further functional investigations. Finally, our findings support the hypothesis that the self-monitoring of errors during language production recruits a primitive brain circuitry that is involved in the feedback-driven control of orofacial and vocal movements.

Key words: speech monitoring; vocal feedback control; ACC; sulcal morphology; fMRI

4.1 Introduction

Producing speech errors leads to increased activity in the anterior cingulate cortex (ACC) region (also known as MCC: midcingulate cortex) (e.g. Chapter 1 or Runnqvist et al. 2021; Gauvin et al. 2016). In current literature, this speech error-related ACC activity is often explained by the theory of domain-general conflict-based monitoring (e.g., Nozari et al. 2011; Gauvin et al. 2016; Gauvin and Hartsuiker 2020). At its core, the conflict monitoring theory postulates that the medial frontal cortical region, including the ACC, is continually engaged in evaluating the levels of conflict during ongoing cognitive processes. When a certain threshold of conflict is surpassed, these structures communicate this information to other frontal brain regions responsible for executive control, prompting necessary adjustments in processing dynamics. In essence, the occurrence of conflict itself signifies a need for heightened cognitive control, precluding the need of external sensory feedback. This theory offers a unified framework that accommodates action monitoring both during planning stages and execution, presenting a continuum of conflict where overt errors represent the most extreme manifestation.

An alternative perspective posits a connection between human speech monitoring and vocal feedback monitoring, drawing parallels across human and non-human primates, as elucidated in studies by Loh and colleagues (Loh et al. 2020; see also Runnqvist et al. 2021). These authors have argued that the pars opercularis of the inferior frontal gyrus (IFG), also known as area 44, governs cognitive control over orofacial and non-speech vocal responses in primates. Simultaneously, the midcingulate cortex specializes in analyzing vocal non-speech feedback, driving adaptive responses. For the uniquely human control of speech vocal information, the engagement of the pars triangularis of the IFG (area 45) and pre-SMA becomes imperative. Direct evidence for this vocal feedback control network has mainly arisen from studies providing external vocal feedback (Loh et al. 2020). Recent findings have revealed the presence of a similar network (BA 44, 45, pre-SMA, ACC) in the context of overt speech errors, but not for contexts of high conflict where ultimately a correct utterance is produced (Runnqvist et al. 2021; Todorović et al. 2023). This hints at the possibility of internal feedback, generated through the proprioceptive or acoustic signal of self-produced speech, in driving the ACC activation related to speech errors.

This study aims to explore this intriguing notion further by capitalizing on interindividual variability in the sulcal morphology of the ACC, a factor underscored in previous research (Amiez and Petrides 2014; Loh et al. 2018; Loh et al. 2020). Specifically, some individuals exhibit a single cingulate sulcus (CgS) in the ACC, while others possess two sulci – a CgS and a paracingulate sulcus (PCgS), forming a paracingulate gyrus (Figure 4.1). This anatomical variation significantly impacts the localization of task-related activations, as evidenced by the work of Loh and colleagues. Notably, the location of

activation related to both tongue movement and external feedback processing are modulated by this sulcal variability. In populations with a PCgS, these activations align with the PCgS, whereas in those with only a CgS, they are confined to the CgS. Furthermore, areas activated during feedback processing notably overlap with tongue-movement-related regions within the PCgS-present population. This observation suggests the involvement of a shared network in governing various self-produced actions that might include speech production. We thus hypothesize that producing a speech error might trigger analogous processes to those involved in external feedback processing during learning tasks. Accordingly, we set out to investigate whether the localization of activation peaks associated with speech error production indeed varies based on the presence or absence of a PCgS – a crucial question that has not yet been addressed in prior research and forms the primary focus of this study.

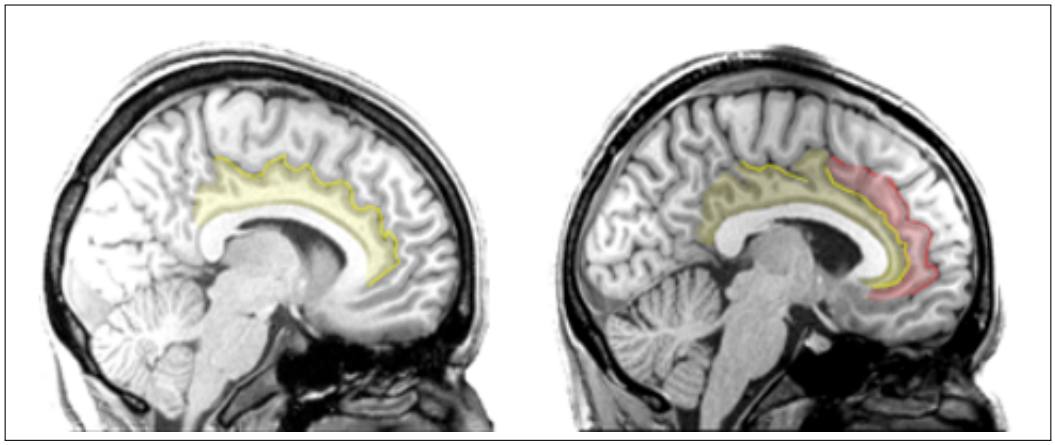


Figure 4.1 – Examples of a hemisphere with only one CgS and a cingulate gyrus (left) and a hemisphere with a CgS and a prominent PCgS and paracingulate gyrus (right). Sulci are displayed with lines of different colors: yellow for the CgS, red for the PCgS; the corresponding color shades overlay the corresponding gyri.

4.2 Methods

4.2.1 Participants

24 participants from Runnqvist et al. 2021 (Chapter 1) were retained for analyses.

4.2.1.1 Procedure

Participants performed an error-eliciting word pair production task (SLIP, spoonerisms of laboratory induced predisposition, Runnqvist et al. 2021). Word pairs appeared on the screen for 748 ms. Participants read these pairs of words silently. When a question mark appeared participants had to produce the last pair that they had read on the screen, aiming at starting their utterance before the appearance of an exclamation

mark. If no question mark appeared (40% of filler trials), word pairs were followed by a blank screen for 340 ms. If there was a question mark it stayed on the screen for 544 ms and was followed by an exclamation mark that remained for 1020 ms. The word pairs were constructed of 320 printed French nouns (the same used in Runnqvist et al. 2016). Priming pairs appeared on the screen before the target pairs to elicit errors by switching the initial letters of the words (e.g., *mole sail* ⇒ *sole mail; mole fence* ⇒ *fole mence*). If participants produced the pair correctly despite the priming, the trial was considered as "correct"; any production that did not match the written word pair was considered as "error". Each participant was presented with 460 unique word combinations (80 targets of which 40 lexical and 40 nonlexical error outcome, 240 primes and 140 fillers). Each participant completed 6 experimental runs in which word pairs were repeated 3 times in different orders (see Runnqvist et al. 2021 for a more detailed description).

4.2.2 Data

Anatomical and functional MRI images from Runnqvist et al. 2021 (Chapter 1) were used in this study.

4.2.2.1 Data Acquisition

Data were gathered at the Marseille MRI Center (Centre IRM-INT@CERIMED, UMR7289 CNRS & AMU) using a 3-Tesla Siemens Prisma Scanner (Siemens, Erlangen, Germany). The data acquisition employed a 64-channel head coil. Functional images, obtained through an EPI sequence, encompassed 54 slices per volume and utilized a multi-band accelerator factor of 3. The repetition time was set to 1.224 seconds, ensuring a spatial resolution of 2.5x2.5x2.5 mm^3, an echo time of 30 milliseconds, and a flip angle of 65°. These functional images covered the entire brain and were acquired during the task execution. Additionally, whole-brain anatomical MRI data were procured through a high-resolution structural T1-weighted image, utilizing an MPRAGE sequence with a repetition time of 2.4 seconds, a spatial resolution of 0.8x0.8x0.8 mm^3, an echo time of 2.28 milliseconds, and a flip angle of 8 °, in the sagittal plane. Before conducting functional imaging, a fieldmap image was also acquired using a dual echo gradient-echo acquisition with a repetition time of 7.06 seconds, a spatial resolution of 2.5 x 2.5.x2.5 mm^3, an echo time of 59 milliseconds, and a flip angle of 90°. The Fieldmap images served a crucial role in the process of correcting distortion and motion in the data. The functional volumes underwent spatial realignment and normalization, leveraging a combination of deformation field, co-registered structural images, and sliced functional images. Subsequently, these volumes were smoothed using an isotropic Gaussian kernel with a full-width at half-maximum of 5 mm. To address issues related to head

movements and functional data outliers, the Artefact Detection Tools (ART) from the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon 2012) were employed. This involved the definition of regressors of no interest. The identification of outlier scans was carried out automatically using ART, which relied on statistical measures. Specifically, outlier thresholds were determined using a 97th percentile criterion, which exceeded normative sample values. These thresholds included a global-signal z-threshold of 5 and a subject-motion threshold of 0.9 mm.

4.2.2.2 Data Processing

In Runnqvist et al. 2021, a univariate analysis of the entire brain was conducted. To achieve this, a general linear model (GLM) was constructed individually for each subject. Within each GLM, 7 regressors were incorporated. These regressors were designed to model variables of interest in Runnqvist et al. 2021: response accuracy, the lexical status of error priming, and the phonetic distance of target pair onsets. In addition to the regressors of interest, regressors of no interest were also included in the analysis. These no-interest regressors were derived from ART (Artifact Detection Tools) text files, with one file per subject. Each of these ART files contained information about outlier scans related to global signal fluctuations and head movements. The regressors of interest were convolved with the canonical hemodynamic response function, and the default SPM autoregressive model AR(1) was applied to account for temporal autocorrelation. Furthermore, to prepare the functional data for analysis, a high-pass filter with a cutoff of 128 seconds was employed. Subsequently, statistical parametric maps for each experimental factor and for each participant were computed at the first level of analysis.

4.2.2.3 Data used in the present study

For the purpose of this study, we used 24 anatomical images, spatially normalized to the avg152 T1 −weighted brain template defined by the Montreal Neurological Institute (MNI) using the default parameters of nonlinear transformation, and 24 functional images (first-level) corresponding to the contrast of the response accuracy (committed errors vs. correct productions).

4.2.3 Analyses

Individual brain anatomies

T1-weighted anatomical scans of 24 participants were annotated manually for the presence of a paracingulate sulcus (PCgS) in the left and right hemisphere separately: three labels were assigned: "prominent", "emerging" and "absent". The "prominent"

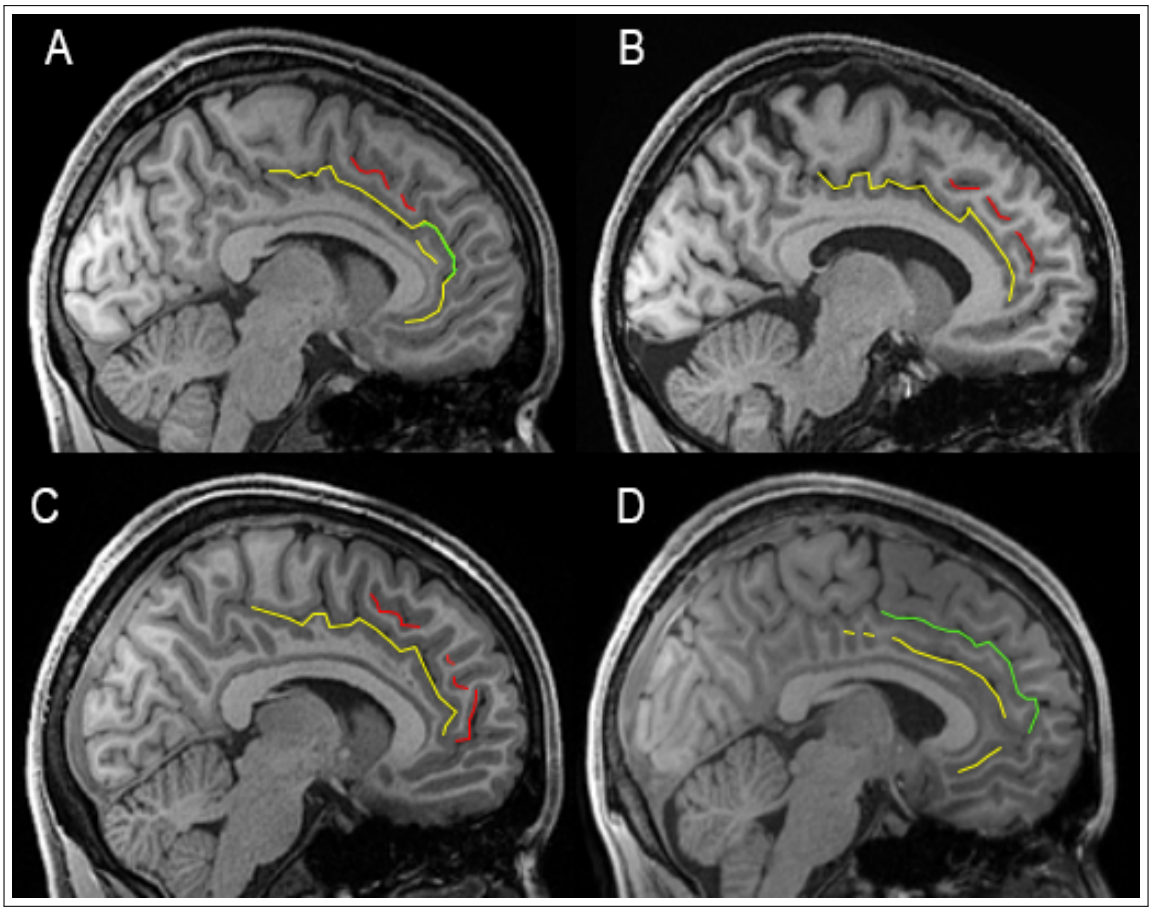


Figure 4.2 – Examples of several emerging cases. The CgS is always displayed in yellow color, the possible PCgS is displayed in red and sulci that could be both CgS and PCgS are in green. The two examples of the top panels are left hemispheres and the two examples of the bottom panels are right hemispheres. Reasons for the classification as "Emerging": A - short and discontinued, B- discontinued , C - discontinued , D - not enough lateral depth.

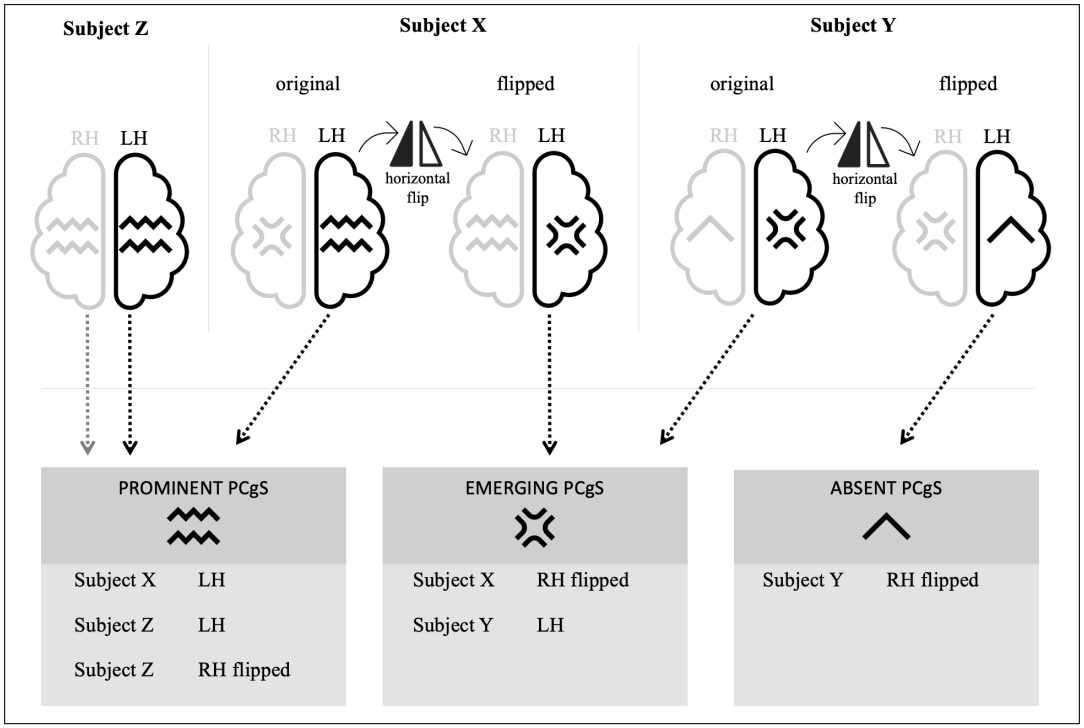


Figure 4.3 – A schematic representation of the process of flipping and dispatch into the groups. RH - right hemisphere, LH - left hemisphere. The flipping procedure was done on T1-weighted scans as well as on functional contrast maps.

