Bridging the neurocognitive mechanisms of inner speech and phenomenology: a multi-mechanistic approach

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2023

Jaydan Pratts

School of Health Sciences

Division of Human Communication, Development and Hearing

Table of contents

Word count = 42,430

List of Figures

List of Tables

Abstract

Inner speech is an internally perceptible speech-like experience which occurs in the absence of actual sound. For many, inner speech represents a part of the daily conscious experience, yet its cognitive and neural underpinnings remain unclear. Two issues in inner speech research are addressed in this thesis. First, no single cognitive model of inner speech generation fully explains the wide range of observed neural correlates. Second, the diverse phenomenology of inner speech and the potential for interactions with its generative mechanisms is often neglected.

The aim of this thesis is to develop an integrated model of inner speech which explains heterogeneous findings and accounts for its diverse phenomenology. The predictive validity of this model is then tested using a range of experimental techniques. Chapter 2 used an fMRI meta-analysis to determine which brain regions are commonly seen across inner speech neuroimaging studies. Evidence for the involvement of both speech production and speech perception regions in inner speech was found, depending on the phenomenology of the elicited inner speech. Chapter 3 found that individuals with aphantasia, which is argued to stem from dysfunction of perceptual imagery mechanisms, exhibited deficits in inner speech which varied across phenomenological dimensions. Chapter 4 used a series of experiments adopting a dual-task paradigm to tax speech perceptual regions and speech production regions, respectively. Results demonstrated that motor and perceptual suppression efficacy vary depending on inner speech phenomenology, indicating a role for both mechanisms dependent on the type of inner speech involved. Chapter 5 used transcranial magnetic stimulation to suppress brain regions associated with speech production and speech perception, respectively. A trend towards stimulation of speech production regions suppressing all phenomenological varieties was observed, although this did not reach statistical significance.

Overall, research findings varied depending on the phenomenology of the elicited inner speech - with the exception of Chapter 5. This supports our argument that inner speech should be interpreted as a dynamic and variable phenomena with varied generative mechanisms and neural correlates. However, the exact generative mechanisms and precisely how they interact with phenomenology could not be determined, with results across Chapters 2 - 5 not being sufficiently consistent with our integrated model, nor with existing models of inner speech.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.

Acknowledgements

I would like to thank the range of people who made this work possible. Dr Bo Yao and Dr Gorana Pobric were exceedingly generous with their time, feedback and dedication. It was an immense privilege to work on inner speech and transcranial magnetic stimulation under the supervision of subject-experts who were always eager to share their knowledge. Dr Yao’s efforts to foster a sense of community within the EPIC laboratory were hugely appreciated and I can safely say that the regular meetups made the lab the envy of the PhD office. I would also like to thank all individuals who participated in my research, especially given that much of it involved sitting in silence, receiving brain stimulation, or both. Academically, there were a host of authors too long to list but whose works were fundamental in my understanding of inner speech and how it is best studied. I would like to express my gratitude to my friends and family who have all played a larger role in this journey than they might realise. Finally, I would like to acknowledge my dear friend Jesse Monteverde. When research and writing-up proved to be a challenge, I was continually inspired by Jesse's incredible journey of losing 30 kg (and counting) and walking across Iceland. His determination and resilience serve as a constant reminder of what can be achieved with hard work and persistence. My trip to his hometown, Lepe, was particularly enjoyable and served as a well-deserved break from the writing of this thesis.

Acknowledgement of Authors' Contributions

Dr Bo Yao and Dr Gorana Pobric supervised all Chapters of this thesis, and the research which constitutes it. The PhD candidate, Mr Jaydan Pratts, recruited all participants and conducted all experimentation, statistical analyses and write-ups. Feedback was given on the thesis draft which was used for further development. Dr Gorana Pobric provided training on the use of transcranial magnetic stimulation, and Dr Bo Yao provided instruction material and feedback on the use of certain statistical techniques.

Chapter 1 - General Introduction

This thesis is written in the Journal Format, also known as the Alternative Format. Chapter 1 is used to provide an overview of previous research, as well as to explain the rationale behind the research direction and approach used in this thesis. Chapters 2 - 5 are manuscripts written in a format suitable for publication in an academic journal. As such, they are written in a manner which is understandable if a chapter is read in isolation, with each containing its own introduction, methodology, results, and discussion. Chapter 6 will be used to provide an overall discussion in which the results of all previous chapters can be considered in the context of our overall thesis aim.

Thesis Aims

Despite an accumulating body of research on the neurocognitive underpinnings of inner speech, there is conflicting evidence on how inner speech is generated and which neural substrates underpin it. In part, this could result from the tendency to interpret inner speech as a single monolithic phenomenon, rather than a phenomenologically diverse experience. The extent to which the phenomenology of inner speech might influence its underlying neurocognitive mechanisms is under-investigated and could explain disparate findings in the research literature. Therefore, the objective of the thesis is to develop and test a framework which bridges the diverse phenomenology of inner speech and its underlying neurocognitive mechanisms. To do this, we will review and integrate existing neurocognitive models of inner speech into a single framework, and then test its predictive validity across a diverse set of paradigms. These aims can be summarised as such:

1. To develop an integrated model of inner speech which explains heterogeneous findings and accounts for the diverse phenomenology of inner speech.
2. To determine the validity of this model by testing its predictive validity using a range of experimental methodologies.

In order to clearly justify the motivations for this thesis, I will first provide some context on inner speech research and previous findings on its neural correlates. I will then explore recent research that suggests that these might vary depending on phenomenology. Subsequently, I will examine the extent to which the methodological techniques commonly used in the research literature capture the varied phenomenology of inner speech. Finally, we will explore the specific aims of each chapter within this thesis.

Defining Inner Speech

Inner speech is an internal, speech-like experience which although inwardly audible, does not involve the production of actual sound. As such, it has been described as a “mental simulation of speech” (Perrone-Bertolotti et al., 2014). It is recognisable as a central part of the day-to-day conscious experience for many individuals, where it is sometimes termed “a little voice in my head” which narrates the conscious experience (Perrone-Bertolotti et al., 2014; Heavey & Hurlburt, 2008). Inner speech can vary widely in the extent to which it is addressed to oneself, the extent to which the inner voice is that of the speaker as opposed to another individual, and the extent to which it is similar to overt speech (Alderson-Day et al., 2018). This reflects research which notes the diverse phenomenology of inner speech (Hurlburt et al., 2013).

Research has implicated inner speech in a wide variety of tasks and cognitions ranging from working memory (D’Esposito, 2007), silent reading (Filik & Barber, 2011) and behavioural self-regulation (Diaz, Berk, & Diaz, 2014), to effective performance in cognitive tasks such as task switching and tracking task goals (Emerson & Miyake, 2003; Miyake, Emerson, Padilla, & Ahn, 2004). Inner speech also forms the basis of a prevalent explanation of auditory verbal hallucinations in which it is proposed that auditory verbal hallucinations emerge from the misidentification of inner speech as being externally generated (Frith, 1987). Given the utility of inner speech and its potential role in disorders, it is imperative that we develop a robust understanding of the cognitive and neural underpinnings of inner speech.

Production of Inner Speech

Early research utilising muscle measurement and electromyography techniques found that inner speech was associated with significant activation of speech musculature (Curtis, 1900; Jacobson, 1932; Max, 1935), this set the foundations for an enduring school of thought in which inner speech is considered a weakened motoric experience reliant on the brain regions utilised in overt speech (Scott, 2013). The introduction of more recent neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) further evidenced a link between inner speech and motor speech production, with activation of the left inferior frontal gyrus (L-IFG), a region implicated in overt speech production (Tourville & Guenther, 2011), being noted across several studies. For example, a PET-study by McGuire et al. (1995) found that uttering sentences in inner speech was associated with significant increases in left inferior frontal gyrus activation. Similarly, an fMRI-study by Lurito et al. (2000) observed left inferior frontal gyrus activation in a phonological judgement task in which participants covertly determined whether two words rhymed. The findings of McGuire et al. (1995) and Lurito et al. (2000) represent a common observation across studies investigating inner speech, with an initial review of the literature identifying 15 further studies which observed increased activation of the L-IFG during the generation of inner speech (Table 1.).

Table 1. List of studies reporting a correlation between inner speech and left inferior frontal gyrus activation. Arrows represent increased or decreased activation.

Study	Paradigm	Observation
Lurito et al. (2000)	Determine whether two words rhyme.	↑ L-IFG
Okada et al. (2018)	Silently articulate a sequence of words.	↑ L-IFG
Okada et al. (2018)	Silently articulate a tongue twister.	↑ L-IFG
Simmonds et al. (2014)	Covertly define a visually presented noun.	↑ L-IFG
Grandchamp et al. (2019)	Covertly define a visually presented object	↑ L-IFG
Kleider-Offutt et al. (2019)	Covertly read visually presented sentence in the voice of previously introduced actor.	↑ L-IFG
Bitan et al. (2007)	Determine whether two words rhyme.	↑ L-IFG
Papathanassiou et al. (1999)	Covertly name verbs relating to visually presented nouns.	↑ L-IFG
Herholz et al. (2012)	Presented with visual cue to well-known lyrical melody. Asked to imagine the remainder of the melody.	↑ L-IFG
Shergill et al. (2001)	Create a sentence ending with the visually presented probe word.	↑ L-IFG
Aleman et al. (2004)	Visually presented with words and asked to determine weak and strong syllables.	↑ L-IFG
Price et al. (2011)	Articulate “one” and “three”.	↑ L-IFG
Zvyagintsev et al. (2013)	Recall a familiar melody which you often hear.	↑ L-IFG
Badcock et al. (2011)	Silently generate word which audio cues are describing.	↑ L-IFG

Linden et al. (2010)	Imagine hearing the voice of a familiar person talking to you.	↑ L-IFG

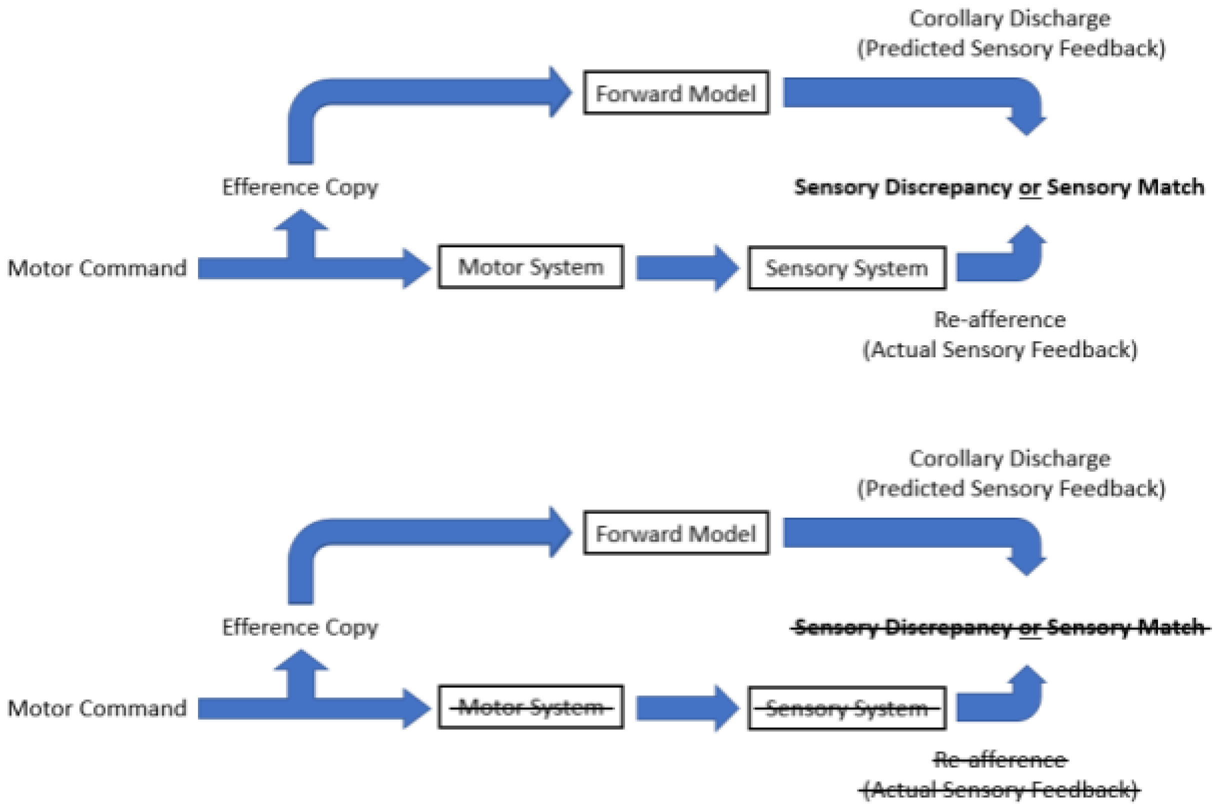
The recurring finding that inner speech shared commonalities with overt speech, whether it be in neural correlates or musculature activation, underpinned a series of models which argued that inner speech was the result of truncated speech motor commands. The Motor Theory of Voluntary Thinking (Cohen, 1986) presents one early model, with its antecedents can be found dating back to the 19th century (Bain, 1859). More recent implementations of the motor theory of inner speech can be found in Grandchamp et al.'s (2019) ConDialInt model of inner speech. Whilst recent models benefit from advances in neuroimaging techniques, the fundamental aspects of the motor theory of inner speech remain relatively unchanged across the years.

The motor theory of inner speech posits that inner speech is a truncated form of overt speech (Feinberg, 1978). More specifically, that inner speech is derived from the efference copies and corollary discharge which support overt speech (Miall and Wolpert, 1996). Efference copies represent a copy of a motor movement command, generated in tandem with the original command, but used for predictive purposes rather than motor movement. While the original motor command propagates to motor areas where it generates a motor movement (e.g. speech via activation of articulators), efference copies propagate through a feed-forward system which uses the efference copy to predict the sensory consequences of the motor movement (Imamizu, 2010). This neural prediction is termed as corollary discharge, and is thought to terminate within the relevant sensory region (Tian et al., 2016). If this predicted sensory consequence matches the actual sensory input being perceived by an individual, it can be deduced that the input was the result of one's own motor commands, and the input can be disregarded. However, if the predicted sensory input is different from the perceived sensory input, the input is likely caused by an external agent, justifying additional attentional resources (Cullent, 2004).

This system whereby observed sensory stimuli are compared to predicted sensory stimuli facilitates the determination of agency (Feinberg and Guazzelli, 1999), as well as motor error-correction (Crapse and Sommer, 2008). A practical example of the usefulness of corollary discharge can be seen in the way we hear our own voice. When we speak, motor commands propagate to our articulators which result in sound. The corollary discharge mechanism uses an efference copy to predict this sensory outcome and suppresses the auditory input that corresponds to our own voice (Greenlee, 2011). This dampening of neural responses to one's voice has practical benefits, dampening perceptions of one’s own voice whilst allowing an individual to remain sensitive to the speech of others (Heinks-Maldonado et al., 2005).

To adapt the corollary discharge model to inner speech, proponents argue that efference copies and corollary discharge can be generated without executing the original motor movement signal, therefore generating an internally perceptible prediction without generating accompanying overt speech (Scott, 2013). It is also argued that the source of the motor movement signals and the accompanying efference copy is the L-IFG (Tourville and Guenther, 2011; Tian et al., 2016). This aligns with previous works dating back to the Broca-Wernicke-Lichtheim-Geschwind model wherein speech production originates from the inferior frontal regions (Broca, 1861). More recently, the DIVA model of speech production (Tourville & Guenther, 2010) also implicates the L-IFG in the initial stages of speech generation. According to the DIVA model, posterior portions of the left IFG (i.e. pars opercularis) subserve a speech sound map which holds a repository of motor programmes for frequently used syllables. This speech sound map directly projects to the motor cortex where the selected motor programme is propagated to the articulatory musculature to produce speech.

Figure 1. Diagram outlining the core pathways involved in corollary discharge.

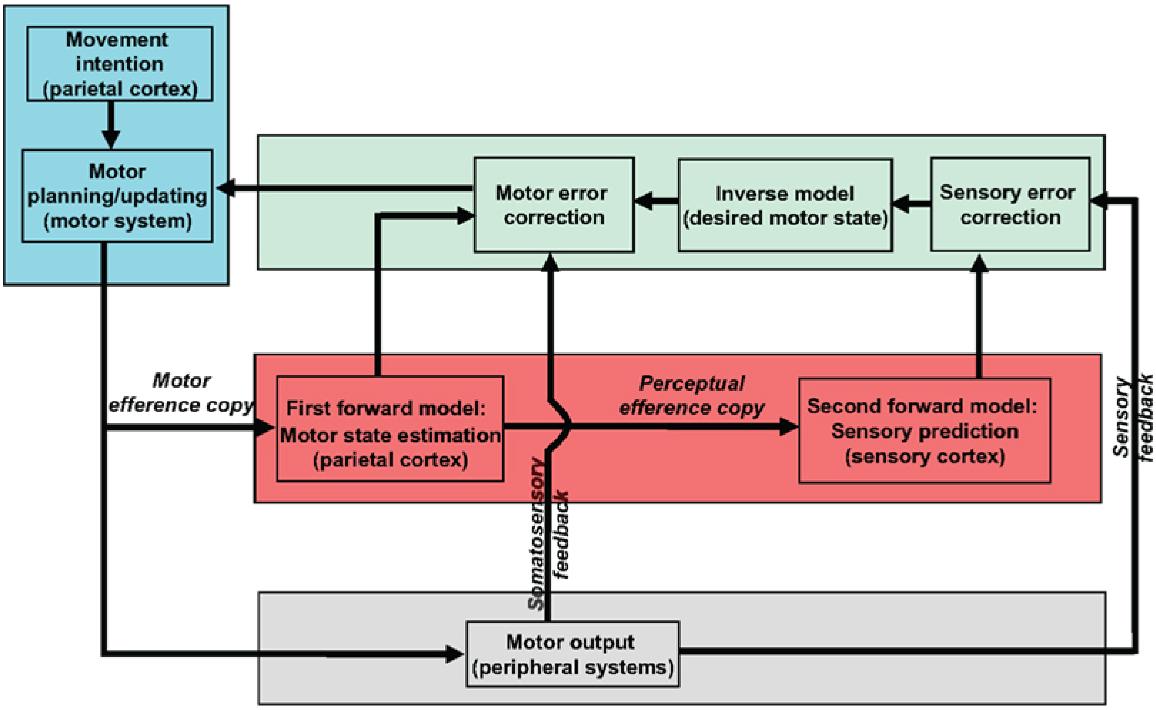


Note: Execution of a typical motor command (above) compared to the modified pathway which subserves inner speech (below).

Evidence in favour of corollary discharge supporting inner speech is not limited to neuroimaging studies observing activation of the left inferior frontal gyrus. One inherent attribute of corollary discharge is that it results in the perceptual and neural attenuation of signals which are deemed to be self-generated (when predicted and observed feedback matches). As such, if inner speech was the result or accompaniment of corollary discharge it should attenuate neural and perceptual responses in certain circumstances. Scott (2013) examined this hypothesis in positing that if inner speech was associated with the generation of corollary discharge, perceived loudness of external noise should be attenuated when matching inner speech was simultaneously generated, but not when the inner speech did not match the external noise. The results of his experiment corroborated this hypothesis with the perceived loudness of the external noise being lower when accompanied by matching inner speech. Jack et al. (2019) presents additional evidence by adopting a similar paradigm with the additional use of electroencephalography (EEG). Participants who generated inner speech which matched an external noise elicited a dampened N1 event-related potential (ERP). Given that the N1 ERP is viewed as an index of auditory processing, this supports the notion that inner speech is accompanied by corollary discharge which attenuates neural responses as well as perceptual responses (Scott, 2013).

A critical assumption which is inherent to a corollary discharge model of inner speech is that brain regions activate in a sequential manner. Inner speech originates in a brain region which generates the relevant motor / efference signal which then travels to a sensory region in order to be inputted into a sensory forward model and attenuate any potential perceptual response. Given that this type of directional pathway has been observed in corollary discharge of other motor-sensory domains (e.g. vision; Sommers & Wurtz, 2004), Tian & Poeppel (2010) adopted magnetoencephalography (MEG) techniques in order to determine whether such a directional pattern of activation could be seen in an inner speech task. Importantly, Tian & Poeppel (2010) propose that corollary discharge involves two forward models, as opposed to the more general single forward model as previously described. After the generation of a motor efference copy, it is inputted into the first forward model within the parietal cortex which estimates the motoric state which results from the motor command (e.g. what is the upcoming position of the vocal tract, lips and tongue). Once the motor state has been estimated, the resulting perceptual efference copy is inputted into a second forward model within the sensory cortex which then predicts the sensory consequences of the estimated motor state. It is this portion of the process, the propagation of a signal from the first forward model to the second forward model, that was examined. In line with the expectation that activation should occur in a specific temporal order, they did observe that activation of the auditory cortex occurred *after* the signal was processed within the parietal cortex.

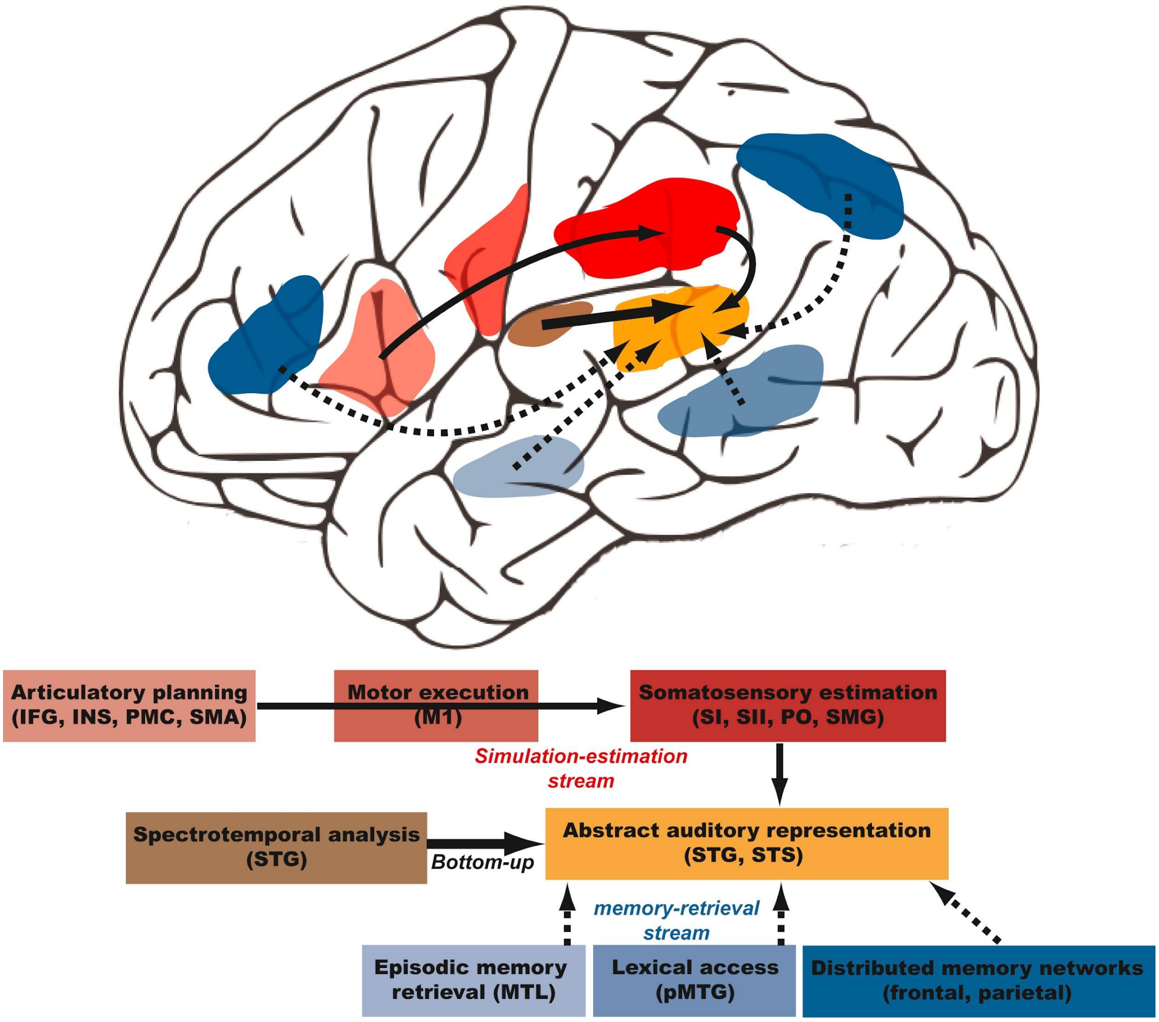
Figure 2. Corollary discharge system proposed by Tian & Poeppel (2010).



The interpretation of inner speech as an outcome of corollary discharge provides a plausible cognitive mechanism which can explain the neuropsychological findings discussed so far. However, the research literature is not wholly consistent with the view that inner speech is invariably the outcome of corollary discharge. For example, whilst adopting a similar rhyme judgement paradigm to Lurito et al. (2000), Rudner, Ronnberg & Hugdahl (2005) did not find significant activation of the left or right IFG, instead observing activation of the left middle temporal gyrus and bilateral superior temporal gyrus. Similarly, Yao, Belin & Scheepers (2011) found that spontaneous inner speech elicited by the reading of direct quotations was associated with increased activation of voice-sensitive regions within the auditory cortex. Hurlburt et al. (2016) observed similar activation, with spontaneous inner speech being associated with voice-sensitive regions rather than the L-IFG. These studies are not unique in running contrary to the notion that inner speech is invariably the result of corollary discharge. As presented in MacKay (1981), inner speech is significantly faster than overt speech. Whilst this does not preclude the possibility that inner speech is generated using motor signals destined for speech articulators, it does suggest that inner speech is not an exact representation of the planned overt speech. It is also notable that some individuals who are congenitally unable to articulate can accurately perform tasks reliant on inner speech, such as rhyme judgement tasks (Bishop & Robson, 1988). One explanation could be that those with congenital disorders have developed alternative strategies which can be used to carry out phonological judgements without the need for inner speech. However, this ability to accurately perform tasks reliant on inner speech whilst experiencing perturbations of overt speech production is not restricted to those with congenital disorders. Some evidence exists showing that the overt articulation of irrelevant speech (articulatory suppression) does not inhibit inner speech to a significant degree (Wheeldon & Levelt, 1995). This presents significant challenges to the corollary discharge interpretation of inner speech given that the irrelevant overt speech would generate irrelevant corollary discharge which should interfere with the main inner speech task.

Although the predominant model of inner speech production, alternatives and adaptations to the corollary discharge model have been proposed. Tian et al. (2016) present a study investigating the mechanisms by which the mental imagery of speech can be generated. They hypothesised that the internal perception of speech can be generated via two mechanisms, a simulation-estimation stream and a memory-retrieval stream.The simulation-estimation stream represents a pathway involving forward models which begins in brain regions associated with articulatory planning (inferior frontal gyrus, insula, premotor cortex, supplementary motor area) and terminates in the auditory cortices (superior temporal sulcus, superior temporal gyrus). This mechanism is broadly equivalent to that proposed in the corollary discharge model of inner speech. In the memory-retrieval stream, memories of speech are retrieved from hippocampal structures or other regions which support long term memory. These retrieved memories are transferred to the sensory cortices where they were originally processed (i.e. auditory cortex) where they are then re-constructed, allowing for an auditory representation to be generated.

Figure 3. Proposed cortical regions involved in each model. (Tian, Zarate & Poeppel, 2016)

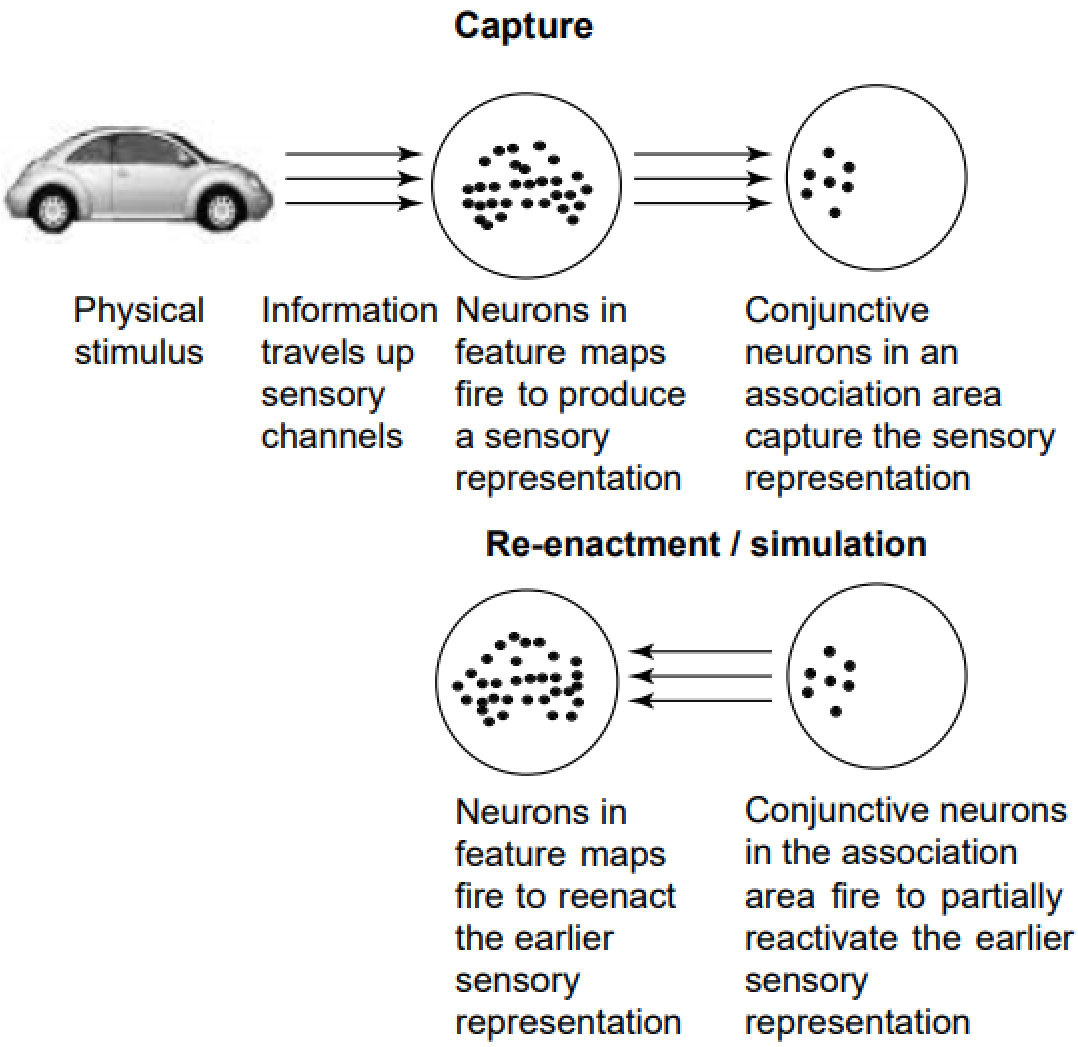


Tian, Zarate & Poeppel (2016) examined the validity of these models by requiring participants to either imagine articulating a word or imagine hearing a word. If an internal representation of speech could be generated by both a simulation-estimation stream and a memory-retrieval stream, it would be expected that both articulation imagery and hearing imagery would result in activation of the superior temporal gyrus and superior temporal sulcus, the regions which likely underlie the perceptible aspects of inner speech. Moreover, they propose that articulation imagery provides a more precise and detailed representation of speech and would therefore lead to increased activation of the superior temporal sulcus and gyrus when compared to hearing imagery. The findings were in line with these hypotheses, with results supporting the existence of both a simulation-estimation stream based on corollary discharge and a memory-retrieval stream based on the re-activation of stored speech memories.

Tian, Zarate & Poeppel (2016) are not alone in proposing an alternative, perceptually-based mechanism of inner speech production. Yao et al. (2011) used functional magnetic resonance imaging (fMRI) to measure brain activity in participants as they silently read sentences in direct and indirect speech. Direct speech was classified as sentences in which the words of a speaker are directly quoted, such as "John said, 'I am going to the store'". Indirect speech were sentences in which the words of the speaker were paraphrased, such as "John said that he is going to the store". Yao et al. (2011) found that silent reading of direct versus indirect speech preferentially activated voice-selective areas in the auditory cortex, implying the spontaneous emergence of an inner voice as participants recreated the described characters voice. In explaining the observed activation of voice-sensitive areas during direct speech reading, Yao et al. (2011) posited that inner speech could emerge from perceptual simulation, a mental reenactment of a memorised perceptual state (Barsalou, 2009).

The system of perceptual simulation posited by Yao et al. (2011) overlaps significantly with the memory-retrieval stream proposed by Tian et al. (2016). The memory-retrieval stream represents a system in which memories of speech are retrieved and reconstructed, allowing for an internally perceptible auditory representation to be generated within perceptual regions. Similarly, the perceptual simulation referenced in Yao et al. (2011) represents a system in which the neuronal patterns fired in response to perceptual stimuli are captured and stored. These stored neuronal firing patterns can be recreated at a later time, allowing for the perceptual experience to be simulated within the relevant sensory region (Barsalou, 2009).

Figure 4. Process by which stimuli are encoded and re-enacted according to a perceptual simulation model (Barsalou et al., 2003).



Further evidence supporting the idea that speech perceptual regions can produce inner speech is presented in Hurlburt el al. (2016). Hurlburt el al. (2016) investigated whether spontaneous inner speech resulted in similar neural activation to task-elicited inner speech, the predominant type of inner speech investigated in the research literature. fMRI region-of-interest (ROI) analysis explored activation of two brain regions typically implicated in inner speech, the L-IFG and Heschl’s gyrus, a region associated with speech perception. Whilst task-elicited inner speech was associated with increased activation of the L-IFG, spontaneous inner speech led to decreased activation of the L-IFG. Further differences were seen in Heschl’s gyrus which saw decreased activation during task-elicited inner speech and increased activation during spontaneous inner speech. In proposing an explanation for this finding, Hurlburt et al. (2016) suggests that inner speech can vary based on its emphasis on either production or perception. Given the activation of the role of the left-IFG in speech production and its activation during task-elicited inner speech, it is proposed that task-elicited inner speech is generated using the speech production regions utilised in overt speech. In contrast, spontaneous inner speech was correlated with activation of speech perception brain regions, rather than speech production brain regions, suggesting a type of inner speech with little or no need for the neural substrates which underlay overt speech production.

Phenomenology of Inner Speech

Hurlburt et al.'s (2016) finding that inner speech which emerges spontaneously exhibits distinct neural activation when compared to task-elicited inner speech suggests that inner speech is a varied and dynamic phenomenon which exhibits interplay between its phenomenology and its neural correlates. This characteristic of inner speech is often neglected in research investigating its neural correlates, which typically examine task-elicited inner speech without assessing the ecological validity of such an approach (Hurlburt et al., 2013). Notably, the lack of consideration for the impact of inner speech phenomenology on underlying neurocognitive mechanisms cannot be attributed to a lack of research on inner speech phenomenology, as several such studies exist. McCarthy-Jones & Fernyhough (2011) developed and tested an 18-item self-report questionnaire, the Varieties of Inner Speech Questionnaire (VISQ; McCarthy-Jones & Fernyhough, 2011), which assesses the phenomenological characteristics of inner speech. The items on the questionnaire aimed to capture four dimensions of inner speech, mostly derived from Vygotsky's (1987) works on the emergence of speech during child development: dialogicality (the extent to which inner speech reflects a back-and-forth dialogue), evaluative/motivational (the extent to which inner speech serves evaluative or motivational purposes), other people in inner speech (whether the inner speech represented the voice of another person, as opposed to oneself) and condensed/expanded (whether the inner speech was abbreviated into smaller verbal images, rather than fully formed prose). Its initial validation on 235 university students found the VISQ to show good test-retest and internal reliability, as well as being able to predict auditory-hallucination proneness and anxiety. However, how each factor maps to underlying neurocognitive mechanisms is less clear given the lack of neuroimaging studies investigating the four VISQ factors.

One phenomenological model which is accompanied by neuroimaging evidence is the ConDialInt model proposed by Grandchamp et al. (2019), which derives several attributes from the Varieties of Inner Speech Questionnaire (McCarthy-Jones &

Fernyhough, 2011). The ConDialInt model maintained the condensation and dialogicality dimensions of the VISQ models, but added an additional dimension measuring intentionality (the extent to which inner speech emerges spontaneously or as a response to task or cognitive demands). Grandchamp et al. (2019) used fMRI to investigate the neural correlates of inner speech subtypes varying across dialogicality and intentionality dimensions, with the condensation dimension not being formally investigated. Dialogicality was examined by asking participants to generate inner speech in the voice of either an avatar, or their own voice. Inner speech in the voice of an avatar resulted in activation of regions in the right hemisphere homologues to those involved in the generation of inner speech in one's own voice, the right inferior parietal cortex and right inferior frontal gyrus. Inner speech derived from mind-wandering, which represented the unintentional inner speech condition, was associated with greater activation in the bilateral inferior frontal cortex and decreased activation in left temporal regions when compared to task-elicited inner speech.