PCgS label accounted for the appearance of two deep parallel sulci longer than 40 mm, the "absent" label was assigned for the presence of only one deep cingulate sulcus in the ACC (Figure 4.1). The "emerging" label was assigned to group ambiguous cases that did not satisfy the conditions of two previous categories: the upper sulcus was not long, laterally deep enough or discontinued to be classified as such. The examples of several "emerging" cases are given in Figure 4.2. 5 participants (21%) were observed with no PCgS neither in the left (LH) nor in the right (RH) hemispheres, additionally 8 were found with no right PCgS; the total number of absent cases was therefore 18 hemispheres. 8 participants (33%) were observed to have left "prominent" PCgS (LPCgS), and 6 participants (25%) were observed to have right "prominent" PCgS (RPCgS), resulting in a total of 14 "prominent" cases. 11 participants were found to have "emerging" PCgS in the left hemisphere and 5 in the right hemispeheres, resulting in 16 hemispheres (Table SC.1). The sagittal view anatomical scans of all subjects could be found in Figure SC.1 in the Supplementary materials 4.6.

Subsequently, all right hemispheres were flipped horizontally to become left hemispheres (using the *flipud* expression of the imCalc function of SPM (SPM Development Team 1991) on MATLAB R2018b (Inc. 2020). This was done to increase the number of observable cases and with assumption that the effect of ACC is bilateral (Runnqvist et al. 2021). Then, all originally left and left-appearing (flipped right) hemispheres were assigned to the three groups based on their cingulate sulcal morphology: (1) promi-

nent PCgS (14 hemispheres); (2) absent PCgS (18 hemispheres); (3) emerging PCgS (16 hemispheres). If the same participant had two prominent PCgS hemispheres, both of them were assigned to the prominent PCgS group. As a next step, mean anatomical scans for each group were created.

Functional activations

The same flipping procedure of the right hemispheres was performed on the first-level contrast statistical maps corresponding to contrast of *errors vs. correct* trials using the *flipud* expression of the imCalc function of SPM (SPM Development Team 1991) on MATLAB R2018b (Inc. 2020) for each subject. As a consequence, all functional data of interest was on the left. 1-sample t-test analysis of variance was performed on each group ("prominent PCgS", "absent PCgS" and "emerging PCgS") separately on accuracy (errors vs. correct) contrast statistical maps through the Statistical Parametric Mapping software (SPM Development Team 1991) on MATLAB R2018b (Inc. 2020). In the subsequent phase of our study, we implemented a masking procedure, wherein all values corresponding to the right hemisphere were set to zero. This step was undertaken with the objective of exclusively scrutinizing clusters situated within the left hemisphere. We constrained our region of interest for significant peaks to the left hemisphere, extending laterally only up to $x = -15$ in the MNI space. The posterior and inferior boundaries were defined based on coordinates obtained from a reference source BioImage Suite Web 2023, specifically for Brodmann area 24 (MNI coordinates: $y = -12, z = 5$).

The resulting level-2 t statistic images in each of the three groups were thresholded using the minimum given by a Bonferroni correction and random field theory to account for multiple comparisons (Worsley et al. 1996). Statistical significance for the group analyses was assessed based on peak thresholds in directed search and the spatial extent of consecutive voxels. For a single voxel in a directed search, involving all peaks within an estimated gray matter of 300 cm3 covered by the slices, the threshold for significance ($p< 0.05$) was set at $t = 5.71$ for "Prominent PCgS" group (n=14), at $t = 5.18$ for "Absent PCgS" group (n=18) and at $t = 5.39$ for "Emerging PCgS" group (n=16). A predicted cluster of voxels with a volume extent > 70.28 mm^3 (5 voxels) for "Prominent PCgS" group (n=14), 68.48 mm^3 (5 voxels) for "Absent PCgS" group (n=18), and 69.24 mm^3 (5 voxels) for "Emerging PCgS" group (n=16) with a t-value > 3 was significant ($p < 0.05$), corrected for multiple comparisons. All subsequently discussed peak activations were retrieved within the aforementioned delimited region of interest in spm results tables, all other activations presented for the group analyses within the significance thresholds but outside the area of interest are reported in Table SC.2 in the Supplementary Materials 4.6.

4.3 Results

		cluster size	*t*	*p*	MNI coordinates		
above the CgS (PCgS if present)	Prominent PCgS	91			x	y	z
			9.70	<.001	-8	14	53
			9.38	<.001	-2	22	50
			6.02	<.05	-2	-1	60
	Emerging PCgS	71					
			7.91	<.001	-2	22	50
			7.11	<.001	-2	9	60
	Absent PCgS	18					
			5.97	<.05	-2	24	50
CgS	Prominent PCgS			none			
	Emerging PCgS			none			
	Absent PCgS	2					
			5.35	<.05	-8	26	28
		5					
			5.75	<.05	-2	34	33

Table 4.1 – Summary of significant peaks identified in three groups categorized based on their spatial localization, either on the CgS or above it.

The proportion of errors in each morphological group was similar: the Prominent PCgS group produced errors on 13% of the trials, the Absent PCgS group on - 12% of the trials and the Emerging PCgS group on - 12% of the trials.

All three group-level analyses yielded significant clusters of activation within the ACC region. The summary of these findings is available in Table 4.1. The functional data was always inspected along with the corresponding structural anatomical images. Two vertical lines were drawn to locate peaks on the anterior-posterior axis: a vertical line through the posterior limit of the genu of the corpus callosum (VPG) and a line passing through the anterior commissure (VAC). For the superior-inferior axis, the main landmarks used were CgS, as well as PCgS for the PCgS-present population (Figure 4.4). The three patterns of activation exhibited a spatial overlap, albeit with quantitative differences that emerged gradually. The "Prominent PCgS" group displayed three significant peaks of activation within a substantial cluster measuring 1419.6 mm^3, situated over the PCgS and extending along the anterior-posterior axis from VAC to VPG. The lateral extent of these peaks ranged from $x = -2$ to $x = -8$ (MNI coordinates).

Conversely, the "Emerging PCgS" group also demonstrated two significant peaks of activation within the same region, specifically above the CgS. The anterior-posterior extent of this cluster mirrored that of the "Prominent PCgS" group, spanning from VAC to VPG. However, the lateral extent did not reach $x = -8$; both peaks were significant at $x = -2$ (MNI coordinates). Furthermore, the cluster size for the "Emerging PCgS" group was smaller, measuring 1107.6 mm^3.

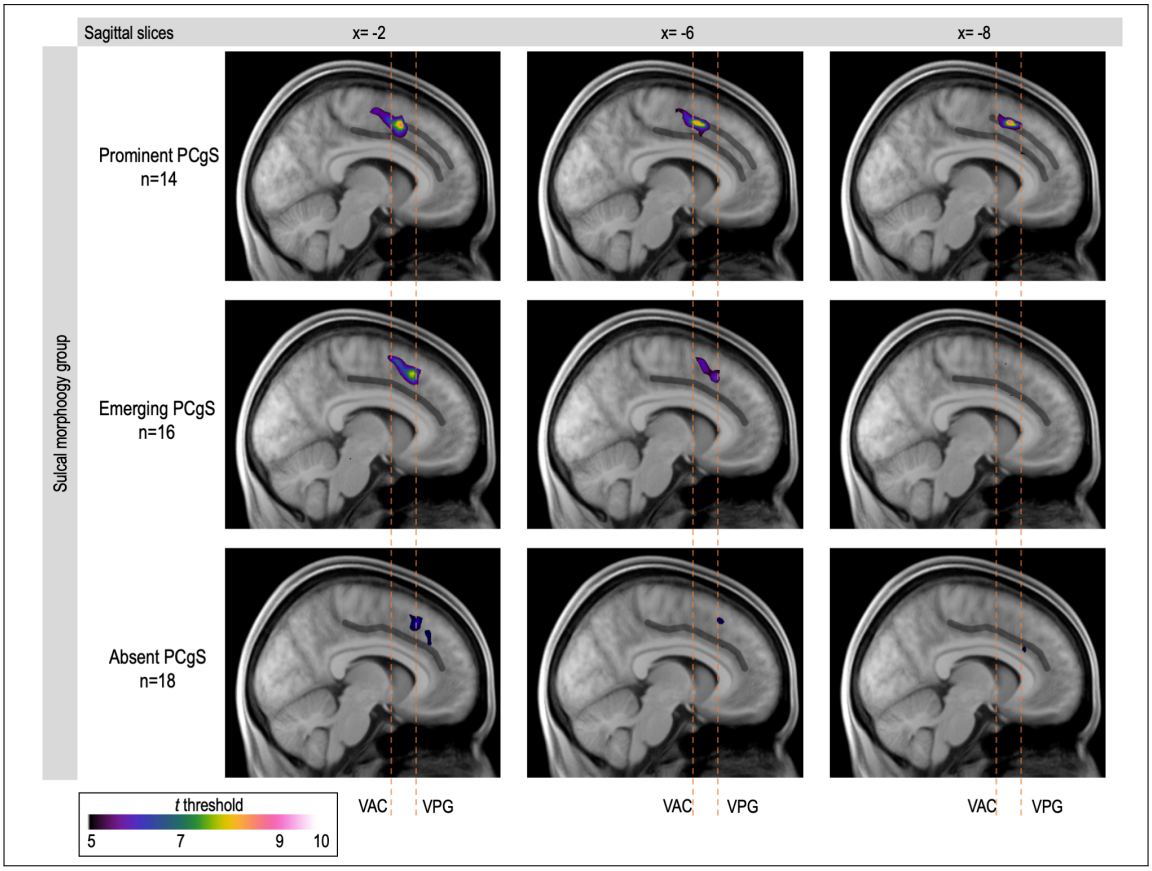


Figure 4.4 – Results for three groups of all right and left hemispheres mixed presented on mean anatomical scans of each group, the CgS and PCgS are drawn in gray lines. VAC refers to the vertical line drawn through anterior commissure, VPG refers to vertical line drawn through the posterior limit of the genu of the corpus callosum

Lastly, the "Absent PCgS" group exhibited a significant peak within a notably smaller cluster measuring 280.8 mm^3, also located above the CgS. However, both the anterior-posterior and lateral extents of this cluster were more limited than two other groups. Notably, the "Absent PCgS" group was distinct in that it showed two significant peaks within the CgS.

4.4 Discussion

The current study aimed to shed light on the neural mechanisms underlying speech error monitoring and its potential connection to vocal feedback processing, focusing on the role of anterior cingulate cortex (ACC) sulcal morphology. We examined BOLD activation corresponding to speech errors contrasted with correct utterances for each of 3 groups formed in function of the morphology of their ACC region : (1) "Prominent PCgS", (2) "Absent PCgS", (3)"Emerging PCgS". Our findings unveil distinct activation patterns within the ACC contingent upon the presence or absence of a PCgS. In individuals possessing a PCgS, error-related activations were predominantly localized along the PCgS, spanning from the VPG line to the VAC line. Conversely, individuals devoid of a PCgS also manifested activations in the region where the PCgS would typically be found in those with a PCgS; however, these activations were considerably smaller in magnitude and extent. Remarkably, the population without a PCgS displayed unique activations along the cingulate sulcus (CgS), a feature not observed in the other groups. The activation pattern observed in the group with an emerging PCgS aligns with this sulcal variation, influencing the localization of task-related activations and occupying an intermediary position between the "Prominent PCgS" and "Absent PCgS" groups. Importantly, it should be noted that the proportion of errors in each morphological group was similar, further supporting the idea that the observed differences in ACC activation patterns are indeed related to ACC sulcal morphology rather than variations in the frequency of speech errors.

Interestingly, the differential activation patterns observed within the CgS/PCgS regions prompt intriguing parallels with prior research by Loh et al. 2020. These authors observed similar modulations by sulcal morphology in the ACC of functional activations related to facial movement and feedback processing across various modalities. This was interpreted in terms of an adaptive feedback control framework. In such framework, our findings would indicate that feedback provided through the speech errors leads to ACC activation, signaling a need for adaptive control. However, the exact nature of this feedback provided through the speech errors remains an open question. The seemingly most obvious option would be that auditory feedback is provided through the speech errors. However, there is compelling evidence showing that a large share of speech errors are detected before articulation is initiated. For instance, when examining the latencies between the initiation of speech errors and their interruption, they exhibit a bimodal distribution: the interruption of an erroneous segment occurs either shortly after the onset of an error or approximately 500 milliseconds later (Nooteboom and Quené 2017). Moreover, several studies using EEG have shown a modulation of the error-related negativity (ERN) for speech errors (e.g. Ganushchak and Schiller 2008a; Riès et al. 2011; Baus et al. 2020). This ERP component has been

source localized to the ACC and peaks around 100 ms after the onset of a response, rendering an explanation in terms of auditory feedback implausible. Therefore, if the ACC activation elicited by speech errors is indeed feedback related, a more plausible source might be proprioception. Consistent with this interpretation, previous studies have shown the ACC is implicated in muscle-spindle feedback control (e.g., Goble et al. 2011). Proprioceptive feedback control would arguably also align with models that integrate somatosensory targets to guide speech production (e.g., Guenther and Hickok 2016; Hickok 2012; Dorokhova et al. 2023). A third possibility, integrating feedback control and conflict monitoring frameworks, would be to consider that conflict arising prior to error commission would generate an internal feedback-like signal within the ACC. This notion aligns with discussions in Botvinick et al. 2001's work proposing that conflict might constitute a component of a broader ACC system primarily focused on reallocating attentional resources. In this scenario, conflict stemming from the competition between alternatives, feedback indicating error commission, and the experience of pain (e.g., Jones et al. 1991), all of which have been shown to activate the ACC, collectively fall into the same category of cues signifying an insufficiency in the current allocation of attention resources to avert unfavorable outcomes. However, what sets conflict apart is its unique potential to proactively prevent impending errors, whereas pure feedback and pain occurrences are reactive responses to the realization of errors. Thus, within this framework, the emergence of conflict can indeed be interpreted as a type of feedback occurring during the planning stages to signal if attention must be reallocated.

Nonetheless, establishing a direct link between conflict and prior research that identified task-related activation shifts for feedback processing and motor areas presents a challenging endeavor. The activity reported in Loh et al. 2020 was primarily measured during the feedback stage, occurring at least 500 ms after the action. While the time delay may not be critically significant, it poses a challenge when interpreting this ACC activation in terms of conflict, which typically implies the presence of competing alternatives, but in this case after an action is completed. Furthermore, the activation of motor areas observed in Loh et al. 2018 presents even greater difficulty in aligning with the concept of conflict, particularly because there were no explicit constraints involved in the task. These challenges underscore the limitations of this proposition.