Methodologies in Inner Speech Research

Grandchamp et al. (2019) represents an example of a study which considers the potential impact of inner speech phenomenology on its neural correlates. However, this relationship between phenomenology and neurocognitive mechanisms remains an uninvestigated aspect of inner speech. This presents challenges when assessing which neurocognitive model of inner speech is best supported by the existing evidence. Prima facie, the corollary discharge interpretation of inner speech might seem best supported by the current research literature given its range of fMRI and EEG evidence. However, a substantial proportion of this research investigates a specific subtype of inner speech - inner speech which is task-elicited and in one's own voice (*see Table 1.*). As evidenced by Hurlburt et al. (2016), Yao et al. (2011) and Grandchamp et al. (2019), the neural correlates of inner speech are sensitive to phenomenology, and the findings of studies investigating a specific subtype of inner speech may not represent the neurocognitive mechanisms responsible for the type of inner speech experience in day-to-day life. As suggested by Yao et al. (2011) and Tian et al. (2016), it is even possible that inner speech is not subserved by a single neurocognitive mechanism, but by multiple mechanisms depending on the attributes of the inner speech.

Given the evidence indicating that the neural correlates of inner speech are dependent on its phenomenological attributes (Grandchamp et al., 2019; Hurlburt et al., 2016; Yao et al., 2011, 2012), it is likely that a complete understanding of the neurocognitive mechanisms driving inner speech will require a range of experimental paradigms which examine a range of inner speech subtypes. As argued in Alderson-Day & Fernyhough (2014), introducing multiple methodologies to inner speech research would also allow for the weaknesses of individual techniques to be offset by the strengths of others. For example, a benefit of the use of fMRI in inner speech research is its ability to determine which brain regions are active during inner speech generation. However, fMRI is an observational technique which cannot determine whether the observed brain activation is a causal driver of the behaviour of interest (Poldrack, 2006). This limitation can be offset via the use of transcranial magnetic stimulation (TMS), a method in which the brain regions observed in the fMRI analysis can be temporarily disrupted, allowing for causal inferences to be made about their relationship to the behaviour (Wassermann and Zimmermann, 2012). Whereas TMS uses highly focal magnetic pulses to target specific brain regions, it is also possible to implement a dual-task interference paradigm in which individuals simultaneously carry out two tasks. By examining whether a concurrent secondary task impacts on the performance of the main task, it is possible to infer whether the two tasks rely on similar cognitive resources (Klingberg and Roland, 1997). For example, examining whether chewing gum impacts on the quality or quantity of inner speech could indicate whether inner speech relies on orofacial motor regions. This presents a complementary addition to both fMRI and TMS approaches: the fMRI methodology captures the location of neural activation in 3-dimensions, but cannot easily determine causality. The TMS methodology can assess the causal role of brain regions in a behaviour, but requires previously collected neuroanatomical coordinates. And dual-task paradigms can be used to deduce the cognitive components supporting a behaviour, but do not provide evidence on the precise brain regions supporting these components.

As explored, the use of multiple experimental methodologies in inner speech research can compensate for the weakness inherent to any individual technique. This can facilitate a more robust understanding of inner speech which can explain its neural correlates, the causal roles which they play in inner speech generation, and how these interact the underlying cognitive mechanisms. The collection of converging evidence from distinct methodologies examining different subtypes of inner speech can also ensure that resulting models are less likely to be the result of paradigm-specific or inner speech-subtype specific dynamics.

Aims of Each Chapter

A key overall aim of the thesis was to develop and test a framework which bridges inner speech phenomenology and its underlying neurocognitive mechanisms. An attribute of this framework is that it should facilitate predictions as to which neurocognitive mechanism subserves which subtypes of inner speech. Examining the accuracy of these predictions represents a useful approach to testing the validity of the framework, and therefore serves as a key part of this thesis. The development and initial examination of the framework was addressed in Chapter 2. Chapters 3, 4 and 5 each aimed to test the predictions made by the framework using distinct research techniques, allowing for converging evidence to be generated. This approach of investigating multiple subtypes of inner speech using distinct experimental methods aims to address some of the weaknesses highlighted in the introductory paragraphs - namely, that previous inner speech research has typically relied on a select number of elicitation techniques and experimental paradigms to make assumptions of inner speech as a whole.

The specific aims per chapter were:

Chapter 2. To analyse the existing neuroimaging literature on inner speech and integrate the disparate findings into a single framework. Initial validation of this framework would then be completed by comparing the neural correlates predicted by the framework, with those observed in a meta-analysis of the neuroimaging literature.

Chapter 3. To test the predictive validity of the framework using a dissociation-based neuropsychological approach. Specifically, to compare the impact of aphantasia on a variety of inner speech subtypes. Differences in the effect of aphantasia were expected depending on the phenomenology of the inner speech in the particular condition.

Chapter 4. To test the predictive validity of the framework using a dual-task interference paradigm. Specifically, to compare the impact of a perceptually-based secondary task and motor-based secondary task on a variety of inner speech subtypes. Graded differences in the effectiveness of each interference condition were expected depending on the phenomenology of the inner speech in the particular condition.

Chapter 5. To test the predictive validity of the framework using non-invasive brain stimulation. Specifically, to examine the impact of virtual-lesions over different brain regions on inner speech performance. Differences in the effect of the virtual-lesion were expected depending on the brain region targeted and the phenomenology of the inner speech in the particular condition.

Chapter 2 - Bridging Phenomenology and Neural Mechanisms of Inner Speech: ALE Meta-Analysis on Egocentricity and Intentionality in a Dual-Mechanistic Framework

Jaydan Pratts[1*], Gorana Pobric[1], Bo Yao[12*]

1 Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, UK
2 Department of Psychology, Fylde College, Lancaster University, UK

*Corresponding Authors:
Jaydan Pratts
Division of Psychology, Communication and Human Neuroscience
Zochonis Building
Brunswick Street
Manchester M13 9PT jaydan.pratts@manchester.ac.uk
United Kingdom

Dr Bo Yao
Department of Psychology
Lancaster University
Lancaster LA1 4YF b.yao1@lancaster.ac.uk
United Kingdom

Data Statement :Analysis, stimuli, output and appendix are made available to the public at https://osf.io/n7rwq/?view_only=d3504cbeb4164d118eaa3797c327cee5

Abstract

The neural mechanisms of inner speech remain unclear despite its importance in a variety of cognitive processes and its implication in aberrant perceptions such as auditory verbal hallucinations. Previous research has proposed a corollary discharge model in which inner speech is a truncated form of overt speech, relying on speech production-related regions (e.g. left inferior frontal gyrus). This model does not fully capture the diverse phenomenology of inner speech and recent research suggesting alternative perception-related mechanisms of generation. Therefore, we present and test a framework in which inner speech can be generated by two separate mechanisms, depending on its phenomenological qualities: a corollary discharge mechanism relying on speech production regions and a perceptual simulation mechanism within speech perceptual regions. The results of the activation likelihood estimation meta-analysis examining inner speech studies support the idea that varieties of inner speech recruit different neural mechanisms.

Keywords: inner speech, corollary discharge, perceptual simulation, meta-analysis, GingerALE

Highlights:

- Activation likelihood estimation meta-analysis of neuroimaging inner speech studies.
- Distinct activation patterns observed when stratifying inner speech by egocentricity and intentionality.
- Activation of speech production regions not observed in all inner speech subtypes.

Introduction

Inner speech is an internal, speech-like experience without the presence of an external sound (Alderson-Day & Fernyhough, 2015). Inner speech has been implicated in a wide variety of cognitive tasks including working memory (Baddeley, 1992; D'Esposito, 2007), silent reading (Filik & Barber, 2011; Yao & Scheepers, 2011; Yao et al., 2011), behavioural self-regulation (Diaz et al., 2014), as well as task switching and goal tracking (Emerson & Miyake, 2003; Miyake et al., 2004). Impairments of inner speech are often associated with mental disorders such as auditory verbal hallucinations in schizophrenia (Frith, 1987) and deficits in metacognition (Langland-Hassan et al., 2017; Morin, 2009). Given the functional role of inner speech in cognition and the negative consequences of its impairments, it is imperative that we develop a robust understanding of the cognitive and neural underpinnings of inner speech. In the current paper, we assess two mechanistic models of inner speech and explore how they can be reconciled with the phenomenology of inner speech in an unifying theoretical framework. This framework is examined and verified through an Activation Likelihood Estimation (ALE) meta-analysis of the existing literature.

Two Mechanistic Models of Inner Speech

The Corollary Discharge Model

The corollary discharge model proposes that inner speech is the predicted perceptual consequence of a planned articulatory movement (Jack et al., 2019; Jacobson, 1932; Scott, 2013; Scott et al., 2013; Watson, 1913). The intent to speak generates an efference copy of the articulatory signal, which enters a forward model to predict what the intended articulation would sound like. This prediction is then perceived internally as inner speech. The model is supported by functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) research, which often reports that tasks thought to elicit inner speech (e.g., rhyme judgement, cued speech and metrical stress judgement) activate brain areas related to speech production, particularly the left inferior frontal gyrus (L-IFG) (Aleman et al., 2005; Curcic-Blake et al., 2013; Lurito et al., 2000; Shergill et al., 2001). Using magnetoencephalography (MEG) in a mental imagery task, Tian & Poeppel (2010) show that kinesthetic estimation of articulatory imagery is followed by increased auditory neural activity ~170ms later, favouring the idea that articulatory signals are subsequently transformed into corollary discharge. The articulation-derived corollary discharge is believed to provide the perceptual content of inner speech, which is found to attenuate the impact of matching overt speech on subsequent speech perception (Scott, 2013), and to reduce the amplitude of the N100 event-related potential (ERP) response to overt speech when they are matched in content and time (Jack et al., 2019).

Although the corollary discharge model explains how inner speech can be generated from articulatory signals, it has several limitations. First, the speed of inner speech is significantly faster than overt speech (Mackay, 1981; Netsell et al., 2016), suggesting that inner speech may not be a fully fledged corollary discharge of the intended articulation. Second, inner speech often incorporates voices of others which contain vocal features distinct from one's own (McCarthy-Jones & Fernyhough, 2011). These 'foreign' vocal features cannot be provided solely by corollary discharge as it is physically constrained by one's own articulator. Third, disruption of the articulatory system does not always lead to disrupted inner speech. Individuals unable to articulate due to congenital disorders can accurately perform tasks reliant on inner speech, such as rhyme judgements (Baddeley & Wilson, 1985; Bishop & Robson, 1989; Smith et al., 2009; Vallar & Cappa, 1987). There is also evidence that overt articulation of irrelevant speech (articulatory suppression) does not inhibit inner speech completely (Wheeldon & Levelt, 1995), suggesting that inner speech may not be driven entirely by articulation. Fourth, brain areas associated with speech production are not always activated during inner speech, especially when it occurs more spontaneously than typically seen in task-elicited inner speech experiments (Hurlburt et al., 2016; Yao et al., 2011). In these studies, spontaneous inner speech predominantly activates the auditory perceptual areas, suggesting that it can be generated perceptually without relying on articulatory signals (Barsalou, 2008; Yao & Scheepers, 2011; Yao et al., 2011, 2021).

The Perceptual Simulation Model

One way for inner speech to be generated perceptually is via *perceptual simulation*, as coined by embodied cognition theories (Barsalou, 1999, 2008). In overt speech perception, neurons in the auditory cortex encode perceptual features of speech in distinct firing patterns. These patterns are captured and stored by neurons in association areas, referred to as *conjunctive neurons* (Barsalou et al., 2003; Barsalou, 2008), which allows the former to be reactivated later to simulate the perceptual experience of the original speech (or part thereof) (Barsalou, 2008). As a variety of captured firing patterns accumulate, they can be integrated or remixed to create new patterns and consequently new perceptual experiences. This would allow for a finite number of captured speech experiences to enable perceptual simulations of potentially an infinite number of novel speech experiences (e.g., an imaginary speech by Donald Trump saying*, “Make psychology great again!”*) (Barsalou et al., 2003).

The perceptual simulation model of inner speech neatly addresses the four limitations in the corollary discharge model. First, inner speech can be simulated at a faster timescale than overt speech as it is not constrained by how fast the articulators can physically move (Oppenheim & Dell, 2010). Second, vocal features that cannot be produced entirely by one’s own articulator, such as vocal features of Darth Vader, or the opposite sex whose pitch is outside of one’s own vocal range, can be perceptually simulated. Third, inner speech can be perceptually simulated in those with speech impairments and during articulatory suppression as it presupposes encoded perceptual experiences of speech rather than articulation-derived corollary discharge. Fourth, perceptual simulations of speech do not depend on speech production, which could explain the lack of activations in speech production areas during spontaneous inner speech in silent reading (Yao et al., 2011) and at rest (Hurlburt et al. 2016).

Nevertheless, the perceptual simulation model is also at odds with some empirical findings. For example, several neuroimaging studies did not observe activation of the auditory cortex during inner speech tasks (De Nil et al., 2000; Gulyás, 2001), which implies a lack of perceptual reactivation. Aziz-Zadeh et al. (2005) found that repetitive transcranial magnetic stimulation (rTMS) of the L-IFG disrupted not only overt speech but also inner speech in a syllable counting task, suggesting that at least some types of inner speech depends on speech production regions. Finally, concurrent white noise has been found to either inhibit (Poulton, 1977), or improve performance on inner speech tasks (Wilding & Mohindra, 1980), suggesting varying dependence of inner speech on perception-related mechanisms. These disparate results highlight that neither corollary discharge nor perceptual simulation alone could offer a full mechanistic account of inner speech.

Reconciling Heterogeneous Findings in Inner Speech Research

To reconcile conflicting findings in the inner speech literature, one needs to recognise that inner speech is *not* a homogeneous, uniform phenomenon, but a multi-dimensional, flexible process manifested in a variety of forms (Hurlburt et al., 2013; McCarthy-Jones & Fernyhough, 2011). Recent evidence suggests that the exact forms of inner speech are at least in part determined by task conditions. For example, an fMRI study by Hurlburt et al. (2016) proposes that inner speech *elicited* by an explicit task (e.g., being asked to imagine saying ‘elephant’) is mechanistically different to inner speech generated *spontaneously* (inner speech captured during resting state). In Regions of Interest (ROI) analyses, they showed that task-elicited inner speech was associated with increased activation of the left IFG and decreased activation in Heschl's gyrus, whereas spontaneous inner speech was associated with increased activation in Heschl's gyrus and no significant effects in the left IFG. They argue that explicit inner speech tasks may rely more on speech production and increase cognitive demands. In contrast, spontaneous inner speech seems to rely less on speech production and may be better explained by a perceptual imagery mechanism.

This notion of a production-perception mechanistic divide in inner speech is also demonstrated by Tian et al. (2016). In their fMRI study, they compared neural activations during imagined articulation and imagined hearing of simple syllables. They found that imagined articulation induced greater activity in a frontal-parietal sensorimotor system resembling a corollary discharge network, with activation encompassing regions involved articulation planning (inferior frontal, premotor, supplementary motor regions), forward model estimation (parietal somatosensory regions), and sound reconstruction (superior temporal regions). In contrast, imagined hearing primarily engaged lexico-semantic (mid temporal regions) and episodic memory networks (mid-frontal, intraparietal and mid-temporal regions). Auditory memories would be retrieved from this distributed network before being reassembled in the superior temporal regions to form an internally perceptible sensation. These findings converge on the idea that various forms of inner speech may be flexibly generated by two distinct neurocognitive mechanisms, with one relying on covert speech production (in line with corollary discharge) and the other on memory-based perceptual imagery (in line with perceptual simulation).

Instead of debating between individual mechanisms of corollary discharge or perceptual simulation, integrating them in a dual-mechanistic model could provide much needed flexibility in accounting for the variety of inner speech and reconciling seemingly contradictory empirical findings. For example, the corollary discharge mechanism can efficiently produce inner speech in one's own voice and at one's own will. This type of inner speech is likely to be used in phonological judgement tasks (e.g., determining whether two words rhyme) and to activate brain areas associated with speech production. Conversely, the perceptual simulation mechanism can better explain inner speech spoken by another person as it bypasses one's physical constraints in articulating other people's voices. Tasks that elicit this kind of inner speech should be more likely to engage areas related to speech perception and memory.

Although this dual-mechanistic model is flexible and could theoretically explain all varieties of inner speech, it remains to be tested across a wider range of task conditions beyond those in Tian et al. (2016) and Hurlburt et al. (2016), and to be reconciled with the diverse phenomenology of inner speech (e.g., VISQ-R; Alderson-Day et al., 2018).

Bridging the Mechanisms and the Phenomenology of Inner Speech

There have been several attempts to characterise the phenomenology of inner speech (Clowes, 2007; Hurlburt et al., 2013; Perrone-Bertolotti et al., 2014). One prominent framework is provided by the Varieties of Inner Speech Questionnaire - Revised (VISQ-R) (Alderson-Day et al., 2018), which builds on an earlier version by McCarthy-Jones & Fernyhough (2011). The VISQ-R characterises the quality of inner speech using five factors: *dialogic*, *condensed*, *other people in inner speech, evaluative/critical* and *positive/regulatory*. The *dialogic* factor represents the extent to which the inner speech is a dialogue or a monologue, the *condensed* factor represents the extent to which inner speech is in abbreviated form or of typical structure and *other people in inner speech* reflects whether the voice is of the speaker or of another person. The final factors, *evaluative/critical* and *positive/regulatory* capture whether the inner speech serves evaluative purposes (e.g., thinking about a previous decision), or positive purposes (e.g., using inner speech to calm oneself), respectively. Whilst these dimensions are well motivated by the traditional Vygotskian model of inner speech (i.e., varying by dialogue and condensation) (Vygotsky, 1987), the extent to which they are supported by proposed neurocognitive mechanisms is largely underspecified.

A recent study by Grandchamp et al. (2019) extended on McCarthy-Jones and Fernyhough (2011), along with other theoretical and empirical works, to develop a neurocognitive model of inner speech which varies along three dimensions: *condensation*, *dialogality* and *intentionality*. In this ‘ConDialInt’ model, *condensation* measures the sensorimotor detail in the representation of inner speech, which can range from being fully detailed to being comparatively abstract. *Dialogality* captures the extent to which inner speech takes the form of a dialogue vs. a monologue. *Intentionality* indicates how deliberately or spontaneously inner speech is generated. The model therefore displays significant overlap with the VISQ and VISQ-R whilst also being integrated into a cognitive framework. In describing the underlying neurocognitive mechanisms involved, Grandchamp et al. (2019) propose a hierarchical predictive control scheme that aims to subsume all subtypes of inner speech. In line with the corollary discharge model, the scheme consists of a series of feedforward and feedback connections that integrate a conceptualisation, a phonetic formulation and an articulatory planning component into a hierarchical whole. The ConDialInt model presents a novel and important contribution to the current understanding of inner speech as it brings both phenomenological and neurocomputational findings into a single framework.

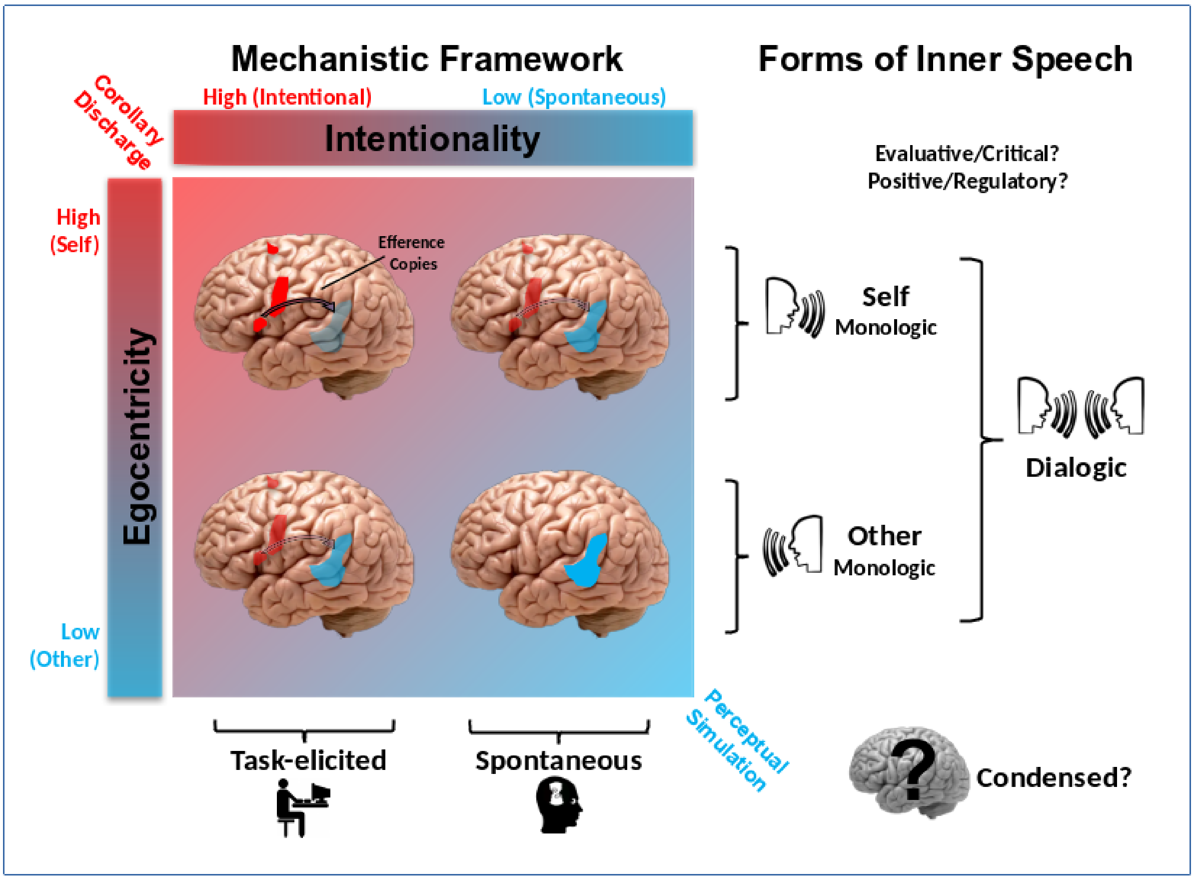
However, some areas of the ConDialInt model remain to be reconciled with existing empirical evidence. For example, as the model is based on the corollary discharge mechanism, it cannot easily explain findings that some inner speech (e.g. spontaneous or in other-voice) does not activate speech production areas (Hurlburt et al., 2016; Raij & Riekki, 2017; Yao et al., 2011). Moreover, the inclusion of a dimension capturing *condensation* introduces further challenges. Conceptually, condensation is unlikely to be orthogonal to the other two dimensions. For example, condensed inner speech is less likely to emerge when the inner speech is highly intentional (i.e. task elicited) as inner-speech elicitation paradigms typically require the generation of richer acoustic, phonological or syntactic information which condensed inner speech can lack (Fernyhough, 2004). Instead, it is plausible that highly condensed inner speech is a phenomenon dependent on the inner speech being spontaneous. Empirically, this correlation between *condensation* and *intentionality* creates challenges in formally manipulating and measuring condensed inner speech (as acknowledged by the authors) as would be needed to comprehensively examine its validity and mechanisms. An additional challenge presented by the dimension of condensation is the difficulty in observing highly condensed and abstract inner speech using current neuroimaging measures, as it is argued to possess only semantic qualities rather than phonological, acoustic or syntactic qualities. While it is possible to record activity of brain regions associated with semantic processing (e.g. anterior temporal lobe; Visser et al., 2010) using neuroimaging methodologies, there would remain significant challenges in distinguishing inner speech activity from other types of non-verbal thought.

The ConDialInt model's dimension of *dialogality* raises further questions which require consideration. First, by categorising inner speech by its dialogicality, the ConDialInt model risks combining types of inner speech with different mechanisms of generation and neural correlates into a single subtype (e.g. corollary discharge and perceptual simulation). For example, inner speech which is highly dialogic could utilise one generative mechanism when the speaker hears his own inner voice, and a different generative mechanism when generating the voice of a second person with distinct vocal characteristics. The interplay between multiple characters inherent to dialogic inner speech also implicates systems such as Theory of Mind (Alderson-Day et al., 2016; 2020), which might play a role in the inner speech experience but do not constitute the speech-like experience, per se. These questions present a hurdle to the empirical investigation of inner speech and could suggest that the dimension of dialogality could be further refined such that the chances of capturing neurocognitively distinct types of inner speech within a single dimension are reduced.

We therefore propose a simpler, more flexible two-dimensional cognitive framework. The original dimension of *condensation* is not included because it is not empirically manipulatable or measurable. The original dimension of *dialogality* is replaced by *egocentricity*. Egocentricity represents the extent to which inner speech can be a recreation of one's own articulation (high egocentricity), or a simulation of another individual's voice (low egocentricity). Egocentricity ensures that mechanistically distinct types of inner speech (i.e. in own-voice & other-voice) are clearly differentiated rather than intermixed, as in the case of dialogic inner speech. Finally, the dimension of *intentionality* is kept unchanged, which captures the extent to which inner speech is elicited by explicit task demands or occurs spontaneously.

Instead of relying on corollary discharge exclusively (Grandchamp et al., 2019), the present framework is dual-mechanistic, additionally incorporating the perceptual simulation mechanism. It straightforwardly predicts the relative contributions of the two mechanisms along the northwest-southeast diagonal of the egocentricity × intentionality space (Figure 1). That is, the more egocentric and intentional inner speech is (i.e. more northwestward in Figure 1), the more strongly it relies on articulation-derived corollary discharge. For example, tasks involving explicit phonological judgements are likely to elicit egocentric and intentional inner speech and to activate speech production areas (Aleman et al., 2005; Lurito et al., 2000). In contrast, the less egocentric or intentional inner speech is (i.e. more southeastward in Figure 2), the more likely it resorts to perceptual simulation. For example, imagery of another person's voice (low egocentricity) and inner speech that emerges spontaneously (low intentionality) would preferentially recruit the perceptual simulation mechanism, given the physical constraints in articulating voices of others and/or a lack of explicit intentions. These types of inner speech are more likely to activate temporal auditory cortices (Alderson-Day et al., 2016; Hurlburt et al., 2016; Marvel & Desmond, 2012; Yao et al., 2011) and may be modulated by oscillatory activity in the auditory cortex rather than in articulatory regions (Yao et al., 2021).

Figure 1. Schematic illustration of the egocentricity × intentionality framework, with its underpinning neural mechanisms and its hypothetical mapping with phenomenological qualities.



Note: Red colour represents the involvement of the corollary discharge mechanism. Blue colour represents the involvement of the perceptual simulation mechanism. Highlighted brain areas on the four brains indicate which and to what extent brain areas would be activated in the four quadrants of this framework (colour saturation levels indicate the strengths of involvement/activation along the northwest-southeast diagonal). Brain areas in red represent speech production (planning) regions including the left inferior frontal gyrus, left premotor cortex and supplementary motor area. Regions in blue represent speech perception regions in the left superior temporal cortex. The black arrow indicates the pathway along which efference copies are sent from the production areas to the perception areas. The fronto-parieto-temporal memory network is expected to be involved in the perceptual simulation mechanism but is not drawn to keep the illustration simple and tidy.

In addition to predicting mechanistic involvement across different inner speech tasks, the proposed framework could also bridge the corollary discharge and perceptual simulation mechanisms and the phenomenological qualities identified in VISQ-R (Alderson-Day et al., 2018). For example, high- and low-egocentric inner speech would be phenomenologically perceived as self- and other-monologic inner speech, respectively. The intermix of the two would support dialogic inner speech. While the dimensions of *dialogicality* and *other people* are more concerned with the sensorimotor features and the agency of inner speech, the dimensions of *evaluative/critical* and *positive/regulatory* primarily capture its cognitive functions. Although the kinds of inner speech used for these functions are yet to be empirically studied, they can nevertheless be represented along the egocentricity and intentionality dimensions. However, the dimension of *condensation* is not explicitly considered in the current framework as it likely correlates with egocentricity and intentionality and cannot be easily manipulated or objectively observed.

Aims & Hypotheses

To verify the proposed framework and its predicted neuroanatomical underpinnings, the present study carried out an Activation Likelihood Estimation (ALE) meta-analysis of the existing neuroimaging literature. Functional activation coordinates are compiled across multiple studies, to identify which brain regions are *consistently* activated as inner speech varies along egocentricity and intentionality. This kind of convergence analysis is more likely to reveal neural correlates inherent to inner speech, as it is less skewed by peripheral processes introduced by specific paradigms (e.g., increased working memory, verbal monitoring, or Theory of Mind). More importantly, it enables us to verify the distinct mechanisms of corollary discharge and perceptual simulation across a wider range of paradigms beyond the studies by Hurlburt et al. (2016) and Tian et al. (2016).

We hypothesised that corollary discharge and perceptual simulation would be differentially engaged to produce a variety of inner speech. Inner speech would primarily engage the corollary discharge mechanism at higher egocentricity and/or intentionality, and rely more on the perceptual simulation mechanism as egocentricity and/or intentionality decrease. In an ALE analysis, we predicted that inner speech with high egocentricity and high intentionality would be associated with more consistent activations in speech production areas. These speech production areas primarily include the L-IFG, the left premotor cortex (L-PMC) and the supplementary motor area (SMA) (Booth et al., 2003; Lurito et al., 2000). Within the L-IFG, we expected greater activation of the *pars opercularis* subregion *(BA44)*. This is because of previous work implicating the *pars opercularis* in phonological processing (Burton et al., 2005) and speech production (Tourville & Guenther, 2011), as well as it serving a putative role in articulatory planning and efference copy generation in previous studies of inner speech (Molnar-Szakacs et al., 2005; Tian et al., 2016). We also predicted involvement of the left superior temporal sulcus and gyrus (L-STS / L-STG) as the terminus of corollary discharge (Tian et al., 2016; Tourville & Guenther, 2011). Inner speech at low egocentricity and/or low intentionality would be associated with more consistent activations primarily in the L-STG/STS but also in the episodic memory network, including the left medial temporal gyrus (L-MTG), the left medial frontal gyrus

(L-MFG) and the superior parietal lobe/precuneus (L-SPL/PC) (Hurlburt et al., 2016; Kleider-Offutt et al., 2019; Linden et al., 2011; Tian et al., 2016).

Materials and Methods

Literature search

The search was planned and conducted in line with PRISMA guidelines for meta-analyses and systematic reviews (Moher et al., 2009). The literature search was conducted using three electronic databases during May 2021 (Pubmed, Web of Science, Scopus) using the search query *("magnetic resonance imaging" OR "mri" OR "fmri" OR "positron emission tomography" OR "pet") AND ("inner speech" OR "auditory imagery" OR "covert speech" OR "speech imagery" OR "inner voice" OR "inner experience")*. Searches were limited to publications mentioning these terms within the title, abstract or author keywords. No further search criteria (e.g. date of publication) was utilised. This yielded 598 results, with 274 remaining after duplicates were removed. Manual searches of the reference sections of resulting articles were conducted in order to include relevant studies which were not captured by the search terms, this yielded a further 22 relevant studies which underwent screening along with the 274 studies, resulting in a total of 296 studies being screened.

Eligibility criteria

Figure 2. Flow diagram of literature review process. Notation boxes in the *screening* section represent the various reasons for study exclusion and the number of studies excluded.

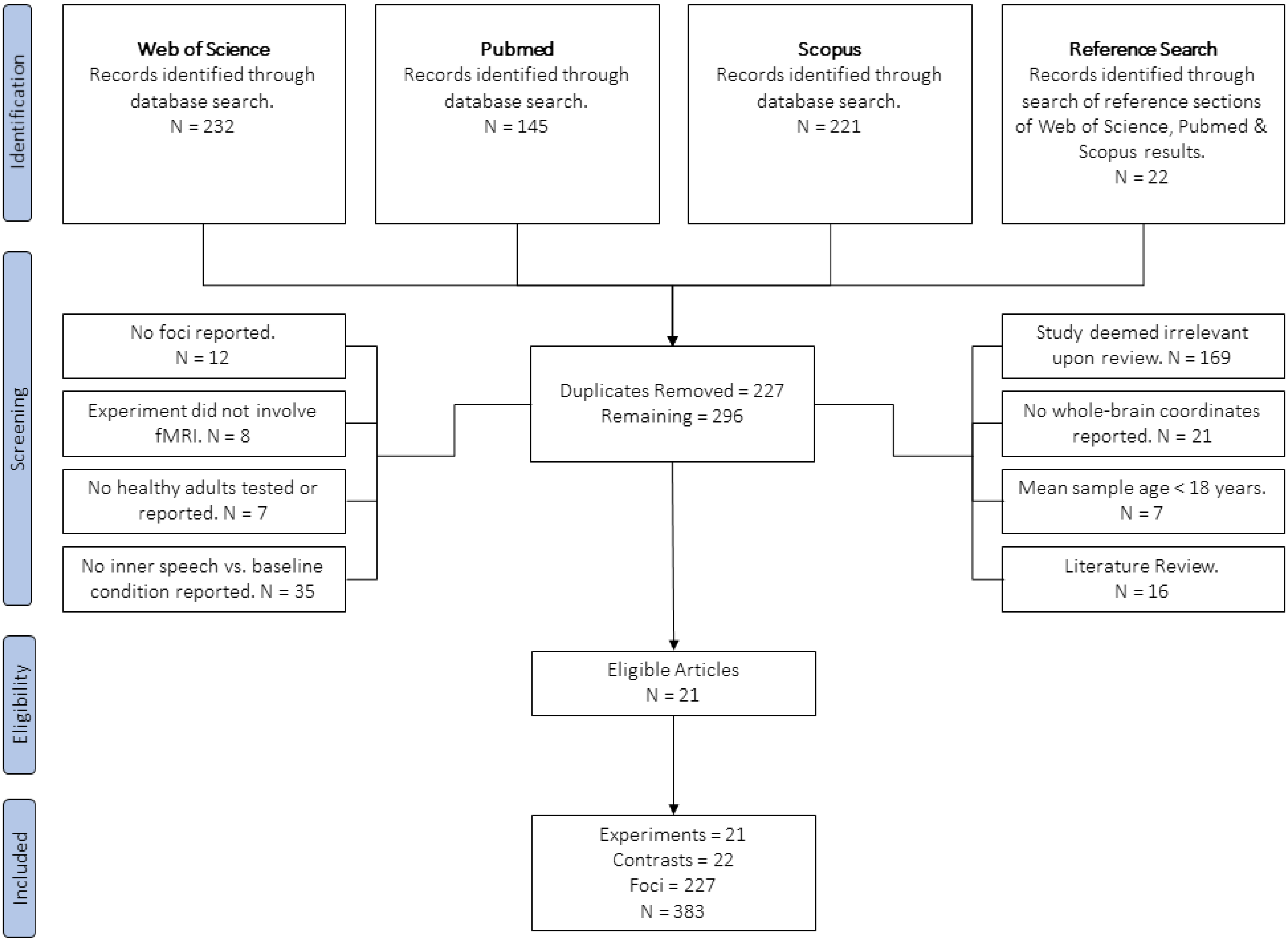


Table 1. List of eligible studies which were included in the meta-analysis.

Study	N	Foci	Task	Baseline	Contrast	Handedness	Intentionality	Egocentricity
Lurito et al. (2000)	5	19	Rhyme judgement	Fixation symbol	Rhyme judgement > Fixation symbol	Right	High	High
Booth et al. (2003)	12	4	Rhyme judgement	Symbol matching	Rhyme judgement > Symbol matching	Right	High	High
Rudner et al. (2005)	12	16	Rhyme judgement	Rest	Rhyme judgement > Rest	Right	High	High
Aparacio et al. (2007)	12	2	Rhyme judgement	String matching	Rhyme judgement > String matching	Right	High	High
MacSweeney et al. (2009)	7	6	Rhyme judgement	Same picture judgement	Rhyme judgement > Same picture judgement	Right	High	High
Ćurčić-Blake et al. (2013)	31	9	Metrical stress judgement	Fixation symbol	Metrical stress judgement > Fixation symbol	Right	High	High
Hernandez et al. (2013)	16	11	Rhyme judgement	Font matching	Rhyme judgement > Font matching	Right	High	High
Aleman et al. (2005)	6	14	Imagine other person reading word and determine metrical stress	Fixation symbol	Imagine voice > Fixation symbol	Right	High	Low
Linden et al. (2011)	7	13	Imagine voice of familiar person	Rest	Imagine voice > Rest	Right	High	Low
Grandchamp et al. (2019)	24	24	Imagine other voice	N.S	Voice imagery > N.S*	Right	High	Low

Kleider-Offutt et al. (2019)	28	9	Read sentence in a learned voice	Rest	Read sentence > Rest	Not specified	High	Low
De Nil et al. (2000)	10	5	Mentally read presented 3-syllable words	Fixation symbol	Read words > Fixation symbol	Right	High	High
Gulyás (2001)	10	1	Internally list alphabet & first verse of national anthem	Rest	Internal listing > Rest	7 Right, 2 Left, 1 Ambidextrous.	High	High
Theys et al. (2020)	11	16	Read visually presented pseudoword	View character string	Silent reading > View character string	Right	High	High
Papathanassiou et al. (2000)	8	19	Generate verbs related to given noun	Rest	Generate verb > Rest	Right	High	High
Wilson et al. (2011)	26	10	Generate words beginning with given letter	Rest	Word generation > Rest	Right	High	High
Alderson-Day et al. (2020)	21	10	Direct quotation speech	Fixation symbol	Direct quotation > Fixation symbol	Right	Low	High
Yao et al. (2011)	16	16	Direct quotation speech	Fixation symbol	Direct quotation > Fixation symbol	15 Right, 1 Left	Low	High
Raij et al. (2017)	51	8	Verbal thought (Direct Experience Sampling)	Fixation symbol	Verbal thought > Fixation symbol	Not specified	Low	High
Grandchamp et al. (2019)	24	16	Use joystick to report verbal mind wandering	N.S	Verbal Mind Wandering > N.S*	Right	Low	High
Okada et al. (2017)	21	7	Silently read the visually presented tongue twister	Fixation symbol	Tongue twister > Fixation symbol	Right	High	High

Okada et al. (2018)	21	13	Silently articulate the visually presented tongue twister	Rest	Tongue twister > Rest	Right	High	High

Contrast Selection & Grouping

Contrasts which compared inner speech to a baseline were selected. For the majority of the studies, these were either an *inner speech* > *rest* or an *inner speech* > *fixation symbol* contrast. Four additional studies utilised a baseline in which participants matched visual symbols (Aparicio et al., 2007; Booth et al., 2003; Hernandez et al., 2013; MacSweeney et al., 2009). Given that visual matching is not known to elicit inner speech, the *inner speech* > *visual matching* contrasts were also included in the analyses.