An alternative framework, as presented in studies such as Zarr and Brown 2016 and Alexander and Brown 2011, which centers on prediction errors activating the ACC, provides an equally suitable explanation for error processing in our study and the feedback-related activations found in the research by Loh et al. 2020. Under this framework, the ACC assumes the role of detecting discrepancies between observed outcomes and internally maintained rules, with the occurrence of a committed error representing a direct instance of such a mismatch, thereby prompting ACC activation. Furthermore,

Zarr and Brown 2016 suggests that ACC activation during feedback, rather than actual performance, is associated with the process of expectation adjustment and updating. A similar conceptual framework has emerged in studies where the observed alterations in paracingulate sulcus (PCgS) morphology have been scrutinized. These alterations, characterized by reduced gyrification and expanded grey matter volume, have consistently displayed a significant inverse relationship with an individual's capacity for reality monitoring and their aptitude for metacognitive introspection concerning performance outcomes (Buda et al. 2011; Garrison et al. 2015). This parallel becomes more apparent when considering the analogy to the concept of prediction error. Just as prediction error involves the assessment of disparities between anticipated outcomes and actual events, reality monitoring engages the process of distinguishing internally generated information from external stimuli. Both cognitive processes are fundamentally rooted in the discernment of incongruities between mental constructs and external reality. Moreover, Zarr and Brown 2016 delve into the concept of various prediction error signals within the ACC, highlighting their research findings that suggest a hierarchical organization extending from rostral to caudal regions. Multiple subregions within the ACC participate in the processing of prediction errors, including the dorsal anterior cingulate cortex (dACC) and the pre-supplementary motor area (pre-SMA) (Zarr and Brown 2016). Furthermore, the research proposes that distinct subregions of the ACC might specialize in the processing of specific categories of prediction errors. Zarr and Brown 2016 identifies three types of prediction errors: performance errors, high-level prediction errors, and low-level prediction errors. Performance errors occur when an individual makes a mistake due to an inadequate understanding of the current environment, while prediction errors occur when an outcome is different from what was expected. High-level prediction errors occur when an individual's expectations about the task or environment are violated, while low-level prediction errors occur when an individual's expectations about sensory input are violated. This proposition has the capacity to encompass all the distinct activations observed within the ACC, including the shifts in motor areas related to eye and tongue movements, feedback processing, and error commission. Additionally, this framework's ability to offer an explanation for all the previously mentioned activities indirectly reinforces the concept of continuity in the brain structures involved in mouth movements and their monitoring, extending to the supervision of language production. This implies that a more fundamental feedback-control circuitry, initially utilized for basic movements and vocalizations, has been repurposed to facilitate the optimization of language production.

In summary, our study's exploration of ACC sulcal morphology and its impact on error monitoring in language production adds a significant layer of complexity to our understanding of neural feedback mechanisms. These findings invite further investigation into the intricate interplay between ACC activation patterns, feedback processing,

and the detection of speech errors, promising new insights into the neural underpinnings of human language production. Furthermore, our findings add to a growing body of researching showing that sulcal morphology influences the functional organization of the human and non-human primate brain (e.g., Hopkins et al. 2017; Lopez-Persem et al. 2019; Leroy et al. 2015; Bodin et al. 2018; Eichert et al. 2021). They reveal a complex relationship between cingulate sulcal morphology and speech monitoring, offering implications beyond fundamental research. For instance, in neurosurgery, precise brain region localization is imperative. Our identification of potential anatomical landmarks within individual brains holds the potential of enabling surgeons to enhance procedural precision and minimize collateral damage for improved patient outcomes. Beyond clinical applications, these findings provide researchers with invaluable references for investigating the neural substrates of speech production and error monitoring, promising fundamental insights into the mechanisms governing language production.

4.5 References

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor [Number: 10 Publisher: Nature Publishing Group]. Nature Neuroscience, 14(10), 1338–1344. https://doi.org/10.1038/nn.2921

Amiez, C., & Petrides, M. (2014). Neuroimaging Evidence of the Anatomo-Functional Organization of the Human Cingulate Motor Areas. Cerebral Cortex, 24(3), 563–578. https://doi.org/10.1093/cercor/bhs329

Baus, C., Santesteban, M., Runnqvist, E., Strijkers, K., & Costa, A. (2020). Characterizing lexicalization and self-monitoring processes in bilingual speech production. Journal of Neurolinguistics, 56, 100934. https://doi.org/10.1016/j.jneuroling.2020.100934

BioImage Suite Web. (2023). BioImage Suite Web Atlas. https://bioimagesuiteweb.github.io/webapp/mni2tal.html

Bodin, C., Takerkart, S., Belin, P., & Coulon, O. (2018). Anatomo-functional correspondence in the superior temporal sulcus. Brain Structure & Function, 223(1), 221–232. https://doi.org/10.1007/s00429-017-1483-2

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Buda, M., Fornito, A., Bergström, Z. M., & Simons, J. S. (2011). A specific brain structural basis for individual differences in reality monitoring. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31(40), 14308–14313. https://doi.org/10.1523/JNEUROSCI.3595-11.2011

Dorokhova, L., Morillon, B., Baus, C., Belin, P., Dubarry, A.-S., Alario, F.-X., & Runnqvist, E. (2023). A robust temporal map of speech monitoring from planning to articulation. https://doi.org/10.31234/osf.io/acwqd

Eichert, N., Watkins, K. E., Mars, R. B., & Petrides, M. (2021). Morphological and functional variability in central and subcentral motor cortex of the human brain. Brain Structure & Function, 226(1), 263–279. https://doi.org/10.1007/s00429-020-02180-w

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

Garrison, J. R., Fernyhough, C., McCarthy-Jones, S., Haggard, M., & Simons, J. S. (2015). Paracingulate sulcus morphology is associated with hallucinations in the human brain [Number: 1 Publisher: Nature Publishing Group]. Nature Communications, 6(1), 8956. https://doi.org/10.1038/ncomms9956

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

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a New Model of Verbal Monitoring. Journal of Cognition, 3(1), 17. https://doi.org/10.5334/joc.81

Goble, D. J., Coxon, J. P., Impe, A. V., Geurts, M., Doumas, M., Wenderoth, N., & Swinnen, S. P. (2011). Brain Activity during Ankle Proprioceptive Stimulation Predicts Balance Performance in Young and Older Adults [Publisher: Society for Neuroscience Section: Articles]. Journal of Neuroscience, 31(45), 16344–16352. https://doi.org/10.1523/JNEUROSCI.4159-11.2011

Guenther, F. H., & Hickok, G. (2016). Neural Models of Motor Speech Control. In Neurobiology of Language (pp. 725–740). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00058-4

Hickok, G. (2012). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135–145. https://doi.org/10.1038/nrn3158

Hopkins, W. D., Coulon, O., Meguerditchian, A., Autrey, M., Davidek, K., Mahovetz, L., Pope, S., Mareno, M. C., & Schapiro, S. J. (2017). Genetic Factors and Orofacial Motor Learning Selectively Influence Variability in Central Sulcus Morphology in Chimpanzees (Pan troglodytes) [Publisher: Society for Neuroscience Section: Research Articles]. Journal of Neuroscience, 37(22), 5475–5483. https://doi.org/10.1523/JNEUROSCI.2641-16.2017

Inc., T. M. (2020). Matlab version: 9.8.0 (r2020a). Natick, Massachusetts, United States. https://www.mathworks.com

Jones, A. K., Brown, W. D., Friston, K. J., Qi, L. Y., & Frackowiak, R. S. (1991). Cortical and subcortical localization of response to pain in man using positron emission tomography. Proceedings. Biological Sciences, 244(1309), 39–44. https://doi.org/10.1098/rspb.1991.0048

Leroy, F., Cai, Q., Bogart, S. L., Dubois, J., Coulon, O., Monzalvo, K., Fischer, C., Glasel, H., Van der Haegen, L., Bénézit, A., Lin, C.-P., Kennedy, D. N., Ihara, A. S., Hertz-Pannier, L., Moutard, M.-L., Poupon, C., Brysbaert, M., Roberts, N., Hopkins, W. D., … Dehaene-Lambertz, G. (2015). New human-specific brain landmark: The depth asymmetry of superior temporal sulcus [Publisher: Proceedings of the National Academy of Sciences]. Proceedings of the National Academy of Sciences, 112(4), 1208–1213. https://doi.org/10.1073/pnas.1412389112

Loh, K. K., Hadj-Bouziane, F., Petrides, M., Procyk, E., & Amiez, C. (2018). Rostro-Caudal Organization of Connectivity between Cingulate Motor Areas and Lat-

eral Frontal Regions. Frontiers in Neuroscience, 11. Retrieved May 19, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2017.00753

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

Lopez-Persem, A., Verhagen, L., Amiez, C., Petrides, M., & Sallet, J. (2019). The Human Ventromedial Prefrontal Cortex: Sulcal Morphology and Its Influence on Functional Organization. The Journal of Neuroscience, 39(19), 3627–3639. https://doi.org/10.1523/JNEUROSCI.2060-18.2019

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

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

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

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), tgab038. https://doi.org/10.1093/texcom/tgab038

SPM Development Team. (1991). Statistical Parametric Mapping (SPM) (Version Version 2) [Software package for fMRI data analysis.]. http://www.fil.ion.ucl.ac.uk/spm/

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S., & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. Neurobiology of Language, 1–47. https://doi.org/10.1162/nol_a_00113

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connectivity, 2(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping, 4(1), 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

Zarr, N., & Brown, J. W. (2016). Hierarchical error representation in medial prefrontal cortex. NeuroImage, 124(Pt A), 238–247. https://doi.org/10.1016/j.neuroimage.2015.08.063

4.6 Supplementary materials

	Left Hemisphere	Right Hemisphere
Subject	PCgS aspect	PCgS aspect
1	**prominent**	absent
2	**prominent**	absent
3	*emerging*	absent
4	*emerging*	**prominent**
5	**prominent**	*emerging*
6	**prominent**	*emerging*
7	*emerging*	absent
8	absent	absent
9	*emerging*	absent
10	*emerging*	**prominent**
11	*emerging*	absent
13	*emerging*	*emerging*
14	*emerging*	**prominent**
16	absent	absent
17	absent	absent
19	absent	absent
20	*emerging*	absent
21	*emerging*	*emerging*
22	**prominent**	**prominent**
23	**prominent**	**prominent**
25	**prominent**	**prominent**
26	*emerging*	absent
27	**prominent**	*emerging*
28	absent	absent

Table SC.1 – Summarized presence/absence of PCgS in both hemispheres for each subject.

	cluster size	t	MNI coordinates		
			x	y	z
Prominent PCgS	242	**13.43**	**-32**	**26**	**-10**
		8.50	**-52**	**9**	**-2**
		7.30	-55	12	8
	13	**9.65**	**-5**	**-4**	**3**
	10	**7.59**	**-42**	**-4**	**53**
	2	5.92	-45	14	20
Emerging PCgS	13	**6.95**	**-40**	**9**	**26**
	15	6.58	-35	29	-2
	12	6.47	-10	-6	6
	6	6.01	-42	-1	48
	1	5.86	-15	-4	10
	1	5.62	-52	6	38
	1	5.44	-42	-38	46
	2	5.43	-42	24	-10
Absent PCgS	1	**6.74**	**-2**	**-14**	**-12**
	32	6.50	-38	26	3
		6.41	-32	22	10

Table SC.2 – Coordinates, *t* values and cluster sizes of peaks of activation in left hemisphere across three groups outside the area of main interest (ACC). Significant *t* values for exploratory search are displayed in bold text.

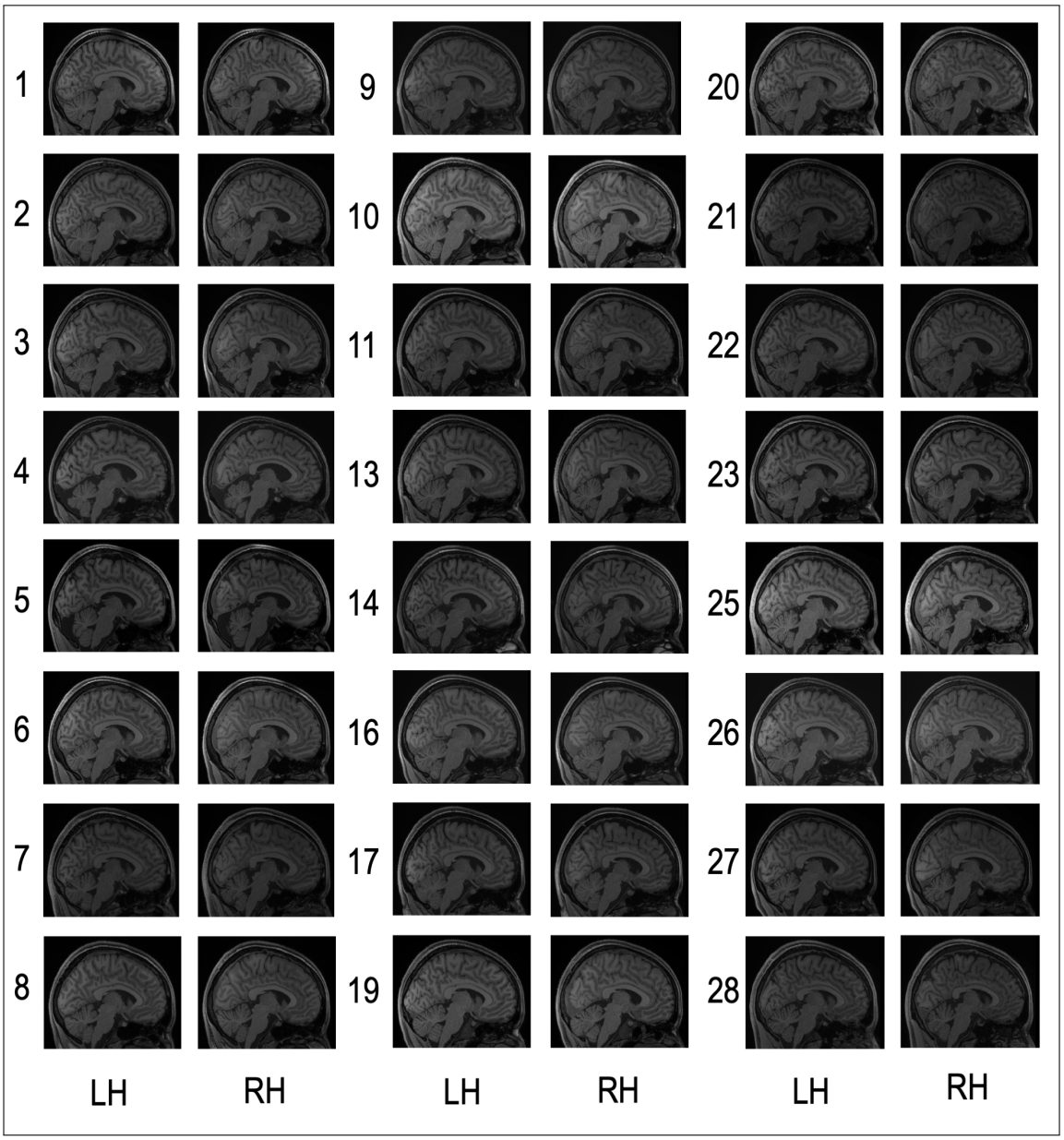


Figure SC.1 – Sagittal slices of right (RH) and left (LH) hemispheres of all analysed subjects at $x = 5$ and $x = -5$ respectively.

General Discussion

The primary goal of this thesis was to amass a substantial body of empirical evidence delving into various facets of self-monitoring in speech. Chapter 1 (Runnqvist et al. 2021) explored speech error monitoring using event-related fMRI, shedding light on cognitive mechanisms and their corresponding neural correlates. Three hypotheses concerning distinct neural correlates for comprehension-based, production-based, and integration-based monitoring were investigated. Additionally, error probability manipulation allowed for the examination of inner channel monitoring and overt errors provided insights into external channel monitoring. The investigation led to several key findings.