Studies were then grouped based on their egocentricity and intentionality *a priori*. Studies were allocated to high and low egocentricity groups, based on whether the paradigm required participants to generate inner speech in their own voice (high egocentricity) or in another person's voice (low egocentricity). This yielded 16 in the high egocentricity group and 4 in the low egocentricity group.

Within the dimension of intentionality, studies in which participants were required, explicitly or implicitly, to generate inner speech were classified as high intentionality studies. For example, De Nil et al. (2000) asked participants to internally read single words and Hernandez et al. (2013) asked participants whether pairs of visually presented words rhymed. Studies in which inner speech occurred spontaneously, either in tasks not reliant on inner speech or in the resting state were classified as low intentionality. For example, studies by Yao et al. (2011) and Alderson-Day et al. (2020) used a reading comprehension task, which does not require the use of inner speech to complete. Inner speech in these tasks emerges from spontaneous perceptual simulations of literary characters when reading direct quotations. Research by Hurlburt et al. (2016) also examined low intentionality inner speech but adopted a different approach. Participants were asked to report their internal state in the moments preceding the sounding of random auditory beeps. fMRI analysis then focused on the moments in which participants reported that they were engaging in inner speech. The division of studies by intentionality yielded 14 studies in the high intentionality group and 4 studies in the low intentionality group. Given that no studies were found which examined inner speech which was low in both intentionality and egocentricity (i.e. spontaneous inner speech in other voices), we could not group studies into the four unique quadrants of the two dimension model.

It is worth noting that the numbers of included studies were unbalanced between the groups defined above. This was primarily because a disproportionately large number of studies used phonological judgement tasks such as rhyme judgement tasks (32%). To ensure our contrasts are not significantly skewed by overrepresented paradigms like rhyme judgements, we ran one set of analyses on the '*unbalanced*' dataset, and re-ran the analyses on a sub-dataset where the numbers of studies were balanced across paradigms. This '*balanced'* dataset contained 2 studies per paradigm-type, with a total of 14 experiments split across 7 paradigm categories (allocations in Appendix). When a particular paradigm-type was employed by more than 2 studies, the experiments with the largest sample sizes were selected. The 7 paradigm-types were: (1) other voice imagery, (2) tongue twister imagery, (3) mind wandering, (4) direct quotation reading, (5) word generation, (6) phonological judgement, (7) single word reading.

Within the results section, analysis using all included studies was labelled as the *unbalanced* dataset. Analysis of the paradigm-adjusted dataset was labelled as the *balanced* dataset.

Activation Likelihood Estimation

Activation likelihood estimation (ALE) analysis was carried out using the BrainMap GingerALE tool, version 3.0.2 (www.brainmap.org). ALE analysis compiles reported activation coordinates across multiple fMRI studies to identify which brain regions are most likely associated with a cognitive task (Eickhoff et al., 2012, Turkeltaub et al., 2012). All MNI coordinates were converted to Talairach space using the icbm2tal transformation implemented in GingerALE (Lancaster et al., 2007). ALE analysis of the *unbalanced* and *balanced* datasets used a cluster-forming threshold of $p < 0.001$ (uncorrected, 1000 permutations), and a cluster-level family-wise error (FWE) corrected threshold of $p < 0.05$, as recommended by Müller et al. (2018). Because the subgroups divided by egocentricity and intentionality each had relatively fewer number of studies, the ALE analysis of the subgroups used a more liberal cluster-forming threshold of $p < 0.01$ (uncorrected, 1000 permutations) and a cluster-level family-wise error (FWE) corrected threshold of $p < 0.05$. The more liberal threshold of $p<.01$ is appropriate for smaller sample sizes, and has been adopted by previous ALE studies (Di et al., 2017; Falcone & Jerram, 2018; Ruiz Vargas et al., 2016).

Given the low number of studies in the low egocentricity (N=4) and low intentionality (N=4) conditions, both in absolute terms and relative to their high egocentricity/intentionality counterparts (N=16 and 14, respectively), we adhered to GingerALE recommendations and did not run any contrast or conjunction analyses. The resulting ALE maps were rendered in MRIcroGL V1.2.2 (https://www.nitrc.org/projects/mricrogl/) with anatomical labelling of significant clusters and peaks being automatically calculated by GingerALE using the Talairach Daemon (http://talairach.org/) and exported to a spreadsheet.

Publication Bias Check: Fail-safe N Analysis

To evaluate how robust the ALE results are against publication bias (i.e. null results not being published, also known as the ‘file-drawer effect’), a fail-safe N analysis was conducted on all datasets. This consists of re-running the GingerALE analysis whilst iteratively adding an increasing number of randomly-generated null-result studies (Acar et al., 2018). The fail-safe N is calculated per ALE cluster. Its value represents the highest number of null studies that can be added to a dataset whilst maintaining the significance of the cluster. Null-result experiments were generated in R, version 4.0.5 (https://www.r-project.org/) using the GenerateNull script (https://github.com/NeuroStat/GenerateNull; as used in Acar et al., 2018). The R script creates a pre-specified number of null-studies matched for the number of participants and foci contained within the real experiment list. Foci within the generated null-studies are distributed randomly throughout the grey matter. Given that there is an estimated upper bound of 30 unpublished studies with null findings per 100 published neuroimaging studies investigating language (Samartsidis et al., 2020), we re-analysed the *unbalanced* pooled dataset (N=22) with up to 7 additional null studies (30%) and re-analysed the *balanced* pooled dataset (N=14) with up to 4 additional null studies (28.6%). Analysis of the datasets divided by egocentricity and intentionality were also re-analysed using the following additional null studies for the unbalanced versions: low egocentricity (Nnull=1; 25%), high egocentricity (Nnull=5; 31.3%), low intentionality (Nnull=1; 25%), high intentionality (Nnull=4; 28.6%). The balanced versions were re-analysed using the following additional null studies: low egocentricity (Nnull=1; 25%), high egocentricity (Nnull=3; 30%), low intentionality (Nnull=1; 25%), high intentionality (Nnull=2; 25%). The clusters which survive the significance thresholds after the addition of ~30% null studies are considered robust against potential file drawer effects.

Outlier Check: Jackknife Analysis

The fail-safe N analysis was complemented by a jackknife analysis to cross-validate that the observed results were not driven by any single study in the dataset (Amanzio et al., 2013; Shao & Tu, 1995). This involved repeatedly re-running the analysis whilst excluding a single, different study each time. The results were then visually analysed and compared to the clusters produced in the original analysis in convergence coordinates and cluster size. Each cluster was scored as a percentage, which represents the proportion of analysis iterations in which the convergence was replicated. Clusters which were present in over 80% of the iterations were considered robust (Yaple & Yu, 2020).

Results

ALE Clusters for Inner Speech - All Studies

The upper panel of Figure 3 shows the ALE results on the *unbalanced* dataset, illustrating the brain areas that displayed significant convergence across all included studies. The associated Talairach coordinates are presented in the ‘unbalanced’ section of Table 2. In total, six clusters were identified. The largest cluster was centred at the left medial frontal gyrus / supplementary motor area (Brodmann Area 6; BA6) and extended across the left superior frontal gyrus (BA6) and left cingulate gyrus (BA24). Additional clusters were centred on the left precentral gyrus (BA6 & BA44), left inferior frontal gyrus (BA45), right insula (BA13) and right culmen.

The lower panel of Figure 3 shows the ALE results on the *balanced* dataset in which all paradigm types were represented by an equal number of studies with the largest sample sizes. The associated Talairach coordinates are presented in the ‘balanced’ section of Table 2. In total, three clusters showed significant convergence. The largest cluster was centred at the left medial frontal gyrus / supplementary motor area (BA6) and extended across the left superior frontal gyrus (BA6) and left cingulate gyrus (BA24). Two smaller clusters were centred on the right insula (BA13) and right culmen.

Figure 3. Areas showing significant ALE statistic across all studies shown at FWE p < 0.05 at the cluster-level.

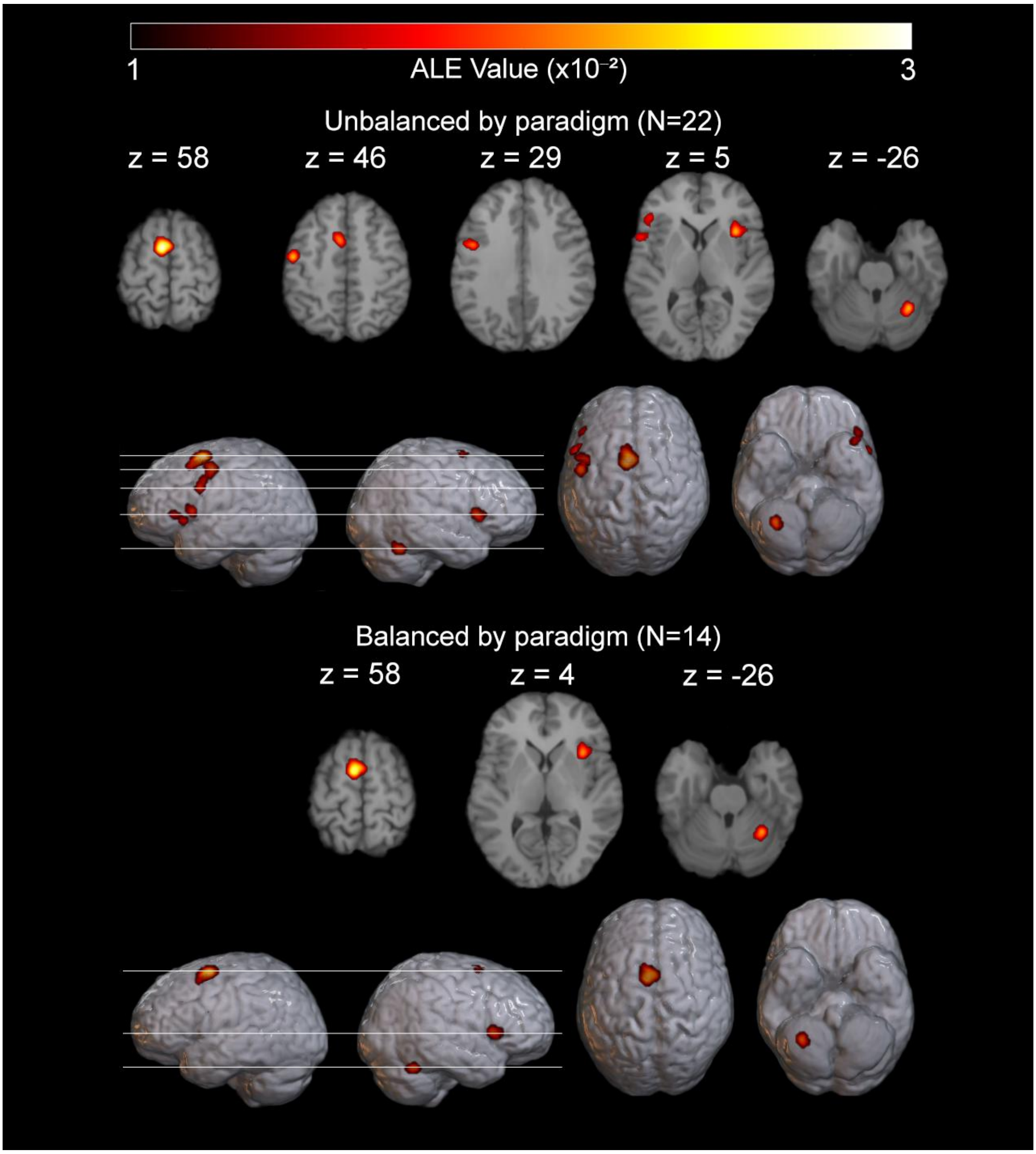


Table 2. Significant clusters across all studies in the balanced and unbalanced datasets.

Cluster	Area		BA	x	y	z	ALE (x10^{-2})	Volume (mm^3)	Fail-Safe N	Jackknife
					Unbalanced by Paradigm (N=22)					
1	**Medial Frontal Gyrus / Supplementary Motor Area**	L	BA6	-5	2	54	3.09	3608	≥7*	100%*
2	Precentral Gyrus	L	BA6	-46	-4	39	2.13	2168	≥7*	100%*
3	**Insula**	R	BA13	35	15	5	2.14	1280	≥7*	95%*
4	**Culmen**	R	-	29	-57	-28	2.17	936	≥7*	95%*
5	Precentral Gyrus	L	BA44	-50	9	8	1.77	840	5	82%*
6	Inferior Frontal Gyrus	L	BA45	-43	20	2	1.63	744	0	73%
					Balanced by Paradigm (N=14)					
1	Medial Frontal Gyrus / Supplementary Motor Area	L	BA6	-5	1	55	2.91	2824	≥4*	100%*
2	Insula	R	BA13	34	15	4	2.04	1128	≥4*	100%*
3	Culmen	R	-	29	-56	-27	2.08	824	≥4*	71%

Note: Coordinates (x,y,z) represent the location of peak ALE statistic per cluster in Talairach space. The Unbalanced subsection shows the results of the dataset that included all eligible studies. The Balanced subsection shows the results of the dataset with an equal number of studies for each included paradigm-type. Area names in bold font represent regions whose activation was replicated in the balanced dataset. Asterixes in Fail-Safe N and Jackknife signify sufficient robustness against publication bias and outliers, respectively, as defined in the Methods section.

Inner Speech As a Function of Egocentricity

The upper panel of Figure 4 shows brain areas which displayed significant convergence for high egocentricity and low egocentricity studies, respectively, in the *unbalanced* dataset. Their Talairach coordinates are reported in the ‘unbalanced’ section of Table 3. The *High Egocentricity* studies converged on five clusters (coloured in red). The largest cluster was centred on the left precentral gyrus (BA6) and extended across the left inferior frontal gyrus and middle frontal gyrus. Additional clusters encompassed the left medial frontal gyrus / supplementary motor area (BA6), left precentral gyrus (BA44), right culmen and right insula (BA13). The *Low Egocentricity* studies converged on a single cluster centred on the right insula (BA13) and extending across the right inferior frontal gyrus (BA44) (coloured in blue).

The lower panel of Figure 4 shows brain areas which displayed significant convergence for high egocentricity and low egocentricity studies, respectively, in the *balanced* dataset. The associated Talairach coordinates are presented in the ‘balanced’ section of Table 3. The *High Egocentricity* studies converged on five clusters (coloured in red). The largest cluster was centred on the left medial frontal gyrus / supplementary motor area (BA6). Two additional clusters centred on the left precentral gyrus (BA6 & BA44), one of which also encompassed the left inferior frontal gyrus. The final clusters were centred on the right insula (BA13) and right culmen. The *Low Egocentricity* studies converged a single cluster centred on the right insula (BA13) and extending across the right inferior frontal gyrus (BA44).

Figure 4. Areas showing significant ALE scores in High Egocentricity (red) and Low Egocentricity studies (blue) at FWE p < 0.05 at the cluster-level.

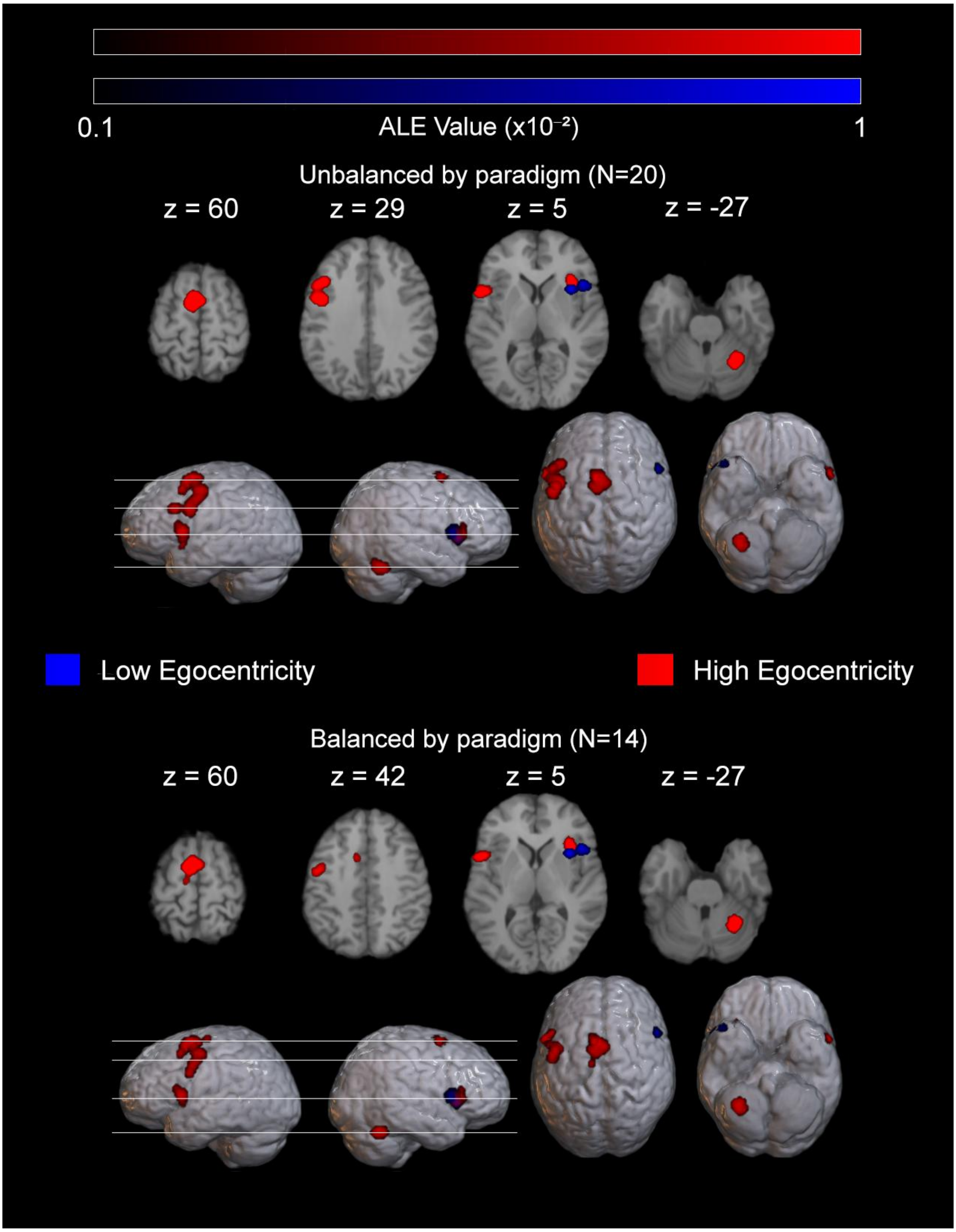


Table 3. Clusters showing significant ALE statistic across High Egocentricity and Low Egocentricity studies, respectively. Shown at FWE $p < 0.05$ at the cluster-level.

Cluster	Area		BA	x	y	z	ALE ($x10^{-2}$)	Volume (mm³)	Fail-Safe N	Jackknife
					Unbalanced by Paradigm (N=20)					
					High Egocentricity					
1	Precentral Gyrus	L	BA6	-45	1	36	1.82	6128	≥4*	100%*
2	Medial Frontal Gyrus / Supplementary Motor Area	L	BA6	-5	2	53	2.85	5376	≥4*	100%*
3	Precentral Gyrus	L	BA44	-51	9	8	1.76	2536	≥4*	100%*
4	Culmen	R	-	29	-57	-26	2.13	2064	≥4*	75%
5	Insula	R	BA13	33	17	5	1.60	1808	≥4*	44%
					Low Egocentricity					
1	Insula	R	BA13	43	13	6	1.13	2712	≥1*	50%
					Balanced by Paradigm (N=14)					
					High Egocentricity					
1	Medial Frontal Gyrus / Supplementary Motor Area	L	BA6	-5	1	54	1.05	5208	≥3*	100%*
2	Precentral Gyrus	L	BA6	-45	-4	40	1.56	3192	≥3*	100%*
3	Precentral Gyrus	L	BA44	-50	9	9	1.61	2072	≥3*	80%*
4	Insula	R	BA13	33	18	6	1.60	1984	≥3*	70%
5	Culmen	R	-	30	-56	-26	2.08	1824	≥3*	70%

					Low Egocentricity					
1	Insula	R	BA13	43	13	6	1.13	2712	≥1*	50%

Note: Coordinates (x,y,z) represent the location of peak ALE statistic per cluster in Talairach space. The Unbalanced subsection shows the results of the dataset that included all eligible studies. The Balanced subsection shows the results of the dataset with an equal number of studies for each included paradigm-type. All activation observed in the unbalanced dataset was replicated in the balanced dataset. Asterixes in Fail-Safe N and Jackknife signify sufficient robustness against publication bias and outliers, respectively, as defined in the Methods section.

Inner Speech As a Function of Intentionality

The upper panel of Figure 5 shows brain areas which displayed significant convergence for high intentionality and low intentionality studies, respectively, in the *unbalanced* dataset. The Talairach coordinates associated with intentionality analyses are presented in the ‘unbalanced’ section of Table 4. *The High Intentionality* studies converged on four clusters. The largest cluster was centred on the left precentral gyrus (BA6) and extended across the left inferior frontal gyrus and middle frontal gyrus. The three additional clusters were centred on the left medial frontal gyrus / supplementary motor area (BA6), left precentral gyrus (BA44) and right culmen / declive. The *Low Intentionality* studies converged on a single cluster centred on the left middle temporal gyrus and extending across the left superior temporal gyrus.

The lower panel of Figure 5 shows brain areas which displayed significant convergence for high intentionality and low intentionality studies, respectively, in the *balanced* dataset. The associated Talairach coordinates are presented in the ‘balanced’ section of Table 4. The *High Intentionality* studies converged on two clusters. The largest cluster was centred on the left medial frontal gyrus / supplementary motor area (BA6), with an additional cluster centred on the left precentral gyrus (BA6). The Low Intentionality studies converged a single cluster centred on the left middle temporal gyrus (BA22) and extending across the left superior temporal gyrus (BA21).

Figure 5. Areas showing significant ALE scores in High Intentionality (red) and Low Intentionality studies (blue) at FWE p < 0.05 at the cluster-level.

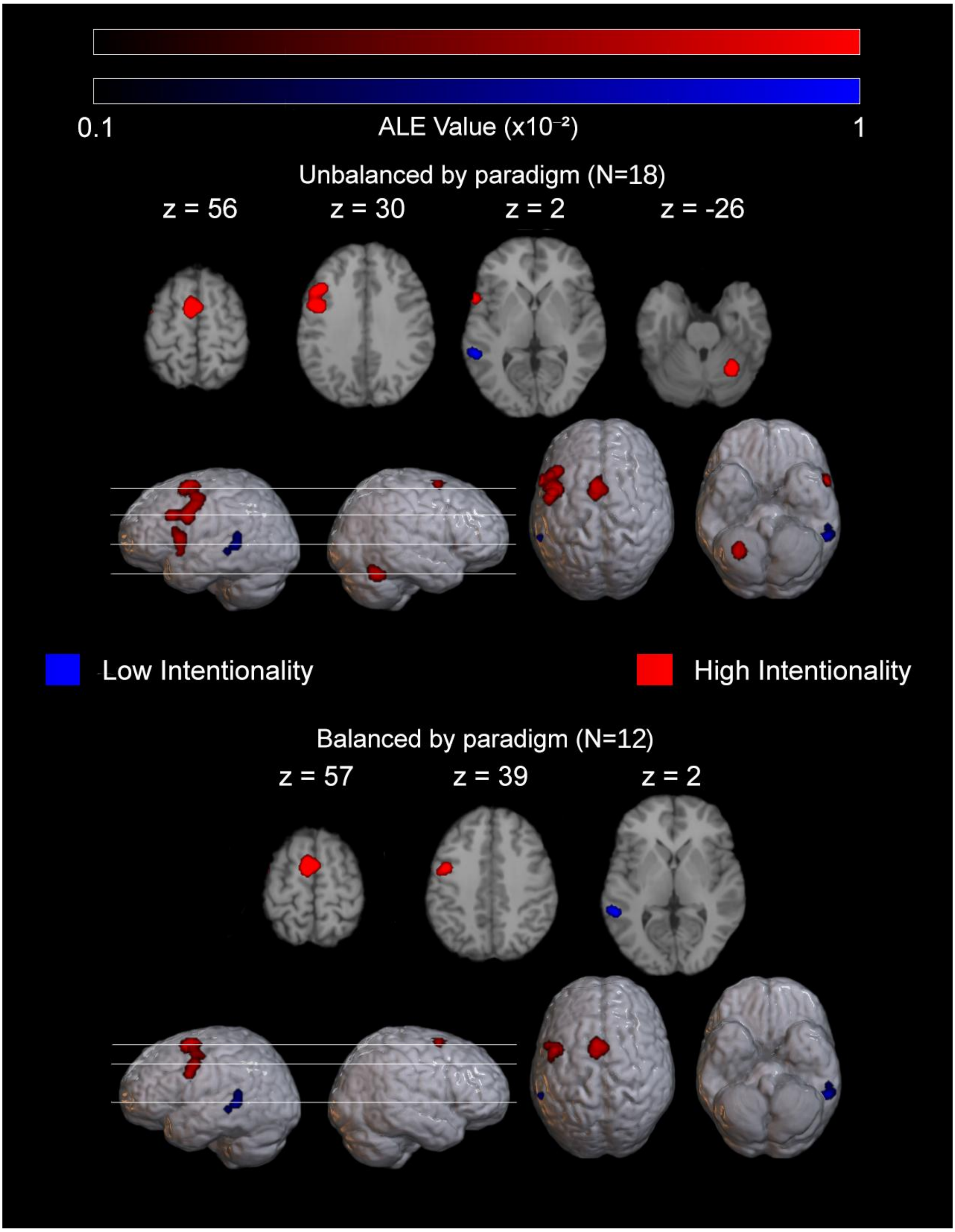


Table 4. Clusters showing significant ALE statistic across High Intentionality and Low Intentionality studies, respectively. Shown at FWE $p < 0.05$ at the cluster-level.

Cluster	Area		BA	x	y	z	ALE (x10^{-2})	Volume (mm^3)	Fail-Safe N	Jackknife
Unbalanced by Paradigm (N=18)										
High Intentionality										
1	**Precentral Gyrus**	L	BA6	-45	3	34	1.82	6320	≥4*	100%*
2	**Medial Frontal Gyrus / Supplementary Motor Area**	L	BA6	-4	1	54	2.15	4152	≥4*	100%*
3	Precentral Gyrus	L	BA44	-52	10	7	1.26	2280	≥4*	100%*
4	Culmen	R	-	28	-58	-26	1.71	1864	≥4*	86%*
Low Intentionality										
1	**Middle Temporal Gyrus**	L	BA22	-56	-43	3	1.03	1440	0	25%
Balanced by Paradigm (N=12)										
High Intentionality										
1	Medial Frontal Gyrus / Supplementary Motor Area	L	BA6	-4	0	55	1.09	3712	≥2*	100%*
2	Precentral Gyrus	L	BA6	-45	-2	38	1.35	3064	≥2*	100%*
Low Intentionality										
1	Middle Temporal Gyrus	L	BA22	-56	-43	3	1.03	1440	0	25%

Note: Coordinates (x,y,z) represent the location of peak ALE statistic per cluster in Talairach space. The Unbalanced subsection shows the results of the dataset that included all eligible studies. The Balanced subsection shows the results of the dataset with an equal number of studies for each included paradigm-type. Area names in bold font represent *regions whose activation was replicated in the balanced dataset. Asterixes in Fail-Safe N and Jackknife signify sufficient robustness against publication bias and outliers, respectively, as defined in the Methods section.*

Discussion

Building on the phenomenological variety of inner speech (Alderson-Day et al., 2018; Grandchamp et al., 2019) and the suggestions of two underlying neural mechanisms (Hurlburt et al., 2016; Tian et al., 2016), the current paper aimed to unify phenomenology and neural mechanisms in a two-dimensional framework. It proposes that inner speech can cognitively vary by egocentricity (in self-voice vs. other-voice) and intentionality (intentional vs. spontaneous), which has the potential to bridge the phenomenological qualities of inner speech (except for condensation) with distinct neurocognitive mechanisms of corollary discharge and perceptual simulation. Specifically, it is hypothesised that inner speech would primarily engage the corollary discharge mechanism at higher egocentricity and/or intentionality, and rely more on the perceptual simulation mechanism as egocentricity and intentionality decrease. Although not directly tested in the present paper, the framework also illustrated that phenomenological qualities of dialogicality, evaluative/critical and positive/regulatory and elicitation methods of inner speech could in principle be accounted for along the dimensions of egocentricity and intentionality.

To validate the utility of this framework, we carried out an ALE meta-analysis to identify neural correlates that converged (1) across all inner speech paradigms and (2) across inner speech at opposite ends of egocentricity and intentionality, respectively. An ALE analysis of all available studies found significant convergence on the left medial frontal gyrus / supplementary motor area (L-MFG / L-SMA; BA6), left precentral gyrus (L-PCG; BA6 & BA44), right insula (R-Insula, BA13), right culmen (R-Culmen) and left inferior frontal gyrus (L-IFG; BA44 / BA45). However, after adjusting the number of studies by paradigm, convergence was only observed over the left medial frontal gyrus / supplementary motor area (L-MFG / L-SMA; BA6), right insula (R-Insula, BA13) and right culmen (R-Culmen), supporting the hypothesis that distinct neural mechanisms could be involved in different forms of inner speech. Specifically, *High Egocentricity* inner speech converged on L-MFG / L-SMA (BA6), L-PCG (BA6 & BA44), R-Insula (BA13) and R-Culmen, whereas *Low Egocentricity* inner speech converged on R-Insula (BA13) only. *High Intentionality* inner speech converged on the L-MFG / L-SMA (BA6), L-PCG (BA6), L-PCG (BA44) and R-Culmen, with only the L-MFG / L-SMA (BA6), L-PCG (BA6) converging after sample size adjustments by paradigms. *Low Intentionality* inner speech converged on the left middle temporal gyrus (L-MTG; BA22) with a substantial portion of the cluster (> 25%;

Supplementary Data) covering the left superior temporal gyrus (L-STG; BA22). Despite a relatively low number of studies available in the inner speech literature, our analyses are robust and sensitive, having accounted for the unbalanced number of studies by paradigm, and cross-validated the results against file drawer effects and outlier studies.

Inner Speech - All Studies

The brain regions which showed significant convergence across all inner speech studies were regions broadly associated with overt speech production. The L-PCG and L-MFG / L-SMA encompass the primary and secondary motor areas and the L-IFG is typically reported in speech production tasks (Frankford et al., 2019). Convergence within these areas therefore suggests that some form of motor planning occurs during the generation of inner speech, as proposed by the corollary-discharge model of inner speech. However, given the proximity of the L-MFG cluster to regions associated with hand/finger movement (Amiez & Petrides, 2014), convergence in L-MFG could also reflect finger movements related to button presses (a common feature in many inner speech tasks), rather than a process inherent to inner speech. The convergence on R-Insula is somewhat unexpected, as it is not commonly discussed within the context of inner speech generation. Nevertheless, other research suggests that the insula could be involved in articulation and could be part of the corollary discharge circuit. For example, a study of macaques showed that stimulation of the insula triggers orofacial motor programmes such as chewing, mouthing, lip smacking and swallowing (Jezzini et al., 2012). In humans, speech production research (Oh et al., 2014) and lesion symptom research (Cereda et al., 2002; Dronkers, 1996; Duffau et al., 2001; Starkstein et al., 1988) has also causally associated the insula with articulation. However, other studies have suggested that activation of the insula reflects interoceptive processing (Marvel & Desmond, 2012; Modinos et al., 2009; Morin & Hamper, 2012). Interoception refers to the processes which underlie self-awareness: such as the detection, filtering and integration of information regarding one's own body (Craig, 2009). Thus, insular involvement in inner speech could represent increased self-awareness associated with inner speech (Morin & Hamper, 2012; Morin & Michaud, 2007; Morin, 2009).

It is worth noting that the above results (N=22) could be skewed heavily towards a select number of paradigms - with phonological judgement elicitation paradigms accounting for more than 30% of studies included. Once the number of studies was balanced between paradigm types (N=14), convergence over L-PCG (BA6 & BA45) and L-IFG (BA44), areas typically associated with speech production and inner speech, was no longer observed. The absence of L-IFG and L-PCG was unlikely to be caused by a mere lack of statistical power, as an ALE analysis of a smaller sample of phonological judgement studies (N=7) observed significant convergence over a cluster encompassing both the L-PCG and L-IFG (Appendix). On balance, the results are likely to reflect a genuine lack of convergence in these areas across different paradigms, as supported by their poor fail-safe N scores. This evidence (or the lack thereof) aligns with the argument made by Hurlburt et al. (2016), Tian et al. (2016) and in this paper, that mechanisms in addition to corollary discharge must be considered when modelling different forms of inner speech.

However, we did not observe significant convergence within the speech perception areas either, which seemed to contradict the hypothesised involvement of perceptual simulation in inner speech. The lack of perceptual convergence may reflect differential levels of perceptual simulation along egocentricity and intentionality. At higher ends of these dimensions, inner speech may be generated predominantly by corollary discharge, which may also attenuate neural activity within speech perception areas (Ford et al., 2021; Hurlburt et al., 2016; Leube et al., 2010; Shergill et al., 2013). At lower ends of these dimensions, perceptual simulation is more strongly engaged and activation is more likely to converge in speech perception areas.

Inner Speech As a Function of Egocentricity

We hypothesised that high egocentricity inner speech would primarily activate the corollary discharge mechanism while low egocentricity inner speech would the perceptual simulation mechanism. We predicted that the former would be associated with more consistent activations in speech production areas such as the L-IFG, L-PMC and SMA, whereas the latter would be associated with converging activations in speech perception areas (e.g., L-STG/STS) and in the memory network (e.g., L-MTG, L-MFG, L-SPL/PC).

The ALE analysis confirmed that high egocentricity inner speech was indeed associated with converging activations in L-IFG, L-PMC and the L-MFG / L-SMA, as well as the right insula and right culmen. The convergence was consistently detected in both the unbalanced and the balanced datasets, suggesting that it was unlikely to be skewed by any particular paradigm.

ALE analysis of low egocentricity inner speech did not reveal significant convergence over speech perception areas or activations in the memory network. Instead, we observed significant convergence in a region encompassing the right insula (R-Insula) and right inferior frontal gyrus (R-IFG) across both datasets. The lack of convergence over speech perception and memory regions could reflect different levels of perceptual simulation along the dimension of intentionality at low egocentricity, and that perceptual simulation may only gain predominance when inner speech is both of low egocentricity and low intentionality. This type of inner speech is under-investigated and as such is not represented in our dataset, with all studies examining low egocentricity but highly intentional inner speech. The observed convergence over the right insula and inferior frontal gyrus are unlikely to reflect articulatory or phonological representations which are found, in a meta-analysis by Vigneau et al. (2011), to be located exclusively in the left hemisphere. Activations of these regions are, however, associated with auditory verbal hallucinations, the majority of which are heard voices in second or third persons and low in egocentricity (Sommer et al., 2008). Further research on non-hallucination participants suggests these right-hemisphere homologues may play a role in detecting unexpected self-voice changes (Johnson et al., 2021). While convergence across these regions could therefore indicate a greater demand on self-monitoring or an inherent inaccuracy of recreating acoustic representations of other voices as compared to one's own voice, there is currently insufficient research to evaluate or elaborate on this potential link.