Firstly, the temporal cortex emerged as a crucial hub in the monitoring of overt speech errors. This observation supports the idea that the temporal cortex plays a pivotal role in comprehending and processing errors during speech production, especially in the context of comprehension-based monitoring.

Secondly, the medial frontal cortex (MFC), previously associated with conflict monitoring, exhibited differential brain responses in the context of external and internal speech monitoring. This finding suggests that the MFC's role may extend beyond conflict resolution, possibly involving feedback control. This aligns with a broader network for vocal cognitive control.

Lastly, the cerebellum's involvement in both internal and external monitoring was highlighted. Notably, the posterior cerebellum (Crus I) was linked to both forms of monitoring, while the superior medial cerebellum was specifically associated with overt errors. This distinction underscores the multifaceted role of the cerebellum in speech monitoring, potentially utilizing different subregions for distinct monitoring functions.

In summary, the study contributes substantial empirical support for the existence of shared and distinct cognitive mechanisms governing internal and external speech error monitoring. These findings emphasize the importance of investigating both facets of speech error monitoring to gain a comprehensive understanding of the complex cognitive and neural mechanisms underpinning speech production. This chapter opened up a series of questions that formed the empirical anchor point of the rest of the thesis. Concretely, the results called for an in-depth investigation on the temporal dynamics of the observed effects, and on how temporal and spatial signatures of monitoring might be modulated by the different level of representations engaged in a given task (more motor related or more cognitive in nature), both for language

production and other actions.

Modulations of Monitoring

Given the dissociation observed in the brain structures involved, which suggest differentiation in monitoring mechanisms, it became crucial to investigate the underlying reasons for this dissociation. Is it related to temporal factors, different levels of representation, or some other factors?

In the second chapter of this thesis (Chapter 2, Dorokhova et al. 2023), an investigation into the temporal dynamics of speech monitoring was undertaken. This study encompassed three critical stages in the speech production process: the initial phases of speech planning, speech motor preparation, and articulation. The findings revealed that monitoring processes intervene at distinct points during speech production, focusing on different types of information. In the early stages, the focus predominantly centers on response appropriateness and meaning. This is exemplified by the observed effect of lexical error probability, which predominantly influences the early pre-response stages. In contrast, as we progress to later stages, the emphasis shifts towards articulatory-phonetic conflicts.

Furthermore, the findings regarding monitoring timing reveal an association with the levels of representation. This observation comes from the observation that competition of alternatives in correct trials, particularly concerning response appropriateness or meaning, is resolved or significantly reduced by the time motor planning and articulation stages are reached. Similarly, when it comes to articulatory−phonetic competition of alternatives in correct trials, variations are observed solely during the speech motor preparation phase. The fact that the effect for distinct levels of representation was found in different timing windows, could imply that mechanisms at play are also different.

However, it's important to note that the EEG study itself does not have the capability to provide detailed insights into the mechanisms of operation for distinct levels of representation. To address this question, an ongoing fMRI study (Chapter A1) is being conducted, utilizing the contrast between levels of representation to shed light on this matter.

These findings align with the sequential nature of the speech production process, wherein meaning-related steps precede the encoding of sound. It is imperative to underscore that the outcomes of this study, which suggest a sequential processing pattern contingent on task relevance when activating linguistic representations for production, warrant a nuanced interpretation within the broader framework of language production models. In fact, a sequential dynamics of monitoring is even compatible with parallel models of language production. As posited by Fairs et al. 2021, there the processing dynamics of language production might unfold in two steps, wherein

initially, all dimensions of a word are subject to parallel activation, a phenomenon often referred to as 'ignition'. This parallel activation can be attributed to the holistic nature of linguistic representations. However, the intricacies of language production extend beyond this initial phase. Subsequent processes, such as selection and monitoring, appear to unfold sequentially during later stages known as 'reverberation'. This sequential nature suggests that, while parallel activation sets the stage for the linguistic entity, the subsequent stages of selection and monitoring may engage in a more step-wise manner.

Additionally, Chapter 2 encompassed a contrast between committed errors and avoided errors in the context of speech production. The results unveiled a collection of robust electrophysiological effects, emphasizing significant distinctions in the underlying monitoring mechanisms when comparing correct trials with error-laden instances. Specifically, a noteworthy observation was the occurrence of the Error-Related Negativity (ERN) phenomenon in overt errors. The presence of an ERN only on trials involving overt errors is also consistent with the findings of Runnqvist et al. 2021 where the commonly attributed neural locus for the ERN, the anterior cingulate cortex (ACC), was absent in the context of internal monitoring. Furthermore, it appears that partially distinct mechanisms are at work for the two monitoring channels: internal monitoring, which occurs before the error becomes perceptible, and external monitoring, which takes place while the error is overt.

Regarding the mechanism underpinning the effects observed in the internal monitoring channel, the collective evidence from our study and previous research indicates a role of predictive internal modeling. This would imply that it is the internal modeling process that generates error signals in a temporally distributed manner, in accordance with the demands of the specific encoding stage in progress. However, it's noteworthy that these error signals are likewise resolved or substantially diminished before the actual response is executed.

Overt errors, on the other hand, induce distinctions in processing across multiple stages, including both early and late speech motor preparation, as well as during articulation, implying a multi-layered error detection system. Differences in processing were observed, encompassing error detection both before the response and during the response itself. This observation suggests that when the error signal generated by predictive internal modeling fails to prevent an error, additional processes, likely reliant on feedback mechanisms, come into play during the articulation phase. In essence, the enduring influence of effects associated with overt errors suggests that errors can be detected at various points in time, potentially involving distinct processes. To be more precise, error detection may commence through internal modeling before the response, followed by proprioceptive feedback, and culminate in feedback-related mismatch detection.

In summary, various phases within the speech production process exhibited distinctive monitoring patterns, implying the presence of a precisely orchestrated temporal hierarchy attuned to the specific cognitive demands inherent at each representation level. Moreover, the empirical evidence robustly supports the role of predictive internal modeling as a central element in the internal monitoring process, while external monitoring mechanisms appear to operate through a diverse array of feedback modalities, as indicated by the findings.

Hierarchical Buildup of Internal Modeling in Motor Actions

The third study (Chapter 3) in our research aimed to investigate the hierarchical buildup of internal modeling during motor actions with varying cognitive demands. We were interested in understanding how the brain's internal modeling network adapts when individuals engage in motor tasks that require different levels of cognitive engagement. To explore this, we designed an experiment where we manipulated the level of cognitive demand in a basic motor task involving different body parts. Participants were tasked with either executing a straightforward movement or performing a movement that involved monitoring, as dictated by an additional rule applied to its execution. This experimental design allowed us to examine how the brain responds to the increasing monitoring load imposed by the task variations.

Frontal and parietal cortex regions demonstrated activation patterns that align with hierarchical gradients, revealing distinct levels of cognitive control during motor tasks.

Within the frontal cortex, our findings revealed a hierarchical gradient that corresponds to the increasing cognitive complexity. Specifically, during monitored movements, we observed increased activation in prefrontal associative regions, notably Brodmann Areas 45, 9, and 10. This contrasted with the predominant activation of the more unimodal regions BA6 and BA4 during simple movement-alone conditions. Monitored movement also engaged medial frontal structures, such as the Anterior Cingulate Cortex (ACC) and the Pre-Supplementary Motor Area (Pre-SMA). These observations align with prior research, underscoring the ACC's pivotal role in self-monitoring across a spectrum of actions. Furthermore, our study shed light on the synergistic relationship between lateral and medial frontal cortex regions, highlighting their collective involvement in monitoring processes and adaptive behavior.

Shifting our focus to the parietal cortex, our investigation revealed a noteworthy shift in activation patterns corresponding to varying levels of cognitive complexity. As anticipated based on existing functional gradients, the cognitively more demanding monitored movements led to increased activation in Brodmann Area 40 (BA40), located caudally. In contrast, the performance of simple movements in isolation primarily activated the somatosensory cortex. This parietal shift further emphasizes the role of

cognitive engagement in influencing the brain's response to motor tasks.

The findings of this study regarding the cerebellum reveal a noteworthy departure from previously documented patterns in the literature. Typically, lobules I-IV in the cerebellum have been associated with basic motor functions. However, this study observed their activation in conditions characterized by increased cognitive complexity. This observation could suggest that the cognitive demands of the task have a notable impact on the motor control process, implicating the involvement of these cerebellar regions in tasks that require a combination of cognitive and motor coordination. Further, Crus I, typically associated with congitive processing was observed in simple movement condition. In light of the hierarchical effects observed in the frontal and parietal cortices, but the absence of such effects in the cerebellum, these findings imply an autonomous organization within cerebral and cerebellar regions that underpin both movement and more intricate cognitive functions. This contradicts earlier findings that advocated for a parallel processing structure encompassing the prefrontal cortex and the posterolateral cerebellum. The cerebellum's role in integrating cognitive and motor functions may have broader implications for our understanding of how the brain manages complex actions, potentially transcending the boundaries of motor control and influencing various aspects of human behavior and cognition. These findings offer valuable insights into the broader mechanisms governing action control. They hint at the possibility that the interplay between cognitive and motor processes extends beyond traditional motor tasks and may have relevance in other domains of complex cognitive-motor actions, such as language production.

The fourth study (Chapter 4) delved into the realm of monitoring within the broader context of action control mechanisms, with a specific focus on analyzing the sulcal morphology of the anterior cingulate cortex (ACC). This investigation involved the annotation of individual anatomical brain scans, followed by a group-level analysis of task-related activations, considering the specific anatomical features of each participant. The study uncovered intriguing relationships between sulcal variations in the ACC and distinct activation patterns triggered by speech errors.

Given that the ACC was involved in processing feedback (Loh et al. 2020, Runnqvist et al. 2021), our study delved into a thorough examination of the sources contributing to its activation. We investigated several potential factors that triggered feedback processing, with a particular emphasis on proprioception and internal modeling. These elements emerged as primary candidates for explaining the observed ACC activation patterns. These elements were discussed as possible drivers behind the observed ACC activation patterns.

Moreover, our study integrated insights gleaned from a body of prior research that specifically focused on prediction errors and reality monitoring. This process of synthesizing existing knowledge allowed us to put forth a comprehensive and unified

theoretical framework, which seeks to elucidate the intricate neurocognitive mechanisms responsible for error processing and the regulation of feedback control within the neural substrates of the human brain. Through this rigorous amalgamation of diverse research strands, our study aimed to provide a profound and scientifically grounded understanding of the multifaceted cognitive processes that underlie error processing and feedback regulation within the ACC.

Future research endeavors

One central question that remains open pertains to the brain substrates involved in monitoring different levels of representation. A study currently in progress, detailed in Chapter A1, manipulated speech error probability at the semantic and articulatory-phonetic levels and, as such,holds promise to shed light on this matter and provide further evidence. The forthcoming analyses encompass Regions of Interest (ROI) and connectivity analyses.

In the context of ROI analysis, two approaches will be used. First, we will utilize theoretically predefined ROIs, as demonstrated in previous studies such as Runnqvist et al. 2021 and Todorović et al. 2023. These pre-established ROIs have proven effective in uncovering specific neural substrates relevant to our research. In addition, we will use the movement task that was carried out with the same participants (specified in Chapter 3) as subject-specific localizer. This approach holds promise in customizing ROIs to the nuanced demands of our research context. Notably, the overlap observed between motor tongue areas and feedback processing regions, as discussed in Chapter 3, provides a unique opportunity for ROI delineation. Furthermore, the monitored movement condition within the task offers valuable insights into the act of monitoring itself, making it a valuable resource for our inquiries.

Regarding connectivity analysis, we aim to link connectivity effects to errors on the manipulated levels of representation or to the magnitude of the effect for monitoring at different representation levels. This approach offers a nuanced understanding of how neural networks adapt and interact in response to monitoring behavior. To this end, we collected resting-state data before and after the task.

Furthermore, future investigations should delve deeper into the role of monitoring as an overarching cognitive ability. The task introduced in Chapter 3 provides a unique perspective on monitoring, and similar paradigms could be employed to unveil motor and cognitive brain signatures across various tasks. Moreover, a more comprehensive understanding of the interplay between cortical structures and the cerebellum is warranted, given the unexpected pattern observed in Chapter 3. A promising future direction for research could involve exploring variations in motor and cognitive loads applied to simple motor tasks involving different body parts. This approach

aims to distinguish the effects of motor and cognitive involvement on cortical and cerebellar regions. By systematically manipulating the demands placed on motor and cognitive systems during motor tasks, we can gain a more nuanced understanding of how these brain regions adapt and interact in response to varying cognitive and motor demands. This research has the potential to shed light on the interplay between motor and cognitive processes within the brain, contributing to a deeper comprehension of their functional organization and dynamics. Additionally, the intriguing lateralization observed in monitored tongue movement, which was absent in tongue movement alone and inverted for hand movement, presents avenues for further investigation. Subsequent research can fine-tune task design based on these observations. One approach is to equalize the cognitive demands across different body parts. For instance, hand movements could involve continuous 180° rotations both clockwise and counterclockwise, while tongue movements could mimic up-and-down motions akin to hand movements.

Furthermore, a promising avenue for future research involves building upon the groundwork laid out in Chapter 4. In that study, we delved into individual morphological factors and their potential correlation with brain function. To deepen our understanding, future investigations could encompass individual annotations of functional data, similar to anatomical scans, enabling a more comprehensive exploration of the intricate interplay between brain structure and function. To begin with, we aim to perform similar analyses as done in Chapter 4 using the data from Chapter 3 with contrast involving movement versus monitored movement, as well as Supplementary Chapter A1 with the contrast of speech errors. Moreover, correct utterances within different representation levels could also undergo the procedure to examine whether the ACC activation loci would differ depending of the level of linguistic representation. And further, more meticulous approach can be adopted by examining individual functional activation coordinates put in relation to the individual anatomy. This relatively novel approach holds great promise for advancing our understanding of the complex processes that unfold within the brain. By establishing a unique connection between individual anatomical characteristics of brain structure and the corresponding functional activations that underpin specific behaviors, this approach offers a pathway to uncovering previously uncharted territories of cognitive research.

Synthesis

The empirical inquiries encapsulated within the present thesis collectively unveil a rich and complex landscape within the realm of monitoring and feedback control in the human brain, offering valuable insights into its multifaceted nature during a spectrum of cognitive and motor tasks. These findings transcend the notion that monitoring processes are rigid and predetermined; instead, they illustrate their remarkable adaptability, which hinges on factors such as the specific task at hand, the level of representation involved, and the cognitive demands imposed.

One of the key threads weaving through these investigations is the pivotal role of internal modeling, prediction errors, and feedback mechanisms. These cognitive elements emerge as fundamental pillars underpinning the brain's ability to scrutinize and detect errors, regardless of whether the context is language production or motor actions. The studies have effectively illuminated how these mechanisms orchestrate the brain's response to discrepancies between expected and observed outcomes.

A striking revelation from this collective body of research is the hierarchical organization of brain regions involved in monitoring and feedback control. The interplay between medial and lateral frontal cortex regions, along with the parietal cortex, appears to provide a common framework for understanding the brain's capacity to monitor, regulate, and adapt behavior across diverse contexts. These hierarchical gradients, extending from primary motor regions to higher-order cognitive centers, offer a glimpse into the choreography of cognitive control processes.

As we consider the future of research in this domain, an exciting avenue emerges for further exploration – the dynamic interplay between monitoring processes and the broader mechanisms governing action control. Unraveling the neural underpinnings of how these two realms interact promises to deepen our understanding of human cognition and behavior.