Inner Speech As a Function of Intentionality

Within the dimension of intentionality we hypothesised that high intentionality inner speech would preferentially recruit the corollary discharge mechanism, with low intentionality inner speech favouring the perceptual simulation mechanism. Similarly, we predicted that the former would be associated with increased convergence in speech production areas such as L-IFG, L-PMC and SMA, whereas the latter would be associated with converging activations in speech perception areas (e.g., L-STG/STS) and in the memory network (e.g., L-MTG, L-MFG, L-SPL/PC).

Largely in line with our predictions, analysis of high intentionality inner speech yielded significant convergence of activation over speech production regions. Specifically, clusters of L-IFG, L-PMC, L-SMA, and parts of the right cerebellum (R-Culmen) were significant in the unbalanced dataset, but only L-PMC and L-SMA were consistently observed in the balanced dataset. The lack of convergence in L-IFG is of interest as it is invariably a part of the corollary discharge network according to computational and neuroanatomical models of speech production (Chen et al., 2011; Tourville & Guenther, 2011). For example, the DIVA model of speech production proposes that the left inferior frontal gyrus contains a speech sound map which serves as a repository of speech motor programs for each phonemic, syllabic or multi-syllabic sound a speaker might want to produce (Tourville & Guenther, 2011) - with the motor commands contained within each motor program then representing the efference copies which are passed into forward models. The lack of observed L-IFG convergence in the balanced dataset could suggest that its involvement is not ubiquitous across all high-intentionality inner speech paradigms. Given our previous proposal that the association between inner speech and the L-IFG was driven by the predominance of phonological-judgement paradigms in inner speech research, we conducted a post-hoc ALE analysing the high-intentionality studies which were classed as phonological-judgement tasks in the *Contrast Selection & Grouping* stage of data analysis. The results of the ALE analysis demonstrated convergence over the L-IFG across phonological judgement tasks (Appendix). While the subdivision of the high-intentionality study pool reduces statistical power, justifying a degree of caution, these preliminary finding raises two questions: (1) Does L-IFG convergence in phonological judgement tasks represent a subprocess specific to phonological judgement (e.g. speech segmentation; Burton, 2001), rather than inner speech *per se*? (2) If L-IFG activation during phonological-judgement does represent the generation of inner speech, does this indicate that other high-intentionality paradigms recruit different neurocognitive mechanisms to generate inner speech?

Analysis of low intentionality studies revealed significant convergence in L-MTG in both unbalanced and balanced datasets, with a substantial portion of the cluster encompassing the L-STG in both datasets. Significant convergence over speech perception brain regions (L-STG) aligns with our proposal that low intentionality inner speech preferentially relies on the perceptual simulation of speech within speech perceptual regions. The inclusion of studies utilising distinct paradigms: mind wandering sampling (Grandchamp et al., 2019; Raij & Riekki, 2017) and direct quotation reading (Alderson-Day et al., 2020; Yao et al., 2011), provides some evidence that the convergence over the L-MTG and L-STG is not attributable to a specific paradigm. This possibility was further examined via jackknife analysis, with results not reaching the predetermined robustness threshold. While typically an indicator of results being driven by an outlier study, the applicability of jackknife analysis to datasets with few studies is uncertain given the large proportion of data being removed with each iteration (e.g. 25% in a four study dataset). It is also noteworthy that these results align with the findings of an additional study not included in the GingerALE analysis, Hurlburt et al. (2016), in which task-elicited inner speech was associated with increased activation of the left inferior frontal gyrus and spontaneous inner speech was associated with increased activation of speech perception brain regions. Notably, Hurlburt et al. (2016) observed increased activation of Heschl's gyrus rather than the L-STG, but this discrepancy can likely be explained by their use of a region-of-interest approach which did not include the L-STG. The convergence of the cluster on portions of the L-MTG is also of interest. Given both the proximity and contiguity of the L-MTG cluster to the L-STG, and some evidence suggesting its involvement in the phonological processing of speech, it is plausible that this role relates to the phonological processing of the elicited inner speech (Ashtari et al., 2004). A role for the L-MTG in inner speech would align with previous findings suggesting that structural and connectivity abnormalities of the L-MTG are involved in the pathogenesis of auditory verbal hallucinations in schizophrenia (Cui et al., 2018; Zhang et al., 2017). However, the exact role the L-MTG plays in inner speech and auditory verbal hallucinations, and its relation to the proximate L-STG/STS, remains unclear.

Methodological Considerations

Activation-likelihood estimation provides a useful approach to address some of the weaknesses of individual neuroimaging studies. By calculating converging regions of neural activation across studies with distinct paradigms, ALE can help distinguish between paradigm-specific correlates which might not directly subserve the investigated behaviour, and paradigm-independent correlates which are more likely parts of the core neural circuit of interest. By pooling together numerous studies, ALE also allows for an increased power to detect true effects (Acar et al., 2018). However, there remain several considerations which should be made when interpreting the meta-analytical data. As explored in the introduction, a fundamental shortcoming within the inner speech neuroimaging literature is the predominance of task-elicited inner speech paradigms and relative lack of spontaneous inner speech experiments. This imbalance was reflected in our dataset, with a small pool of low intentionality experiments. A similar challenge exists within the egocentricity dimension. Despite inner speech experiences in day-to-day life often following a dialogic structure (Fernyhough, 1996, 2004), a comparatively small number of studies investigated low egocentric or dialogic inner speech as compared to high egocentric inner speech. This underlines an apparent tendency within the inner speech neuroimaging literature to adopt paradigms based on the ease of their implementation as opposed to their similarity to day-to-day inner speech. In both low egocentricity and low intentionality, this led to a smaller pool of studies than ideal and prevented more comprehensive analysis into the effects of specific paradigms and the contrasting of dimensions.

A further limitation of this activation-likelihood estimation study, and the ALE technique more generally, is that they analyse fMRI or PET data which are inherently correlational. Whilst this can be used to identify relationships between neural activation and behaviour, the degree to which a behaviour is caused by that neural activation cannot be easily determined using these observational techniques. The results of these analyses could therefore serve as an empirical and theoretical basis on which future, causal research may be based. One avenue for further causal research could involve the use of brain stimulation techniques to disrupt processing within the speech production and speech perception regions, individually, as performance in various inner speech tasks is recorded. Using the model and predictions laid out in this paper, specific hypotheses can be made as to which tasks would be impaired by suppression of speech production regions as compared to speech perception regions.

In spite of these shortcomings, the ALE findings are the result of best efforts given the current state of the literature, and serve to highlight the importance of interpreting inner speech as a phenomenon which can vary in its phenomenology, sensorimotor properties and neural correlates. While providing evidence in support of a model which explains a diverse range of findings within the neuroimaging literature, the meta-analysis also underscores the need for future research to incorporate a more diverse range of analytical techniques and elicitation paradigms in order to fully elucidate the mechanisms by which it can be generated.

The Utility of the Current Framework

The conceptual aim of the current framework was to explain the mechanisms by which inner speech can be generated and the variables that influence these mechanisms. The results of the ALE analyses broadly support these motivations. First, we provide evidence that a framework classifying inner speech across egocentricity and intentionality dimensions can allow for the identification of different neural circuits during inner speech generation. In turn, these various neural circuits indicate that inner speech is generated via multiple, distinct mechanisms. Second, by centering the framework around two fundamental dimensions (egocentricity and intentionality) inherent to all varieties of inner speech, the framework also allows for existing studies to be placed within the two dimensions post-hoc. This helps identify which subtypes of inner speech are well documented within the research literature, and which subtypes of inner speech remain under-investigated (e.g. low egocentricity and low intentionality subtypes). Finally, although not directly tested in the ALE analyses, the framework allows for diverse phenomenologies to be easily mapped into the two dimensions. This can then be used to generate predictions on the neural correlates and cognitive mechanisms associated with the generated inner speech.

The lack of reliable L-IFG convergence challenges the predominant view that inner speech is invariably generated by the motor speech production system (Alderson-Day & Fernyhough, 2015; Jones & Fernyhough, 2007). L-IFG involvement in inner speech could instead be restricted to specific paradigms, with a preliminary analysis indicating that phonological judgement tasks are strongly associated with the L-IFG. This is notable as phonological judgements are commonly used to reliably induce inner speech in research settings. Future studies should weigh the convenience of phonological judgements as an inner speech induction technique against the possibility of them demonstrating distinct neural and cognitive mechanisms when compared to other inner speech subtypes.

It is also notable that the current framework failed to predict the lack of convergence over the L-STG/STS in low egocentricity studies, with convergence instead being observed over the R-IFG. Given that the R-IFG is not commonly implicated as a region causally involved in the generation of inner speech, further research elucidating the neural and cognitive mechanisms driving low egocentricity inner speech is required. As the pool of low egocentricity studies consisted entirely of studies which were also high in intentionality, it remains unclear as to whether the observed neural correlates are specific to studies which are both low in egocentricity and high in intentionality, or whether they are a feature of low egocentricity inner speech more broadly. The investigation of inner speech which is low in both egocentricity and intentionality represents a compelling area for future research given the current paucity of research and its regular occurrence within day-to-day inner speech experiences (McCarthy-Jones & Fernyhough, 2011).

The broader development and testing of the framework also exposed a relative lack of research investigating the exact mechanisms and neural correlates driving perceptual simulation in inner speech. This is of interest as the concept of perceptual simulation has received wide attention in explaining other types of sensory imagery, such as visual imagery (Ranganath & D’Esposito, 2005; Reddy et al., 2010). The precise involvement of different neural networks (perception, memory, lexical) in the perceptual simulation of speech therefore remains a topic requiring further consideration and empirical investigation.

The results of the ALE analyses yield distinct patterns of neural activation than observed in Grandchamp et al. (2019). Grandchamp et al. (2019) observed consistent L-IFG activation throughout their investigation of inner speech across dialogality and intentionality dimensions - therefore lending support to a purely corollary discharge approach. However, our ALE analyses found convergence over the L-IFG to be particularly unreliable, as determined by both observed convergence across conditions and fail-safe N /

jackknife analyses. Whilst the reasons for this divergence are difficult to determine without carrying out further analysis, it is notable that Grandchamp et al. (2019) predominantly utilised highly intentional inner speech tasks, some of which also involved semantic processing. For example, both the *monologal self-voice inner speech condition* and the *monologal other-voice inner speech condition* required participants to generate definitions for a visually presented object. Within these conditions, it is plausible that activation of the L-IFG could reflect semantic processing during object name retrieval (Krieger-Redwood & Jefferies, 2014) rather than inner speech, *per se*. However, the involvement of the L-IFG in the *verbal mind wandering condition* remains less clear given the lack of a significant semantic component to the task.

It is also notable that Grandchamp et al. (2019) reported minimal activation of L-STG and L-MTG during their low intentionality task, which is at odds with Hurlburt et al. (2016) and our ALE analysis of low intentionality studies. Grandchamp et al. (2019) propose that the absence of L-STG / L-MTG activation in their study could be explained by their inclusion of verbal mind wandering experiences which were more condensed than that used in Hurlburt et al., (2016). Although plausible, it is unclear from a neurocognitive perspective why condensed inner speech would not result in any activation of speech perceptual regions when compared to an implicit baseline, nor is it clear the extent to which the analysed experiences were actually condensed. We judge Grandchamp et al. (2019) alternative explanation to be more likely, that the lack of L-STG / L-MTG activation in low intentionality inner speech was caused by insufficient statistical power to detect the effect. It is also plausible that the task methodology, which required participants to report the timing of the mind wandering experiences *after* the 30-second trial, produced timing data which is not accurate enough to isolate verbal mind wandering experiences from other cognitions during fMRI modelling and analysis. Nevertheless, Grandchamp et al. (2019) dimension of condensation does remain an area which is worthy of further elucidation and could explain some of the divergent findings within our analyses. Given that it was excluded from our framework, in part, due to ambiguity in implementation, testing, and evidence, it is a concept worth revisiting when a larger corpus of research is available.

Given the finding that L-IFG activation is not an invariable feature across all forms of inner speech, and that there are more general differences in neural correlates across egocentricity and intentionality of inner speech, we argue our current framework is of significant utility when compared to models which posit that a motor-route of generation subsumes all inner speech subtypes. Our pragmatic approach views inner speech as a dynamic phenomenon which varies in its phenomenological attributes and mechanisms of generation. Whilst we provide one framework which seeks to explain the exact relationship between phenomenological attributes and neurocognitive mechanisms, it is clear that further research on less studied inner speech subtypes (e.g. spontaneous inner speech in other voices) is vital to refining the model and developing a complete understanding of how inner speech is implemented in the brain. There is also a need for research to investigate the exact *causal* mechanisms by which corollary discharge and perceptual simulation operate, beyond much of the available research, including our own, investigating these mechanisms at a correlational level. The benefits of a more complete understanding of inner speech are not limited to basic research, but could have a tangible impact on translational studies. For example, accurate and reliable functional mapping of the brain regions involved in inner speech generation could maximise the efficacy of brain stimulation interventions of auditory verbal hallucinations, a therapeutic approach which has yielded mixed results to date (Moseley et al., 2015).

Conclusion

In line with studies highlighting the diverse nature of inner speech (Alderson-Day et al., 2018; Hurlburt et al., 2016), the results of the ALE meta-analysis further demonstrated that distinct neural mechanisms were differentially engaged for inner speech that varies along its egocentricity and intentionality. In particular, speech production areas implicated in the motor-route of generation are consistently engaged in inner speech with high egocentricity and high intentionality, but not with inner speech with low egocentricity and intentionality. The current study makes three important contributions: First, it provides evidence that varieties of inner speech are supported by more than one neural mechanism. Second, it provides a flexible and useful cognitive framework that bridges between the diverse phenomenology of inner speech and the two underlying neural mechanisms. Third, we demonstrated that our current understanding of inner speech is highly skewed by paradigms that require explicit phonological judgements. It is crucial that we test different types of inner speech across a range of paradigms to triangulate the neurocognitive mechanisms that causally produce various forms of inner speech, as well as auxiliary mechanisms that underpins inner speech (e.g., working memory, attention, verbal monitoring, Theory of Mind, etc.). In conclusion, the present study provides a novel contribution to the research literature by showing that different neural mechanisms are engaged for inner speech that varies in its egocentricity and intentionality. It also provides a flexible cognitive framework that bridges the phenomenology of inner speech and its underlying neural mechanisms. The study highlights the importance of testing different types of inner speech across a range of paradigms to better understand the neurocognitive mechanisms that causally produce and support inner speech.

Chapter 3 - Differential Impacts of Aphantasia Across Inner Speech Subtypes

Jaydan Pratts[1*], Gorana Pobric[1], Bo Yao[12*]

1 Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, UK
2 Department of Psychology, Fylde College, Lancaster University, UK

*Corresponding Authors:
Jaydan Pratts
Division of Psychology, Communication and Human Neuroscience
Zochonis Building
Brunswick Street
Manchester M13 9PT jaydan.pratts@manchester.ac.uk
United Kingdom

Dr Bo Yao
Department of Psychology
Lancaster University
Lancaster LA1 4YF b.yao1@lancaster.ac.uk
United Kingdom

Abstract

Despite the ubiquity of inner speech, the mechanisms by which it is generated remain debatable. Two models which aim to explain inner speech generation are the corollary discharge model and the perceptual simulation model. In Pratts et al. (2023), we propose a dual-mechanistic framework which states that both of these mechanisms are differentially engaged in inner speech generation, depending on the phenomenology of the inner speech. Here, we investigate this framework by examining a population with aphantasia, a condition that negatively affects the ability to generate visual imagery. As both visual imagery and the perceptual simulation model rely on the reactivation of perceptual experiences, aphantasia could represent a natural model of impaired perceptual simulation. This would allow us to examine whether perceptual simulation underpins specific phenomenological subtypes of inner speech. Across two experiments, we demonstrate that aphantasia also negatively impacts measures of inner speech, with the impact of aphantasia varying across phenomenological dimensions of inner speech.

Keywords: inner speech, corollary discharge, perceptual simulation, aphantasia

Highlights:

- Aphantasia is not limited to deficits in visual imagery abilities.
- Varieties of inner speech are differentially affected by aphantasia.
- Some evidence in favour of a dual-mechanistic framework is observed.

Introduction

Models of Inner Speech Generation

Inner speech is an internal, speech-like experience without the presence of an accompanying external sound (Alderson-Day & Fernyhough, 2015). Inner speech plays a large role in cognition, being linked to working memory (Baddeley, 1992; D'Esposito, 2007), silent reading (Filik & Barber, 2011; Yao & Scheepers, 2011; Yao et al., 2011) and goal tracking (Emerson & Miyake, 2003; Miyake et al., 2004). The impairment of typical inner speech is often implicated in mental disorders such as auditory verbal hallucinations in schizophrenia (Frith, 1987), highlighting the importance of thoroughly understanding its neural and mechanistic underpinnings.

The cognitive mechanisms by which inner speech is generated remains the topic of some debate. One approach to understanding the cognitive underpinnings of inner speech, and likely the predominant approach, is based on the corollary discharge model. This model proposes that inner speech is the predicted perceptual consequence of intended articulation (Jack et al., 2019; Jacobson, 1932; Scott, 2013; Scott et al., 2013; Watson, 1913). The intent to speak generates an efference copy of the articulatory signal. This efference copy originates in speech production regions before entering a forward model to predict what the intended articulation would sound like. The resulting prediction is then received and perceived in speech perception regions. Given the reliance of this model on speech production regions to generate the initial motor commands, the model is said to use a motor-route in order to generate inner speech.

Although the corollary discharge model aligns with the finding that tasks which require the generation of inner speech (e.g. rhyme judgement) elicit activity in speech production brain regions (Lurito et al. 2000; Shergill et al. 2001), it has several limitations. For example, inner speech often incorporates the voices of other individuals, which inherently contain vocal features distinct from one's own (McCarthy-Jones & Fernyhough, 2011). These 'foreign' vocal features cannot be provided solely by corollary discharge as it is physically constrained by one's own articulator. There is also increasing heterogeneity in the data, with recent neuroimaging studies associating inner speech with activation of speech perception regions, rather than speech production regions (Hurlburt et al., 2016; Yao & Scheepers, 2011; Yao et al., 2011, 2021).

One way for inner speech to be generated in perceptual regions is via perceptual simulation, also termed the perceptual-route (Pratts et al., 2023). The perceptual simulation model posits that as an individual is exposed to a particular sensory stimulus, the neurons in the relevant sensory cortices encode the distinct firing patterns which are evoked. These patterns are captured and stored, allowing them to be reactivated at a later time. By reactivating these firing patterns, the stored perceptual experience (or part thereof) can be simulated and internally perceived (Barsalou, 2008). As a variety of captured firing patterns accumulate, they can be integrated or remixed to create new patterns and consequently new perceptual experiences (Barsalou, 2008). Interpreting inner speech through the lens of perceptual simulation would address some of the limitations of the corollary discharge model described earlier. For example, vocal features that cannot be accurately reproduced using one's own articulator, such as vocal features of Darth Vader, or some of the opposite sex whose pitch is outside one's own range, can be perceptually simulated. Perceptual simulations of speech do not depend on speech production, which could explain the lack of activations in speech production areas during spontaneous inner speech in silent reading and at rest (Yao et al., 2011; Hurlburt et al., 2016).