References

Alario, F., Chainay, H., Lehéricy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production [Publisher: Elsevier]. Brain Research, 1076(1), 129–143. https://doi.org/10.1016/j.brainres.2005.11.104

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor [Number: 10 Publisher: Nature Publishing Group]. Nature Neuroscience, 14(10), 1338–1344. https://doi.org/10.1038/nn.2921

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex [Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . .]. Neural Computation, 27(11), 2354–2410. Retrieved September 29, 2023, from https://direct.mit.edu/neco/article-abstract/27/11/2354/8116

Amiez, C., & Petrides, M. (2014). Neuroimaging Evidence of the Anatomo-Functional Organization of the Human Cingulate Motor Areas. Cerebral Cortex, 24(3), 563–578. https://doi.org/10.1093/cercor/bhs329

Amiez, C., & Petrides, M. (2018). Functional rostro-caudal gradient in the human posterior lateral frontal cortex. Brain Structure & Function, 223(3), 1487–1499. https://doi.org/10.1007/s00429-017-1567-z

Avants, B., Epstein, C., Grossman, M., & Gee, J. (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

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. Journal of Verbal Learning and Verbal Behavior, 14(4), 382–391. https://doi.org/10.1016/S0022-5371(75)80017-X

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends in Cognitive Sciences, 12(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? Nature Reviews Neuroscience, 10(9), 659–669. https://doi.org/10.1038/nrn2667

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. Trends in Cognitive Sciences, 22(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2021). Matrix: sparse and dense matrix classes and methods [R package version 1.3-4]. https://CRAN.R-project.org/package=Matrix

Baus, C., Santesteban, M., Runnqvist, E., Strijkers, K., & Costa, A. (2020). Characterizing lexicalization and self-monitoring processes in bilingual speech production. Journal of Neurolinguistics, 56, 100934. https://doi.org/10.1016/j.jneuroling.2020.100934

Behzadi, Y., Restom, K., Liau, 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

Binkofski, F., & Buccino, G. (2006). The role of ventral premotor cortex in action execution and action understanding. Journal of Physiology, Paris, 99(4-6), 396–405. https://doi.org/10.1016/j.jphysparis.2006.03.005

BioImage Suite Web. (2023). BioImage Suite Web Atlas. https://bioimagesuiteweb.github.io/webapp/mni2tal.html

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

Bodin, C., Takerkart, S., Belin, P., & Coulon, O. (2018). Anatomo-functional correspondence in the superior temporal sulcus. Brain Structure & Function, 223(1), 221–232. https://doi.org/10.1007/s00429-017-1483-2

Borich, M. R., Brodie, S. M., Gray, W. A., Ionta, S., & Boyd, L. A. (2015). Understanding the role of the primary somatosensory cortex: Opportunities for rehabilitation. Neuropsychologia, 79(Pt B), 246–255. https://doi.org/10.1016/j.neuropsychologia.2015.07.007

Boschin, E. A., Mars, R. B., & Buckley, M. J. (2017). Transcranial magnetic stimulation to dorsolateral prefrontal cortex affects conflict-induced behavioural adaptation in a Wisconsin Card Sorting Test analogue. Neuropsychologia, 94, 36–43. https://doi.org/10.1016/j.neuropsychologia.2016.11.015

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Botvinick, M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature, 402(6758), 179–181. https://doi.org/10.1038/46035

Buda, M., Fornito, A., Bergström, Z. M., & Simons, J. S. (2011). A specific brain structural basis for individual differences in reality monitoring. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31(40), 14308–14313. https://doi.org/10.1523/JNEUROSCI.3595-11.2011

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. Nature Human Behaviour. https://doi.org/10.1038/s41562-022-01516-2

Chambers, J., Hastie, T., & Pregibon, D. (1990). Statistical Models in S. In K. Momirović & V. Mildner (Eds.), Compstat (pp. 317–321). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-50096-1_48

Choi, E. Y., Drayna, G. K., & Badre, D. (2018). Evidence for a Functional Hierarchy of Association Networks. Journal of Cognitive Neuroscience, 30(5), 722–736. https://doi.org/10.1162/jocn_a_01229

Cisek, P., & Kalaska, J. F. (2004). Neural correlates of mental rehearsal in dorsal premotor cortex. Nature, 431(7011), 993–996. https://doi.org/10.1038/nature03005

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

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: i. segmentation and surface reconstruction. NeuroImage, 9(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Davare, M., Andres, M., Cosnard, G., Thonnard, J.-L., & Olivier, E. (2006). Dissociating the role of ventral and dorsal premotor cortex in precision grasping. The Journal of Neuroscience, 26(8), 2260–2268. https://doi.org/10.1523/JNEUROSCI.3386-05.2006

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a Neural System for Error Detection and Compensation [_eprint: https://doi.org/10.1111/j.1467-9280.1994.tb00630.x]. Psychological Science, 5(5), 303–305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production [Place: US Publisher: American Psychological Association]. Psychological Review, 93, 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist Models of Language Production: Lexical Access and Grammatical Encoding. Cognitive Science, 23(4), 517–542. https://doi.org/10.1207/s15516709cog2304_6

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. NeuroImage, 33(1), 127–138. https://doi.org/10.1016/j.neuroimage.2006.05.056

Diedrichsen Lab. (2023). Diedrichsen Lab Atlas Viewer. https://www.diedrichsenlab.org/imaging/AtlasViewer/viewer.html

D'Mello, A. M., Gabrieli, J. D. E., & Nee, D. E. (2020). Evidence for Hierarchical Cognitive Control in the Human Cerebellum. Current biology: CB, 30(10), 1881–1892.e3. https://doi.org/10.1016/j.cub.2020.03.028

Dorokhova, L., Morillon, B., Baus, C., Belin, P., Dubarry, A.-S., Alario, F.-X., & Runnqvist, E. (2023). A robust temporal map of speech monitoring from planning to articulation. https://doi.org/10.31234/osf.io/acwqd

Eichert, N., Watkins, K. E., Mars, R. B., & Petrides, M. (2021). Morphological and functional variability in central and subcentral motor cortex of the human brain. Brain Structure & Function, 226(1), 263–279. https://doi.org/10.1007/s00429-020-02180-w

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018a). Fmriprep. Software. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018b). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods. https://doi.org/10.1038/s41592-018-0235-4

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), tgab040. https://doi.org/10.1093/texcom/tgab040

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalography and Clinical Neurophysiology, 78(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Fiehler, K., Ullsperger, M., & Von Cramon, D. Y. (2005). Electrophysiological correlates of error correction [Place: United Kingdom Publisher: Blackwell Publishing]. Psychophysiology, 42(1), 72–82. https://doi.org/10.1111/j.1469-8986.2005.00265.x

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, Supplement 1, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning? Cognition, 176, 107–130. https://doi.org/10.1016/j.cognition.2018.03.009

Friston, K. (2011). What Is Optimal about Motor Control? Neuron, 72(3), 488–498. https://doi.org/10.1016/j.neuron.2011.10.018

Ganushchak, L. Y., & Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring: An ERP study. Brain Research, 1125(1), 104–115. https://doi.org/10.1016/j.brainres.2006.09.096

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

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

Garrison, J. R., Fernyhough, C., McCarthy-Jones, S., Haggard, M., & Simons, J. S. (2015). Paracingulate sulcus morphology is associated with hallucinations in the human brain [Number: 1 Publisher: Nature Publishing Group]. Nature Communications, 6(1), 8956. https://doi.org/10.1038/ncomms9956

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

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a New Model of Verbal Monitoring. Journal of Cognition, 3(1), 17. https://doi.org/10.5334/joc.81

Goble, D. J., Coxon, J. P., Impe, A. V., Geurts, M., Doumas, M., Wenderoth, N., & Swinnen, S. P. (2011). Brain Activity during Ankle Proprioceptive Stimulation Predicts Balance Performance in Young and Older Adults [Publisher: Society for Neuroscience Section: Articles]. Journal of Neuroscience, 31(45), 16344–16352. https://doi.org/10.1523/JNEUROSCI.4159-11.2011

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, 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

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. Software. https://doi.org/10.5281/zenodo.596855

Grafton, S., & Hamilton, A. (2007). Evidence for a Distributed Hierarchy of Action Representation in the Brain. Human movement science, 26, 590–616. https://doi.org/10.1016/j.humov.2007.05.009

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

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. The Journal of Neuroscience, 37(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Mohr, B., & Pulvermüller, F. (2019). Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. NeuroImage, 199, 206–216. https://doi.org/10.1016/j.neuroimage.2019.05.071

Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum [Publisher: eLife Sciences Publications Limited]. Elife, 7, e36652.

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

Guenther, F. H., & Hickok, G. (2016). Neural Models of Motor Speech Control. In Neurobiology of Language (pp. 725–740). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00058-4

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

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

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). Journal of Memory and Language, 52(1), 58–70. https://doi.org/10.1016/j.jml.2004.07.006

Hartsuiker, R. J., & Kolk, H. H. J. (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

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., & Amunts, K. (2007). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. Human Brain Mapping, 30(2), 392–402. https://doi.org/10.1002/hbm.20512

Hickok, G. (2012). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135–145. https://doi.org/10.1038/nrn3158

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

Hopkins, W. D., Coulon, O., Meguerditchian, A., Autrey, M., Davidek, K., Mahovetz, L., Pope, S., Mareno, M. C., & Schapiro, S. J. (2017). Genetic Factors and Orofacial Motor Learning Selectively Influence Variability in Central Sulcus Morphology in Chimpanzees (Pan troglodytes) [Publisher: Society for Neuroscience Section: Research Articles]. Journal of Neuroscience, 37(22), 5475–5483. https://doi.org/10.1523/JNEUROSCI.2641-16.2017

Hothorn, T., Bretz, F., & Westfall, P. (2021). multcomp: simultaneous inference in general parametric models [R package version 1.4-25]. https://CRAN.R-project.org/package=multcomp

Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-Scale Gradients in Human Cortical Organization. Trends in Cognitive Sciences, 22(1), 21–31. https://doi.org/10.1016/j.tics.2017.11.002

Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: Implications for higher-order cognitive functions [Place: Germany Publisher: Springer]. Psychological Research, 73(4), 527–544. https://doi.org/10.1007/s00426-009-0235-1

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T., & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool [Number: 6766 Publisher: Nature Publishing Group]. Nature, 403(6766), 192–195. https://doi.org/10.1038/35003194

Inc., T. M. (2020). Matlab version: 9.8.0 (r2020a). Natick, Massachusetts, United States. https://www.mathworks.com

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. Frontiers in Psychology, 2. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. Cognition, 92(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

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

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

Jones, A. K., Brown, W. D., Friston, K. J., Qi, L. Y., & Frackowiak, R. S. (1991). Cortical and subcortical localization of response to pain in man using positron emission tomography. Proceedings. Biological Sciences, 244(1309), 39–44. https://doi.org/10.1098/rspb.1991.0048

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends in Cognitive Sciences, 18(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. PLOS Computational Biology, 13(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2021). lmerTest: tests in linear mixed effects models [R package version 3.1-4]. https://CRAN.R-project.org/package=lmerTest

Kwon, S. E., Yang, H., Minamisawa, G., & O'Connor, D. H. (2016). Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. Nature Neuroscience, 19(9), 1243–1249. https://doi.org/10.1038/nn.4356

Lamb, S. M., & Newell, L. E. (1966). Outline of stratificational grammar. Georgetown University Press.

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

Leroy, F., Cai, Q., Bogart, S. L., Dubois, J., Coulon, O., Monzalvo, K., Fischer, C., Glasel, H., Van der Haegen, L., Bénézit, A., Lin, C.-P., Kennedy, D. N., Ihara, A. S., Hertz-Pannier, L., Moutard, M.-L., Poupon, C., Brysbaert, M., Roberts, N., Hopkins, W. D., … Dehaene-Lambertz, G. (2015). New human-specific

brain landmark: The depth asymmetry of superior temporal sulcus [Publisher: Proceedings of the National Academy of Sciences]. Proceedings of the National Academy of Sciences, 112(4), 1208–1213. https://doi.org/10.1073/pnas.1412389112

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. Cognition, 14(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). Speaking: from intention to articulation. The MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. Cognition, 42(1-3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. BEHAVIORAL AND BRAIN SCIENCES, 76.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' Acceptance of Real-Time Speech Exchange Indicates That We Use Auditory Feedback to Specify the Meaning of What We Say. Psychological Science, 25(6), 1198–1205. https://doi.org/10.1177/0956797614529797

Lockwood, D. G. (1972). Introduction to stratificational linguistics. Harcourt Brace Jovanovich.

Logan, G. D., & Crump, M. J. C. (2010). Cognitive Illusions of Authorship Reveal Hierarchical Error Detection in Skilled Typists [Publisher: American Association for the Advancement of Science]. Science, 330(6004), 683–686. https://doi.org/10.1126/science.1190483

Loh, K. K., Hadj-Bouziane, F., Petrides, M., Procyk, E., & Amiez, C. (2018). Rostro-Caudal Organization of Connectivity between Cingulate Motor Areas and Lateral Frontal Regions. Frontiers in Neuroscience, 11. Retrieved May 19, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2017.00753

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

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. Frontiers in Human Neuroscience, 8. Retrieved May 8, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00213

Lopez-Persem, A., Verhagen, L., Amiez, C., Petrides, M., & Sallet, J. (2019). The Human Ventromedial Prefrontal Cortex: Sulcal Morphology and Its Influence on Functional Organization. The Journal of Neuroscience, 39(19), 3627–3639. https://doi.org/10.1523/JNEUROSCI.2060-18.2019

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. Neuroreport, 12(9), 1851–1855. https://doi.org/10.1097/00001756-200107030-00018

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging. NeuroImage, 31(4), 1453–1474. https://doi.org/10.1016/j.neuroimage.2006.02.004

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 [Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . .]. Journal of Cognitive Neuroscience, 33(3), 422–444. Retrieved September 25, 2023, from https://direct.mit.edu/jocn/article-abstract/33/3/422/95552

Mesulam, M. M. (1998). From sensation to cognition. Brain: A Journal of Neurology, 121 (Pt 6), 1013–1052. https://doi.org/10.1093/brain/121.6.1013

Miall, R. C., & Wolpert, D. M. (1996). Forward Models for Physiological Motor Control. Neural Networks, 9(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Moller, J., Jansma, B. M., Rodriguez-Fornells, A., & Munte, T. F. (2006). What the Brain Does before the Tongue Slips. Cerebral Cortex, 17(5), 1173–1178. https://doi.org/10.1093/cercor/bhl028

Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. Cognition, 4(2), 177–187. https://doi.org/10.1016/0010-0277(76)90003-2

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. The Journal of Neuroscience, 33(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Nooteboom, S. G., & Quené, H. (2008). Self-monitoring and feedback : A new attempt to find the main cause of lexical bias in phonological speech errors [Accepted: 2017-06-09T19:35:25Z]. Journal of Memory and Language, 58, 837–861. Retrieved September 8, 2021, from http://localhost/handle/1874/350522

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

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

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1844), 20200531. https://doi.org/10.1098/rstb.2020.0531

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. Cerebral Cortex (New York, N.Y.: 1991), 6(3), 342–353. https://doi.org/10.1093/cercor/6.3.342

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

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. Journal of Speech and Hearing Research, 35(3), 537–544. https://doi.org/10.1044/jshr.3503.537

Postma, A., & Noordanus, C. (1996). Production and Detection of Speech Errors in Silent, Mouthed, Noise-Masked, and Normal Auditory Feedback Speech [Publisher: SAGE Publications Ltd]. Language and Speech, 39(4), 375–392. https://doi.org/10.1177/002383099603900403

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behavior Research Methods, 39(4), 859–862. https://doi.org/10.3758/BF03192979

R Core Team. (2022). R: a language and environment for statistical computing [Version 4.2.2, released on 2022-10-31]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving

speech motor control. Neurology, 64(4), 700–706. https://doi.org/10.1212/01.WNL.0000152156.90779.89

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

Riès, S., Nadalet, L., Mickelsen, S., Mott, M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2020). Pre-output Language Monitoring in Sign Production. Journal of Cognitive Neuroscience, 32(6), 1079–1091. https://doi.org/10.1162/jocn_a_01542

Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. Current Opinion in Neurobiology, 12(2), 149–154. https://doi.org/10.1016/s0959-4388(02)00308-2

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. Cognition, 64(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2020). Self-Monitoring in Speaking: In Defense of a Comprehension-Based Account [Number: 1 Publisher: Ubiquity Press]. Journal of Cognition, 3(1), 18. https://doi.org/10.5334/joc.61

Roland, P. E. (1984). Organization of motor control by the normal human brain. Human Neurobiology, 2(4), 205–216.

Runnqvist, E. (2023). Self-Monitoring: The Neurocognitive Basis of Error Monitoring in Language Production [Num Pages: 23]. In Language Production. Routledge.

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

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), tgab038. https://doi.org/10.1093/texcom/tgab038

Saadon-Grosman, N., Arzy, S., & Loewenstein, Y. (2020). Hierarchical cortical gradients in somatosensory processing. NeuroImage, 222, 117257. https://doi.org/10.1016/j.neuroimage.2020.117257

Sanes, J. N., & Donoghue, J. P. (2000). Plasticity and primary motor cortex. Annual Review of Neuroscience, 23, 393–415. https://doi.org/10.1146/annurev.neuro.23.1.393

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. The Journal of the Acoustical Society of America, 98(5), 2428–2442. https://doi.org/10.1121/1.413277

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime User's Guide.

Severens, E., Janssens, I., Kühn, S., Brass, M., & Hartsuiker, R. J. (2011). When the brain tames the tongue: Covert editing of inappropriate language [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2011.01190.x]. Psychophysiology, 48(9), 1252–1257. https://doi.org/10.1111/j.1469-8986.2011.01190.x

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

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

SPM Development Team. (1991). Statistical Parametric Mapping (SPM) (Version Version 2) [Software package for fMRI data analysis.]. http://www.fil.ion.ucl.ac.uk/spm/

Strijkers, K. (2016). A Neural Assembly–Based View on Word Production: The Bilingual Test Case [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lang.12191]. Language Learning, 66(S2), 92–131. https://doi.org/10.1111/lang.12191

Szirtes, J., & Vaughan, H. (1977). Characteristics of cranial and facial potentials associated with speech production. Electroencephalography and Clinical Neurophysiology, 43(3), 386–396. https://doi.org/10.1016/0013-4694(77)90261-9

Teghipco, A., Okada, K., Murphy, E., & Hickok, G. (2023). Predictive Coding and Internal Error Correction in Speech Production. Neurobiology of Language (Cambridge, Mass.) 4(1), 81–119. https://doi.org/10.1162/nol_a_00088

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers in Psychology, 1, 166. https://doi.org/10.3389/fpsyg.2010.00166

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S., & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. Neurobiology of Language, 1–47. https://doi.org/10.1162/nol_a_00113

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

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. NeuroImage, 39(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. https://doi.org/10.1038/nature01710

Tunik, E., Rice, N., Hamilton, A., & Grafton, S. (2007). Beyond grasping: Representation of action in human anterior intraparietal sulcus. NeuroImage, 36 Suppl 2, T77–86. https://doi.org/10.1016/j.neuroimage.2007.03.026

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

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

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connectivity, 2(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Wickham, H. (2016). ggplot2: elegant graphics for data analysis [R package version 3.4.3]. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: a grammar of data manipulation [R package version 1.0.7]. https://CRAN.R-project.org/package=dplyr

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. Current Biology, 11(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration [Publisher: American Association for the Advancement of Science]. Science, 269(5232), 1880–1882. Retrieved September 8, 2021, from http://www.jstor.org/stable/2889276

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping, 4(1), 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

Zarr, N., & Brown, J. W. (2016). Hierarchical error representation in medial prefrontal cortex. NeuroImage, 124(Pt A), 238–247. https://doi.org/10.1016/j.neuroimage.2015.08.063

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

Additional Content

A1 The brain basis of speech monitoring at different levels of representation

Dorokhova Lydia
Anton Jean-Luc
Nazarian Bruno
Sein Julien
Peirolo Morgane
Belin Pascal
Runnqvist Elin

Outline

Abstract

In an event-related fMRI study we investigated the neural bases underpinning the optimizing of language production and explored how these might change as a function (a) of the level of language at which potential errors originate, or (b) of whether or not an error is probable but ultimately avoided or probable and overtly committed. 20 participants were scanned 4 times on separate weeks while performing a speech production task designed to elicit speech errors. For correct productions we contrasted trials with high versus low error probability of a semantic source, or an articulatory phonetic source. We also contrasted overtly committed errors with correct trials.Results of an initial whole brain analysis showedthat all three contrasts resulted in significant premotor (BA6) activation. Moreover, the semantic error probability contrast was associated with differential activations in left inferior frontal gyrus. The articulatory-phonetic error contrast was associated with differential activations in the left inferior temporal gyrus. Finally, overt errors compared to correct trials engaged inferior frontal gyrus bilaterally, as well as medial frontal BA 8, and several cerebellar regions. These results are to be completed with several complementary analyses, such as hypotheses driven ROI analyses using both theoretically predefined coordinates and functional subject specific localizers. However, already at this initial stage the results suggest that both representation specific and aspecific activations can be identified in structures related to monitoring and executive control, potentially reflecting the existence of multiple underlying monitoring mechanisms.

Key words: language production; error-probability; speech monitoring; semantics; articulatory-phonetics; fMRI

A1.1 Introduction

Speech production is a complex process that comprises multiple stages, including conceptualization, lemma retrieval, and encoding. These stages collectively work to process crucial information, ensuring our speech is coherent and error-free. Errors in speech production are exceedingly rare. While there's a consensus on the presence of self-monitoring processes, both internal, as demonstrated by the speed of self-repairs (Levelt 1983; Hartsuiker and Kolk 2001), and external, as shown in studies involving altered feedback (Tourville et al. 2008; Niziolek and Guenther 2013), several intricacies of this process remain unresolved.

However, despite substantial progress in understanding self-monitoring mechanisms, numerous intricacies within this process remain unresolved. One pivotal question revolves around the interplay of these mechanisms and their potential relationship with linguistic levels ranging from cognitive processing to articulation. In a recent study by Runnqvist et al. 2021, a comprehensive examination of theories pertaining to speech monitoring shed light on this matter. This study proposed the involvement of different brain structures, including the temporal cortex, medial frontal cortex, and cerebellum, linked to various monitoring theories. These theories include the comprehension-based theory, conflict-based theory, and forward modeling theory, each providing unique insights into the complex monitoring process during speech production.

Nonetheless, a significant knowledge gap persists regarding how monitoring operates at various levels of linguistic representation. Errors in speech production can manifest at distinct stages, such as the mispronunciation of 'cat' as 'gat,' indicating impaired articulatory control, or the substitution of 'cat' with 'dog,' signifying a semantic error and hinting at issues within higher-level processes like lexical retrieval. This suggests that the observed in Runnqvist et al. 2021 monitoring mechanisms may be specialized for specific representations within the broader speech production process.

To address these unresolved questions, we conducted a functional magnetic resonance imaging (fMRI) study. Our investigation aimed to uncover the neural underpinnings involved in optimizing language production and explore potential variations in these mechanisms based on (a) the level at which errors originate within the language production process and (b) whether errors are probable yet avoided or probable and overtly committed. We recruited 20 participants who underwent fMRI scanning on four separate occasions, each while engaged in a speech production task designed to elicit speech errors. Our analyses included contrasting trials with high versus low error probability stemming from either a semantic source or an articulatory phonetic source. Additionally, we examined overtly committed errors and their neural signatures in comparison to correct trials.

A1.2 Methods

A1.2.1 Participants

The study obtained ethical approval from the regional ethical committee Comité de Protection des Personnes Sud Méditerranée I, with the approval filed under identification number 2017-A03614-49. A total of twenty right-handed young adults (12 women) participated in the study, with age ranging from 20 to 33 years (mean age = 25.5 years). Participants were remunerated for their involvement. None of the participants reported a history of language or neurological disorders.

A1.2.2 Stimuli

The target stimuli consisted of 112 printed French nouns, forming 56 pairs. Altering the initial consonants of words within a pair always resulted in a lexical outcome, leading to new pairs (e.g. *barque manque* ⇒ *marque banque*). Manipulation of consonant onset articulatory-phonetic distance was employed: half of the pairs shared two features (such as place, manner of articulation, or voicing; e.g., [m] and [b] both being labial and voiced), while the other half lacked shared features. Three phonological primes were used to prime the pairs, with the order of initial consonants inverted (e.g., ***m**ode **b**louse, **m**oiche **b**aille, **m**otte **b**aie* for the target pair ***b**arque **m**anque*). In half of the target pairs, the primed error outcome of the first word was additionally primed with a synonym placed after the phonological primes (e.g., *trace pose* for *barque manque*, where *trace* is a synonym for *marque*). Participants encountered all pairs, yet only half were semantically primed, while the rest contained morphologically similar fillers in lieu of semantic primes. This distribution was reversed across two distinct lists, equally dispatched in between participants (10 participants per list) with each participants consistently encountering the same list. Filler word pairs were interspersed between each bundle of prime pairs up to the target pairs. Among these fillers, 53% were presented for silent reading, while the remaining fillers were accompanied by a question mark and subsequently exclamation marks. This design aimed to reduce the predictability of the task sequence.

A1.2.3 Procedure

Prior to commencing the fMRI experiment, participants were introduced to the task and underwent training outside the scanner. This training aimed to reduce head movements during speech and enhance response speed. Then, participants were positioned in the fMRI scanner in front of the screen, with a head-mounted microphone. The instruction of the task was recalled. Pairs of words were displayed on the screen

for a duration of 693 ms, succeeded by a question mark that remained visible for 539 ms. Subsequently, three exclamation marks (!!!) persisted on the screen for 1001 ms. Participants were instructed to silently read the word pairs and promptly vocalize the last pair upon the appearance of the question mark. The inclusion of exclamation marks following the question mark served to further expedite participants. The question mark was succeeded by a blank screen, with the duration jittered between 462 ms and 1232 ms. All words and symbols were presented in white on a black background. The illustration of the procedure is given in Figure A1.1. In a single experiment session, the list was presented to participants two times with different order, partitioned into four runs, each lasting for 5 minutes. Participants were encouraged to take a brief break before starting the subsequent run. Each participant engaged in 4 experiment sessions (16 runs), with a span of one week between each session. In total, one participant was presented with the 56 target pair list 8 times (twice per each out of 4 sessions).

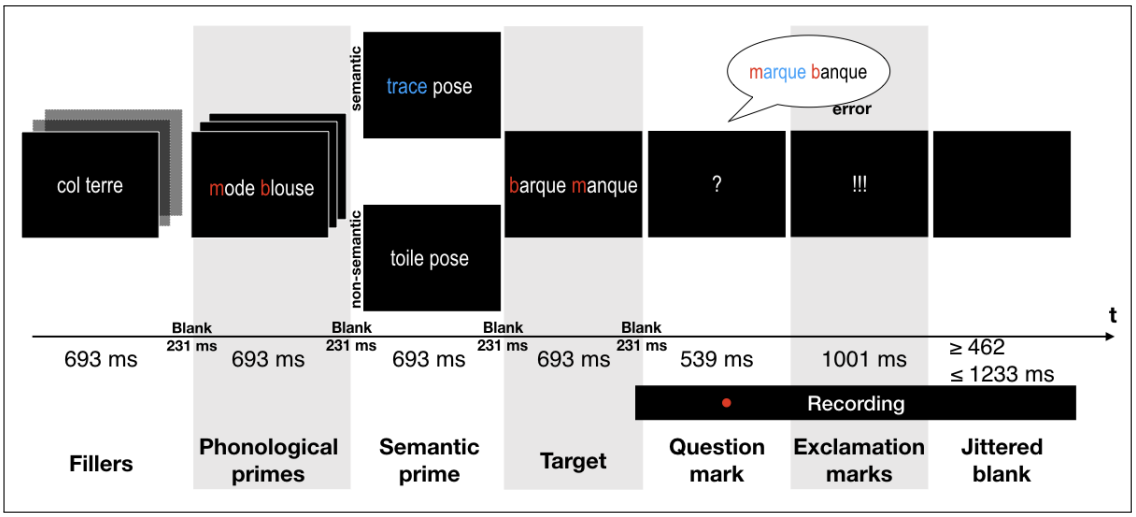


Figure A1.1 – Task Design. Color fonts have been utilized for descriptive purposes, with red representing articulatory-phonetic manipulation and clear blue indicating semantic manipulation. All words were presented to participants in white.

A1.2.4 MRI Data Acquisition

A 3-Tesla Siemens Prisma Scanner (Siemens, Erlangen, Germany) was used to collect data using a 64-channel head coil at the Marseille MRI Center (Centre IRM-INT@CERIMED, UMR7289 CNRS & AMU). Functional BOLD images (EPI sequence, 72 slices per volume, multi-band = 4, repetition time = 1.386 s, spatial resolution = 2 × 2 × 2 mm^3, echo time = 33.4 ms, flip angle = 56°) covering the whole brain were acquired during the task performance. Four runs were acquired using a BOLD protocol containing 155 images. Whole-brain anatomical MRI data were acquired using high-resolution structural T1 - weighted image (MPRAGE sequence, repetition time = 2.3 s, spatial resolution = 0.8 × 0.8 × 0.8 mm^3, echo time = 3.1 ms, inversion time = 0.9 s flip angle = 9°) in the sagittal

plane. Prior to functional imaging, whole-brain Fieldmap images were acquired with a spin-echo EPI sequence with the same spatial parameters as the BOLD images and acquired twice with opposite phase encode directions along the anterior-posterior axis with the following parameters: TR/TE = 7220/59 ms, voxel size = 2 x 2 x 2 mm^3, slices = 72, flip angle = 90/180°.

A1.2.5 Image processing and analyses

The data was preprocessed using fMRIPrep 20.2.6 (Esteban et al. 2018b; Esteban et al. 2018a; RRID:SCR_016216), which is based on Nipype 1.7.0 (K. Gorgolewski et al. 2011; K. J. Gorgolewski et al. 2018; RRID:SCR_002502).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al. 2010, distributed with ANTs 2.3.3 Avants et al. 2008, RRID:SCR_004757), 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, RRID:SCR_002823, Zhang et al. 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al. 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) 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, RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym),

Functional data preprocessing

For each of the 4 BOLD runs found per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band references (SBRefs). A B0-nonuniformity map (or fieldmap) was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` Cox and Hyde 1997 (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) 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 `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with nine degrees of freedom 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) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al. 2002). First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. 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. Two global signals are extracted within the CSF and the WM. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the anatomical denoising approach (aCompCor: two probabilistic masks (CSF and WM) are generated in anatomical space. Components are calculated separately within the WM and CSF masks. For each CompCor decomposition, the 12 first components with the largest singular values are retained. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. 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` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964).

Subsequently, the data was analyzed using the Statistical Parametric Mapping software (SPM Development Team 1991) on MATLAB (Inc. 2020).The preprocessed BOLD data were smoothed with an isotropic three-dimensional Gaussian kernel (FWHM = 5 mm).For each run and each subject, a general linear model (GLM) was designed with 8 conditions: "semantically related distance 1", "semantically related distance 3"; "semantically unrelated distance 1"; "semantically unrelated distance 3"; "error"; "silence"; "prime" and "filler". The regressors were convolved by canonical hemodynamic response. The default SPM autoregressive model AR(1) and high-pass filter (128s) were applied. 48 nuisance regressors were also included in the model : 24 head movement derived and 24 aCompCor (12 CSF and 12 WM) and the two global signals within CSF and WM masks. To isolate the effect of *overt errors*, the "error" condition

was contrasted with the "semantically related distance 1", "semantically related distance 3"; "semantically unrelated distance 1"; and "semantically unrelated distance 3" conditions. To isolate *semantically driven error probability* we contrasted "semantically related distance 1" and "semantically related distance 3" with "semantically unrelated distance 1" and "semantically unrelated distance 3". To isolate *articulatory-phonetic driven error probability* we contrasted "semantically related distance 3" and "semantically unrelated distance 3" with "semantically related distance 1" and "semantically unrelated distance 1". These first level contrast images were then entered in a second level whole brain analysis. The resulting statistical parametrical map was thresholded at $p < 0.001$ (voxel level, uncorrected), and a minimum cluster size of 25 voxels was used. Effects were reported as significant at p.<05 after a False Discovery Rate (FDR) correction.

A1.3 Results

The results of the whole brain analysis suggest a partial dissociation in brain structures sustaining semantic and articulatory-phonetic driven error probability. The clusters of activation for the contrasts of interest are reporrted in Tables A1.1; A1.2; A1.3. The articulatory phonetic contrast resulted in a significant cluster in premotor cortex (right BA6) and right inferior temporal gyrus (ITG).

BA	*q*FDR-corr	*k*E	T	x	y	z
R 6	0.005	79	6.88	44	0	48
L 7	0.171	26	6.76	-32	-48	60
R ITG	0.029	52	5.61	48	-62	-12
R 37	0.171	27	5.18	32	-50	-20
R 5/precuneus	0.147	33	5.05	12	-42	56
L CrusI CB	0.171	27	4.68	-40	-50	-30

Table A1.1 – High articulatory-phonetic error probability (close versus distant consonant onsets). Significant clusters of activity at uncorrected p<.001

As for semantically driven error probability, it showed a cluster in the premotor cortex (right BA6), as well as a cluster in the left inferior frontal gyrus (BA44).

BA	*q*FDR-corr	*k*E	T	x	y	z
L 44	0.000	234	5.63	-44	16	8
R 6	0.049	37	5.38	32	0	56
R 7	0.131	25	4.84	32	-58	60

Table A1.2 – High semantic error probability (semantically related versus unrelated error priming). Significant clusters of activity at uncorrected p<.001

Finally, the contrast isolating overt errors resulted in a left lateralized cluster in premotor cortex (BA6), as well as a series of regions in inferior frontal gyrus (right 44, bilateral BA45, left 47), a medial frontal cluster (BA8) and three clusters in the right cerebellum (lobules VI and VIII).

BA	*q*FDR-corr	*k*E	T	x	y	z
L 47	0.000	773	7.45	-40	26	-6
L 6	0.000	782	7.44	-48	4	50
R 45	0.000	471	7.39	46	20	2
L 8	0.000	1235	6.66	-2	20	56
R Lobule VI CB	0.000	175	6.27	30	-72	-26
R 44/9	0.000	79	5.63	52	16	28
R Lobule 8 CB	0.001	70	5.43	26	-74	-50
R Lobule 8 CB	0.034	31	5.03	36	-58	-54
L 45	0.000	79	4.99	-54	22	10

Table A1.3 – Overt errors versus correct productions. Significant clusters of activity at uncorrected p<.001

A1.4 Future prospects

A1.4.1 Region of Interest Analysis

As a next step of this study we will perform two Region of Interest (ROI) analyses, a crucial step to assess our hypotheses and to allow for comparability with the previous literature (e.g., Runnqvist et al., 2021; Todorovic et al., 2023). To define these critical ROIs, we will use two complementary approaches. Firstly, we will use established, theoretically predefined ROIs, a methodology well-applied in prior studies like Runnqvist et al. 2021 and Todorović et al. 2023. These pre-established regions have consistently demonstrated their efficacy in unveiling specific neural substrates relevant to our research.

In parallel, we are embarking on an exploration of a subject-specific localizer approach, building upon the task introduced in Chapter 3. This approach holds the potential to define ROIs customized to the specific nuances of our research context. Two compelling reasons underpin this endeavor:

Firstly, a noteworthy overlap was revealed between motor tongue areas and feedback processing regions (Loh et al. 2020), as elucidated in Chapter 4. Given the pivotal role of feedback processing in our study, this overlap offers a unique vantage point for the delineation of ROIs. Secondly, the monitored movement condition within the task present a significant opportunity to identify regions that are part of a broader action monitoring network, comprising both motor and more cognitive actions..

A1.4.2 Resting-State Functional Connectivity

In the present study, resting-state (RS) data was collected before and after the SLIP task in three out of four sessions (3*10 minutes pre-task, 3*10 minutes post-task). To explore task related functional connectivity, we will first calculate connectivity maps using the same ROIs as seeeds as those specified for the SLIP task (e.g., anatomically predefined and subject specific movement monitoring task localizer) in a ROI to whole brain analysis. Next, we will compare the RS data collected before and after the SLIP task introducing several behavioral measures as covariates: global error rates, semantic relatedness effect on errors and phonetic distance effect on errors; as well as response latencies on semantically related and phonetically close correct trials.

A1.5 References

Avants, B., Epstein, C., Grossman, M., & Gee, J. (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

Behzadi, Y., Restom, K., Liau, 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

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

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: i. segmentation and surface reconstruction. NeuroImage, 9(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018a). Fmriprep. Software. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018b). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature Methods. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, Supplement 1, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, 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

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. Software. https://doi.org/10.5281/zenodo.596855

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

Hartsuiker, R. J., & Kolk, H. H. J. (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

Inc., T. M. (2020). Matlab version: 9.8.0 (r2020a). Natick, Massachusetts, United States. https://www.mathworks.com

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

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. PLOS Computational Biology, 13(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

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

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. Cognition, 14(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

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

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. The Journal of Neuroscience, 33(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

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), tgab038. https://doi.org/10.1093/texcom/tgab038

SPM Development Team. (1991). Statistical Parametric Mapping (SPM) (Version Version 2) [Software package for fMRI data analysis.]. http://www.fil.ion.ucl.ac.uk/spm/

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S., & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. Neurobiology of Language, 1–47. https://doi.org/10.1162/nol_a_00113

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. NeuroImage, 39(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

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

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

A2 Behavioral pilots

Prior to perform neuroimaging studies, four behavioral pilots were tested to validate the efficiency of the manipulations. All the pilots were using the SLIP protocol.

A2.1 Pilot 1 - Lexicality + Articulatory-Phonetc Manipulation

A2.1.1 Stimuli

120 pairs of words for the SLIP task in French were created. Half of them resulted in a lexical outcome if the initial consonants were exchanged (e.g., *balade sonnet* ⇒ *salade bonnet*, stroll sonnet ⇒ salad bonnet), and half of them resulted in non-words (*ballet sapeur* ⇒ *sallet bapeur* ballet sapper ⇒ non-words). Furthermore, the lexical manipulation was crossed with an articulatory-phonetic (AP) manipulation. Initial letters either shared two or zero shared features (SF) among manner, place and phonation. Half of the stimuli were distant and half close, equally dispatched in between lexical and non-lexical condition. Syllabical structure was also controlled: half of the stimuli pairs contained monosyllabic words, the other half was made of bisyllabic words. However, this effect was not exploited.

A2.1.2 Participants

Participants were 20 native French speakers aged between 19 and 27 y.o. (mean 21; 13 female). They all received monetary compensation in a form of a gift card of 10€.One participant was excluded from the analyses because their response times were too slow and their utterances did not fit into the recording window.

A2.1.3 Procedure

The experiment is consisted of pairs of words appearing on the screen. The order of the trial was as follows: fillers (randomly between one and three, stayed for 700 ms), three phonological (AP) primes (700 ms each) and a target. Target remained on the screen for 700 ms and then was followed by a blank screen (200 ms), question mark

(500 ms) and an exclamation mark (1000 ms) to hurry the participant up. With the appearance of the question mark the audio recording started and lasted 2000 ms (see Figure A2.1). The participants were given the instruction to read aloud the last pair that they saw on the screen when question mark appeared. They were also asked to respond as quickly as possible. To randomize the task some fillers were followed by a question mark too, but they were not recorded. Participants were seated in front of the computer screen with a headband microphone. The task lasted for 35 minutes with a pause each 12 minutes. 714 items were repeated twice, fillers were randomized for each repetition.

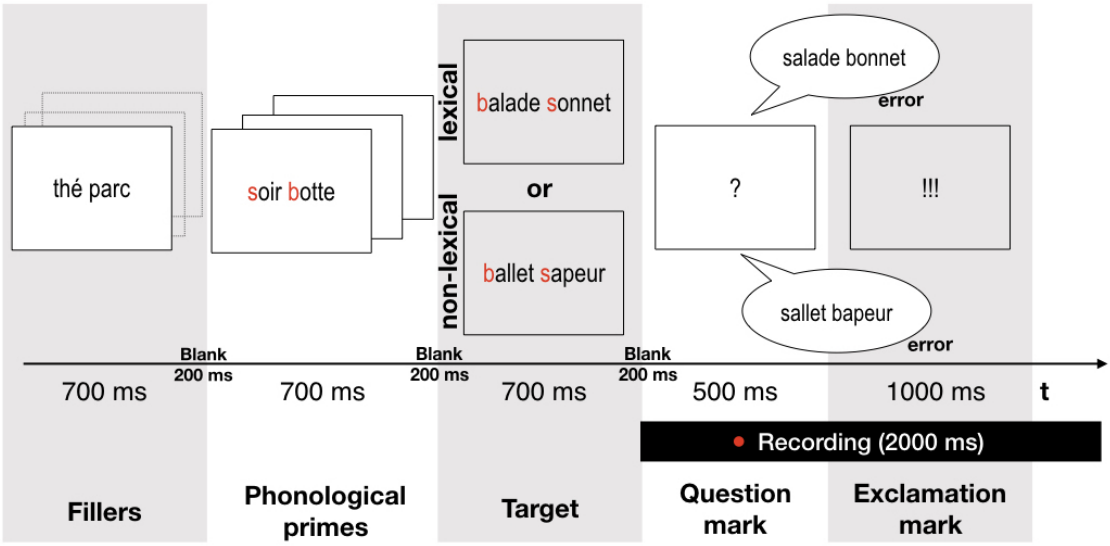


Figure A2.1 – Pilot 1 experimental design and procedure. Red text indicate manipulated initials in the AP condition.

A2.1.4 Data

Data of 19 participants initially presented 4560 trials, wherein each of the 120 pairs was repeated twice, resulting in 240 trials per participant. The lexicality and AP conditions were evenly distributed in halves of the total trial count (2280 per condition). After filtering the full omissions, the dataset comported the total number of 4418 exploitable trials (lexical: 2209, non-lexical: 2209, close AP: 2198, distant AP: 2220).

The data analysis was conducted using the RStudio R Core Team 2022 and key analytical tasks were performed utilizing specific packages including `'Matrix'` (version 1.5-1, Bates et al. 2021), `'lme4'` (version 1.1-34, Bates et al. 2015), `'lmerTest'` (version 3.1-3, Kuznetsova et al. 2021), `'multcomp'` (version 1.4-25, Hothorn et al. 2021), `dplyr` (version 1.1.2, Wickham et al. 2021), `ggplot2` (version 3.4.3, Wickham 2016), each of which facilitated critical statistical and visualization procedures. For the analysis of errors, we employed a mixed-effects logistic regression model using the `glmer` function of `lme4` package in R Core Team 2022.

```
glmer(RelatedErrors ~ Lexicality + AP + Lexicality * AP +
(1+Lexicality|Subject)+(1|Item), family = binomial)
```

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (Chambers et al. 1990 via R documentation) prior to conducting the statistical analysis in R Core Team 2022.

A2.1.5 Results

Participants made priming-related errors in 223 trials (5.05%). More erroneous production trials were observed in a 'close AP' condition (144 trials out of 2198, 6.5%) in comparison to 'distant AP' (79 trials out of 2220, 3.6%). As well as more errors were committed for lexical outcome (166 out of 2209, 7.5%) comparing to non-lexical outcome condition (57 out of 2209, 2.6%). The effects of lexicality and articulatory-phonetic condition were significant (z = 4.07, p<.001; z=-2.9, p<.01 respectively) on priming-related errors without interaction of the variables (z=-.005, p=.9).

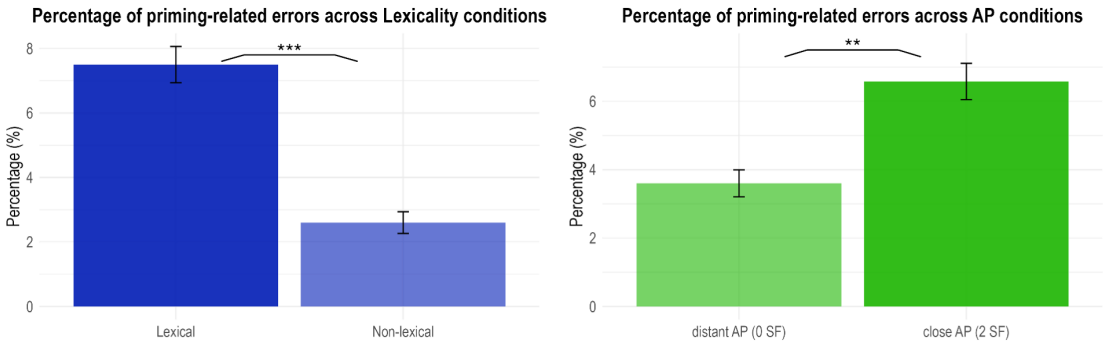


Figure A2.2 – Percentages of priming related errors by Lexicality and AP conditions. Each portion is displayed with respect to the number of trials within the condition (i.e., the potential maximum number of errors). SF stands for shared features.

The success of this manipulation in the present behavioral pilot led to its application in the EEG experiment detailed in Chapter 2 (Dorokhova et al. 2023).

A2.2 Pilot 2 - Semantic + Articulatory-Phonetc Manipulation

The SLIP task can also be adjusted to induce semantic priming, as for example in Motley and Baars 1976. Aside phonological priming there were semantic primes related to the lexical outcome of the pair (e.g., ill bishop, striken priest, target: pick soap ⇒ sick pope). In the original design, Motley and Baars 1976 primed the target pair as a noun group with synonymic noun groups. Unfortunately, in this case it is not possible to manipulate phonetic distance. Thus, I simplified the original design to keep the AP manipulation.

A2.2.1 Stimuli

There were 56 pairs of words created for this experiment. As in the previous experiment the AP distance of the consonant onsets of the word pairs was manipulated: half of the stimuli shared two features and the other half zero. For half of the target word pairs, the primed error outcome of the first word of each pair was additionally primed with a synonym that was placed after the AP primes (see Figure A2.4). Participants saw all the pairs, but only a half was semantically primed. The other half contained morphologically similar fillers on the place of semantic prime (neutral priming). These two conditions were inverted in two lists, participant saw only one list.

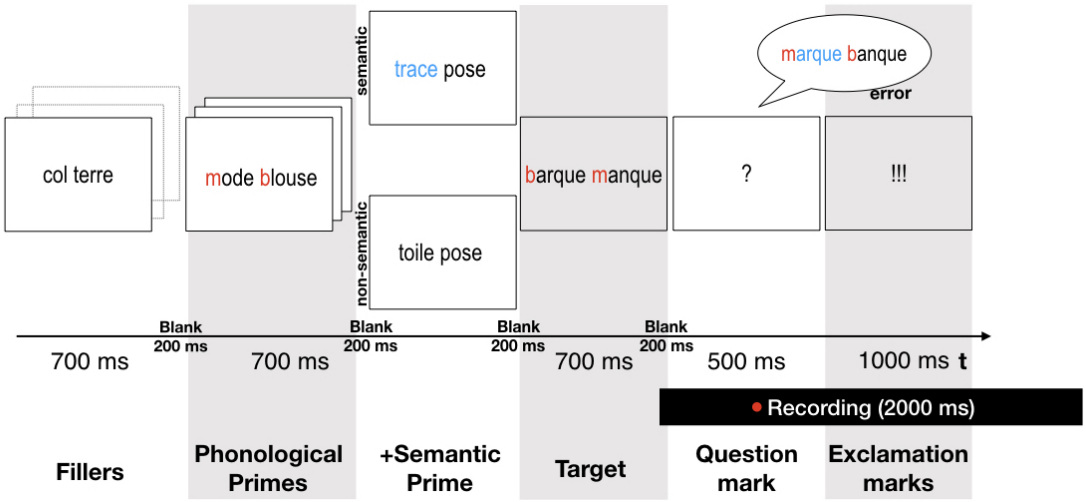


Figure A2.3 – Pilot 1 experimental design and procedure. Red text indicate manipulated initials in the AP condition, while light blue text indicate synonymical relation in between items : *trace* (trace) is a synomym of the word *marque* (mark).

A2.2.2 Participants

Participants were 36 native French speakers aged between 19 and 30 y.o. (mean 24 y.o.; 25 female). They all received monetary compensation in a form of a gift card of 10€. 19 participants passed the experiment under laboratory condtion modality, while 17 participants passed the same experiment via FindingFive platform (FindingFive Corporation, a 501(c)(3) nonprofit organization) online.

A2.2.3 Procedure

The procedure was similar to the first experiment, but instead of the third phonological prime there was a semantically related prime that preceded the target and stayed on the screen for 700 ms. The total number of 408 items were repeated 4 times, fillers were randomized for each repetition.

A2.2.4 Data

Data of 36 participants initially presented 8064 trials, wherein each of the 56 pairs was repeated 4 times, resulting in 224 target trials per participant. Semantic and AP conditions were evenly distributed in halves of the total trial count (4032 per condition). After filtering the full omissions, the dataset comported the total number of 7978 exploitable trials (semantic: 3986, neutral: 3992, close AP: 3986, distant AP: 3992). For the analysis of errors, we employed a mixed-effects logistic regression model using the `glmer` function of `lme4` package in R Core Team 2022. The modality (online/offline) was included as a fixed effect and an interaction term. However, the interaction of modality was tested separately for semantic manipulation and AP manipulation due to the model convergence failure when including both in one.

```
glmer(RelatedErrors ~ Semantic + AP Distance + Modality
Semantic * AP Distance + Semantic * Modality +(1+Semantic |Subject)
+(1|Item), family = binomial)

and

glmer(RelatedErrors ~ Semantic + AP Distance + Modality
Semantic * AP Distance + AP Distance * Modality +(1+Semantic |Subject)
+(1|Item), family = binomial)
```

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (Chambers et al. 1990 via R documentation) prior to conducting the statistical analysis in R Core Team 2022.

A2.2.5 Results

Participants made errors related to the priming manipulation in 271 trials 2,58% of the total number. Error percentages across both groups were similar: online group: 116 errors out of 3752 trials, 3.09%; offline group: 155 errors out of 4226, 3.7%). More errors were made for word pairs whose initial consonants were AP close (197 trials out of 3986, 4.9%) compared to those that were distant(74 trials out of 3992, 1.9%). More errors were made following the semantic priming condition (148 out of 3986, 3.7%) than after the neutral priming (123 out of 3992, 3%).

Both effects of semantic and articulatory-phonetic priming were significant ($z = -2.08, p < .05$; $z = -2.8, p < .01$ respectively) on priming-related errors with interaction of the variables ($z = -2.3, p < .05$). The effect of modality was not significant ($z = -0.364, p = .71575$) and did not interact with the other variables (AP: $z = .534, p = .59$; semantic: $z = 1.560, p = .11$).

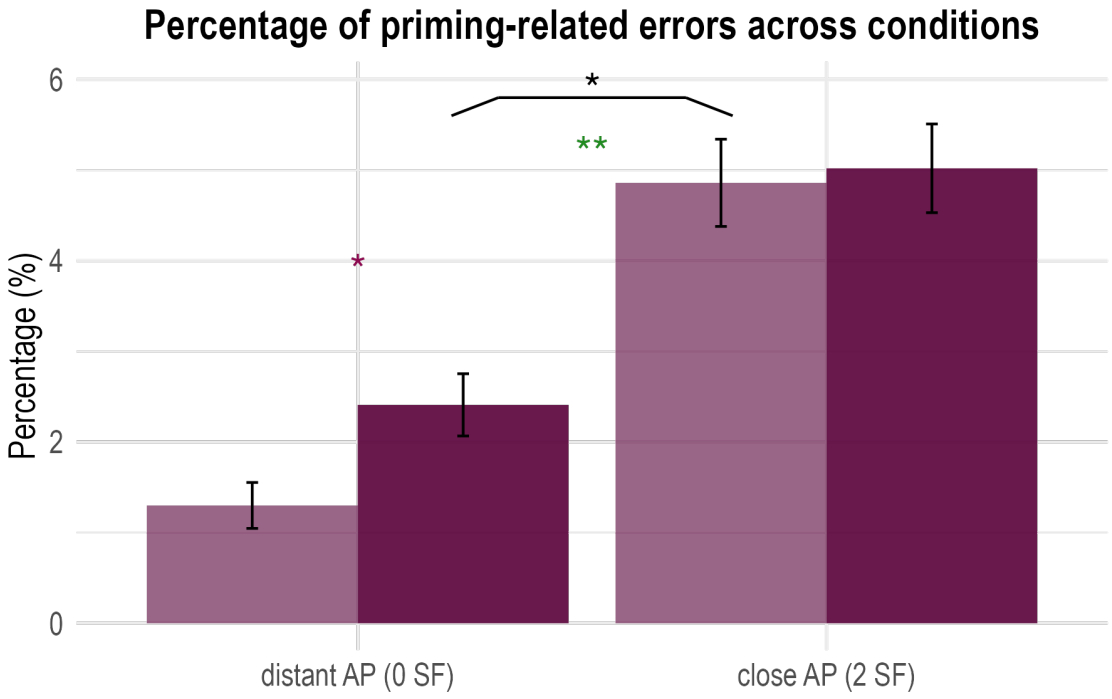


Figure A2.4 – Percentages of priming related errors across Semantic and AP conditions. Each portion is displayed with respect to the number of trials within the condition (i.e., the potential maximum number of errors). Main effects of a variable are displyed in color: semantic in merlot color, AP in green, interaction in black. SF stands for shared features.

A2.3 Pilot 3 - Doubled Semantic Priming and Articulatory-Phonetic Manipulation

In order to conduct fruitful fMRI studies, it was important to implement highly impactful manipulations. While our previous experiments exhibited a substantial articulatory-phonetic distance effect, I aspired to attain a similar level of potency in our semantic manipulation. Consequently, I have opted to intensify the semantic effect by incorporating an additional semantically related prime into the task.

A2.3.1 Stimuli

In this pilot, we employed a set of 48 word pairs. The erroneous outcome of the first word of the pair, which was paired with two synonyms that were displayed as semantic primes to the participants. Similar to our prior experiments, half of the stimuli belonged to the 'close' articulatory-phonetic (AP) condition. All participants were exposed to all word pairs, but the presentation varied: half of the pairs were presented with both articulatory-phonetic and semantic primes, while the other half was only primed with articulatory-phonetic information. To maintain consistency in the non-semantic (neutral) condition, we substituted semantic primes with morphologically similar pairs. Consequently, we generated two lists where the halves of the 48 pairs were inverted in terms of their conditions.

A2.3.2 Participants

Participants were 33 native French speakers aged between 18 and 35 y.o. (mean 24, 26 female). They all received monetary compensation of 12€ on their PayPal account.

A2.3.3 Procedure

This experiment was conducted online on FindingFive platform (FindingFive Corporation, a 501(c)(3) nonprofit organization). The procedure of the manipulation was replicated from Motley and Baars 1976. The order of the trials was as follows: three fillers, semantic prime 1, AP prime 1, semantic prime 2, AP prime 2, target (see Table A2.1). All the primes always appeared twice on the screen, phonological primes were only presented for reading, semantic primes were followed with a question mark to be read aloud. The question mark appeared randomly either after the second appearance of the prime or after its first appearance. Fillers were also sometimes followed by a question mark, but the responses were not recorded. 576 items (384 manipulation items + 192 fillers) were repeated 3 times, fillers were randomised for each repetition. The task lasted for 51 minutes with pauses each 12 minutes.

			Semantic condition		Neutral condition	
PROCEDURE		Time (ms)	Type trial	Screen	Type trial	Screen
reading		700	filler	papa baigneur	filler	papa baigneur
produce		700	filler	papa baigneur	filler	papa baigneur
		500		?		?
		1200		!!		!!
reading		700	filler	tomate cratère	filler	tomate cratère
reading		700	semprime	poème talent	Nprime	permis talent
produce		700	semprime	poème talent	Nprime	permis talent
		500		?		?
		1200		!!		!!
reading		700	phoprime	salut bétail	phoprime	salut bétail
reading		700	phoprime	salut bétail	phoprime	salut bétail
produce		700	semprime	chanson robot	Nprime	chaussure robot
		500		?		?
		1200		!!		!!
reading		700	semprime	chanson robot	Nprime	chaussure robot
reading		700	phoprime	saphir billet	phoprime	saphir billet
reading		700	phoprime	saphir billet	phoprime	saphir billet
produce		700	target	bonnet salade	target	bonnet salade
		500		?		?
	recording	1200		!!		!!

TABLE A2.1 – Pilot 3 experimental design and procedure. Blue cell backgound indicate synonymical relation in between items : *chanson* (song) and *poème* (poem) is a synonym of the outcome error word *sonnet* (sonnet)

.

A2.3.4 Data

Data of 33 participants 4752 trials, resulting in 144 target trials per participant. Semantic and AP conditions were evenly distributed in halves of the total trial count (2376 per condition). After filtering the full omissions or occurences where audio data was lost, the dataset comported the total number of 4686 exploitable trials (semantic: 2344, neutral: 2342, close AP: 2310, distant AP: 2376). For the analysis of errors, we employed a mixed-effects logistic regression model using the `glmer` function of `lme4` package in R Core Team 2022.

```
glmer(RelatedErrors ~ Semantic + AP Distance + Semantic * AP Distance
+(1+Semantic |Subject) +(1|Item), family = binomial)
```

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (Chambers et al. 1990 via R documentation) prior to conducting the statistical analysis in R Core Team 2022.

A2.3.5 Results

Participants made errors related to the priming manipulation in 82 trials 1,75% of the total number. More errors were made for word pairs whose initial consonants were AP close (61 trials out of 2310, 2.64%) compared to those that were distant(21 trials out of 2376, 0.9%). More errors were made following the semantic priming condition (46 out of 2344, 1.9%) than after the neutral priming (36 out of 2342, 1.54%). None of the effects

were significant (semantic: $z = 0.11$, $p = .9$; marginal for AP distance: $z = 1.8$, $p = .06$,) on priming-related errors with no interaction of the variables ($z = .89$, $p = .36$).

A2.4 Pilot 4 - replication of Motley and Baars 1976 in French

The pilot is a full replication of Motley and Baars 1976 in French. As far as the try to reinforce semantic effect by adding supplementary semantic prime did not have expected results and even weakened the effect of phonetic distance manipulation, I replicated the experiment from Motley and Baars 1976 where they found significantly different number of spoonerisms through the conditions. With the aim to get the contrast of monitoring at different stages of speech processing, this manipulation had to be crossed with AP distance manipulation. However, it was impossible to control first letters of each word in AP distance and then expect them to form valid noun groups. Therefore, the decision was taken to cross semantic manipulation with a syntactic one because it was indeed possible in French.

A2.4.1 Stimuli

32 pairs of words were created, 16 of them were pairs of nouns that gave a SLIP outcome as noun + adjective (e.g., *lande barge* moor barge ⇒ *bande large* wide band); the other half was noun + adjective without syntactic change (e.g., *raison murale*wall-mounted reason ⇒ *maison rurale* rural house). The translation is given with adjective - noun order of English for simplicity, the used pairs in French always had a grammatically correct noun + adjective order. As far as all the outcomes formed a logical noun group, two synonymic noun groups were found to take place as semantic primes (e.g., *foyer rustique, demeure paysanne* rustic hearth, peasant residence for *maison rurale* riral house). No control over 'distant AP' and 'close AP' was possible in this configuration.

A2.4.2 Participants

Participants were 20 native French speakers aged between 19 and 29 y.o. (mean 23; 15 female). They all received monetary compensation of 10€ on their PayPal account

A2.4.3 Procedure

This experiment was conducted online on FindingFive platform (FindingFive Corporation, a 501(c)(3) nonprofit organization). The order of the trials is exactly the same as

in Pilot 3 but the number of fillers before each trial is randomized from 1 to 3. Moreover, fillers in this experiment are nouns and adjectives. 385 items were repeated 4 times. The experiment lasted around 48 minutes and contained 3 pauses each 12 minutes.

A2.4.4 Data

Data of 20 participants 2560 trials, resulting in 128 target trials per participant. Semantic and syntactic conditions were evenly distributed in halves of the total trial count (1280 per condition). After filtering the full omissions or occurences where audio data was lost, the dataset comported the total number of 2544 exploitable trials (semantic: 1278, neutral: 1266, syntactic change: 1269, no-change: 1275). For the analysis of errors, we employed a mixed-effects logistic regression model using the `glmer` function of `lme4` package in R Core Team 2022.

```
glmer(RelatedErrors ~ Semantic + Syntax + Semantic * Syntax
+(1+Semantic |Subject) +(1|Item), family = binomial)
```

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (Chambers et al. 1990 via R documentation) prior to conducting the statistical analysis in R Core Team 2022.

A2.4.5 Results

Participants made errors related to the priming manipulation in 54 trials 2.11% of the total number. Similar amount of errors was found across semantic conditions (semantic: 28 out of 1278, 2.19%; neutral:26 out of 1266, 2.05%). Slight difference in error patterns was observed across syntactic change condition (syntactic change: 23 out of 1269, 1.81% ; no-change: 31 out of 1275, 2.43%) None of the effects were significant (semantic: $z = .77$, $p = .43$; syntactic: $z = .29$, $p = .76$,) on priming-related errors with no interaction of the variables ($z = .05$, $p = .95$).

The absence of the AP effect in Pilot 3 and 4 was interpreted by the fact that the first articulatory-phonetic prime is firstly, relatively distant from the target word, and secondly, it is distant from the other AP prime. Consequently, the impact of the phonetic distance manipulation is less pronounced when the primes do not immediately precede the target word and are not presented consecutively. The manipulation that was therefore retained for the fMRI study described in Chapter A1 is the manipulation from Pilot 2.

A2.5 References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2021). Matrix: sparse and dense matrix classes and methods [R package version 1.3-4]. https://CRAN.R-project.org/package=Matrix

Chambers, J., Hastie, T., & Pregibon, D. (1990). Statistical Models in S. In K. Momirović & V. Mildner (Eds.), Compstat (pp. 317–321). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-50096-1_48

Dorokhova, L., Morillon, B., Baus, C., Belin, P., Dubarry, A.-S., Alario, F.-X., & Runnqvist, E. (2023). A robust temporal map of speech monitoring from planning to articulation. https://doi.org/10.31234/osf.io/acwqd

Hothorn, T., Bretz, F., & Westfall, P. (2021). multcomp: simultaneous inference in general parametric models [R package version 1.4-25]. https://CRAN.R-project.org/package=multcomp

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2021). lmerTest: tests in linear mixed effects models [R package version 3.1-4]. https://CRAN.R-project.org/package=lmerTest

Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. Cognition, 4(2), 177–187. https://doi.org/10.1016/0010-0277(76)90003-2

R Core Team. (2022). R: a language and environment for statistical computing [Version 4.2.2, released on 2022-10-31]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Wickham, H. (2016). ggplot2: elegant graphics for data analysis [R package version 3.4.3]. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: a grammar of data manipulation [R package version 1.0.7]. https://CRAN.R-project.org/package=dplyr