Nevertheless, the perceptual simulation model is also at odds with some empirical findings. For example, several neuroimaging studies do not find inner speech to be correlated with auditory cortex activation (De Nil et al., 2000; Gulyás, 2001), which implies a lack of perceptual simulation. Moreover, suppression of speech production regions using repetitive transcranial magnetic stimulation (rTMS) has been found to inhibit inner speech, suggesting that at least some types of inner speech depend predominantly on speech production regions (Aziz-Zadeh et al., 2005). These disparate results suggest that neither corollary discharge nor perceptual simulation offer a full mechanistic account of inner speech when considered in isolation.

A Dual-Mechanistic Framework of Inner Speech

In Pratts et al. (2023), we propose that in order to reconcile conflicting findings in the inner speech literature, one needs to recognise that inner speech is not a homogeneous, uniform phenomenon, but a multi-dimensional, flexible process manifested in a variety of forms (Hurlburt et al., 2013; McCarthy-Jones & Fernyhough, 2011). We therefore proposed a framework containing two phenomenological dimensions by which inner speech can vary: egocentricity and intentionality. Egocentricity represents the extent to which inner speech is a recreation of one's own voice (high egocentricity), or a simulation of another individual's voice (low egocentricity). The dimension of intentionality captures the extent to which inner speech is elicited by explicit task demands (high intentionality) or occurs spontaneously (low intentionality).

Instead of subscribing to either the corollary discharge model or perceptual simulation model exclusively, the framework predicts that the two mechanisms play a role in generating in speech, but the relative contributions of the two mechanisms vary depending on the position of the inner speech along the egocentricity × intentionality space. The more egocentric and intentional inner speech is, the more strongly it relies on corollary discharge. In contrast, the less egocentric or intentional inner speech is, the more likely it resorts to perceptual simulation. For example, imagery of another person's voice (low egocentricity) and inner speech that emerges spontaneously (low intentionality) would preferentially recruit the perceptual simulation mechanism. Initial evidence for the dual-mechanistic framework was gathered via a series of fMRI meta-analyses which observed differential activation of brain regions in inner speech studies varying in intentionality and egocentricity. However, two methodological limitations justify further validation of the framework. First, the meta-analyses were constrained by the limited number of fMRI studies investigating less intentional and egocentric varieties of inner speech, with the majority of available studies eliciting highly intentional and egocentric inner speech. Second, the fMRI meta-analyses provided correlational evidence, with the causal relationship between the brain regions and inner speech generation remaining unclear. Further research which explicitly tests the interaction between different varieties of inner speech and their neurocognitive mechanisms is therefore required.

Aphantasia: A Natural Model For Perceptual Simulation Dysfunction?

One approach to investigating the dual-mechanistic framework can be found within the broader mental imagery literature. Like inner speech, mental imagery is the ability to generate internal representations of sensory percepts in the absence of an external accompanying stimulus (Kosslyn et al., 2001). For example, the ability to close one's eyes and internally visualise an apple. This ability is fairly ubiquitous among the general population and is not limited to a single sensory modalities (e.g. visual imagery, tactile imagery, auditory imagery). Despite heterogeneity in phenomenology and function, research investigating mental imagery of different sensory modalities has generally converged on a single general mechanism explaining the process by which the imagery is generated. While terminology varies across studies and modalities, this mechanism is broadly synonymous with perceptual simulation as it involves the simulation of a percept via reactivation of neurons within the sensory area (Dijkstra et al., 2020; Kuhl et al., 2010; McNorgan, 2012; Schmidt et al., 2014; Wheeler et al., 2006). This model is supported by a range of evidence, including neuroimaging studies which illustrate very similar neural representations when imagining and during perception (Dijkstra et al., 2019; Lee et al., 2012; Zatorre et al., 1996).

Given that the generation of mental imagery is argued to rely on a similar mechanism to perceptual simulation, examining the ability of individuals with deficits in mental imagery to generate phenomenological varieties of inner speech could help deduce which subtypes of inner speech, if any, are generated using perceptual simulation. Such a population can be found in those with aphantasia, the congenital inability to generate visual imagery (Zeman et al., 2015). Given that aphantasia stems from aberances in the generation of imagery rather than deficits in metacognition (i.e. blind to a successfully generated image; Keogh & Pearson, 2018), the aetiology of aphantasia could lay in deficits in reactivating neurons within the sensory areas. If correct, aphantasia would be expected to affect all types of mental imagery given their shared mechanism of generation. More importantly, aphantasia could be used as a natural model of perceptual simulation dysfunction, allowing us to investigate the hypotheses outlined in the dual-mechanistic framework.

Aim & Hypotheses

The dual-mechanistic framework argues that two distinct mechanisms of inner speech generation exist, one based on corollary discharge (motor-route) and the other based on perceptual simulation (perceptual-route). Aphantasia would be predicted to be correlated with deficits in the types of inner speech which rely on perceptual simulation. As per the framework described earlier, this would be types of inner speech which are less egocentric and less intentional. Inner speech types which primarily rely on the corollary discharge should be comparatively spared, this would be inner speech which is more egocentric and intentional.

In this two part study, we first used a brief imagery questionnaire to examine our initial hypothesis that aphantasia is not limited to deficits in the visual domain. We hypothesise that aphantasia instead represents a more general deficit in imagery generation abilities which could indicate its usefulness as a natural model of perceptual simulation dysfunction (pre-experiment). Then, we used a more comprehensive battery of inner speech tasks to determine the impact of aphantasia on four phenomenological subtypes of inner speech (high & low egocentricity, high & low intentionality) relative to a control group (main experiment). This battery of tasks consists of the following:

- Vividness of Visual Imagery Questionnaire.
- Scenario rating task (high intentionality inner speech with high & low egocentricity conditions).
- Mind wandering task (low intentionality inner speech with high & low egocentricity conditions).

According to our hypotheses, Experiment 1 should demonstrate that aphantasia is significantly correlated with deficits in other types of sensory imagery. In Experiment 2, we hypothesise that performance in inner speech tasks which primarily rely on corollary discharge, as determined by Pratts et al.'s (2023) framework, are preserved relative to tasks which rely on perceptual simulation. Performance on the scenario rating task and mind wandering task should produce varying results depending on the specific condition involved. This will be further discussed in the methodology section.

Pre-experiment: Is aphantasia correlated with non-visual imagery deficits?

In the pre-experiment, we set out to determine whether individuals with aphantasia retain their ability to generate mental imagery in other sensory modalities. We hypothesised that individuals with aphantasia would exhibit decrements in their ability to generate imagery across different sensory modalities (i.e. not only visual imagery) as they rely on a mutual generative mechanism. Given our proposal that some forms of inner speech rely on similar mechanisms of generation to visual imagery, we hypothesised some decrease in inner speech performance. This hypothesis is further addressed in the main experiment.

Methods

Participants

Our participants consisted of 75 individuals recruited from the *r/aphantasia* sub-forum on *Reddit,* an online discussion website with sub-forums targeting specific interests or communities. Ethical approval for the study was granted by the local ethics board (Ethics Committee Approval Code: 2019-5370-10112). Participants were informed that participation was entirely voluntary and that no identifiable data would be collected. Participants agreeing to take part in the study were then able to access a questionnaire via an attached URL link. The questionnaire was hosted on Google Forms and did not collect names, e-mail addresses or any other type of self-identifying information. Given the exploratory nature of the first experiment, information regarding gender and age was not collected.

The questionnaire was also posted on the *r/samplesize* sub-forum on Reddit which served as the control sample. The control group consisted of 164 respondents.The *r/samplesize* sub-forum is dedicated to discussion and participation in online experiments and questionnaires. The choice of a different sub-forum as a control sample, as opposed to in-person testing using a locally available sample or online crowdsourcing provider was made as it was likely to present a more demographically comparable group and sufficient for the exploratory purposes of this study.

Materials & Procedure

The experiment consisted of a brief questionnaire drawn from the Plymouth Sensory Imagery Questionnaire (Psi-Q; Andrade et al. 2014). The Psi-Q consists of seven sections, each with five questions. Each section is dedicated to a specific sensory or interoceptive modality (sight, sound, smell, taste, tactile sensation, bodily sensation, emotions). Each item is in the form of a question asking the participant to “Imagine the sight/sound/smell/etc of…”, to which participants respond by selecting a vividness score ranging from 0 to 10. Given our specific interest in the auditory modality, and the exploratory nature of this study, the questionnaire was altered in order to ensure specificity and brevity.

The adapted questionnaire consisted of seven sections, six of which consisted of just the first two items. The section addressing auditory imagery contained four items. As the original Psi-Q primarily focuses on imagery of non-vocal sounds, the four items were adjusted such that they each addressed a different variety of auditory imagery: inner speech in one’s own voice, imagery of another person’s voice, imagery of a musical melody, and imagery of non-vocal noises. Participants responded to each item by using a 1-5 scale in which 1 corresponded with “No image at all…” and 5 corresponded with “Perfectly clear…”.

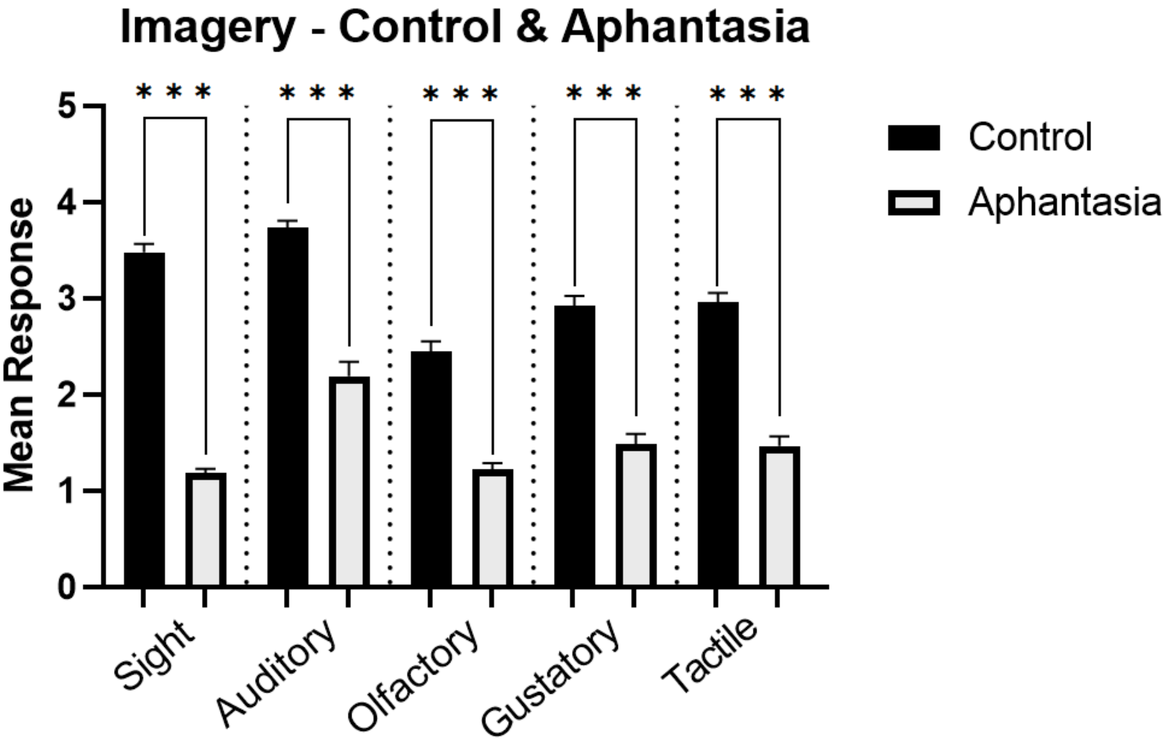
Data Analysis

Analysis of questionnaire data was completed using R (version 4.1.2) and the RStudio IDE (Build 443). *Tidyverse* (Version 1.2.1; Wickham et al., 2019) and *Rstatix* (Version 0.3.0; Kassambara, 2019) packages were used for data cleaning and analysis. Analysis compared the vividness scores of the control group to the aphantasia group on a per sensory modality basis. Participants' responses in the various items within each sensory domain were averaged in order to give a single vividness value from 1-5 for each of the 5 modalities. Given the unequal sample sizes and non-gaussian distributions, Mann-Whitney U tests were used.

2.2. Results

Consistent with their self-identification, primary data analysis found that participants with aphantasia scored significantly lower [median = 1.0, SD = 0.70; *U* = 855, p <.001] in the visual imagery subsection than the control group [median = 3.5, SD = 1.12]. A significant reduction in vividness score was also seen in all other sensory modalities: olfactory imagery [*U* = 2741, p <.001], gustatory imagery [*U* = 2522, p <.001], tactile imagery [*U* = 2303, p <.001] and auditory imagery [*U* = 2364, <.001].

Figure 1. Mean response per sensory category in control and aphantasia groups.



*Note: Error bars represent standard error of the mean. *** = p <.001.*

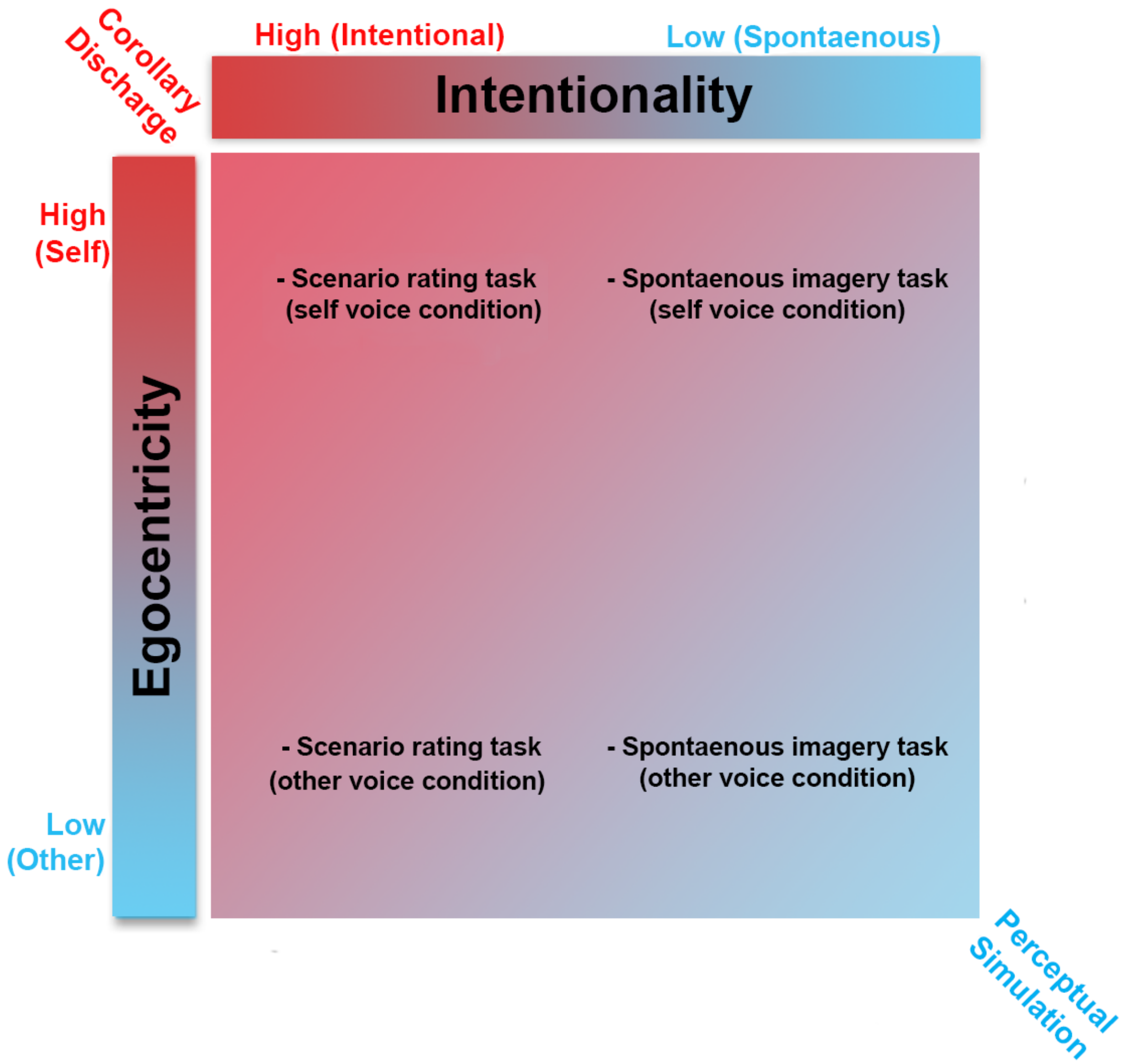
Discussion

The pre-experiment examined whether aphantasia was limited to the inability to generate visual imagery, or whether aphantasia was characterised by a reduced ability to generate mental imagery across sensory modalities. The results of the adapted questionnaire demonstrate that aphantasia is linked to a decreased ability to generate imagery in a variety of sensory modalities. Visual, auditory, olfactory, gustatory and tactile imagery were all significantly less vivid in the aphantasia group when compared to the control group. Given the high degree to which a self-identified lack of visual imagery is concomitant with a decreased ability to generate imagery in other sensory modalities, a common aetiology seems likely. As proposed earlier, this aetiology could relate to issues in the production mechanisms underlying imagery. In this case, aphantasia represents a good candidate for a natural model of perceptual simulation dysfunction. Given the preliminary nature of the experiment and the brevity of the modified Psi-Q used, we did not conduct additional analysis examining whether aphantasia selectively impacted specific subtypes of inner speech. Instead, we observed sufficient evidence that on aggregate, aphantasia impacts auditory imagery and inner speech, and could therefore represent a useful approach to investigating the hypotheses laid out by the dual-mechanistic framework.

Main Experiment: Performance of participants with aphantasia in a battery of inner speech tasks.

After confirming the initial hypothesis that aphantasia is associated with decreased imagery abilities in a range of sensory domains, a further sample of participants was recruited and asked to complete a more comprehensive battery of inner speech tasks. This battery was designed to test each phenomenological variant of inner speech identified in Pratts et al. (2023). We hypothesised that aphantasia would predominantly affect conditions in which the elicited inner speech subtype was reliant on perceptual simulation; those conditions are those which are lower in egocentricity and intentionality, as seen in the figure below.

Figure 2. Placement of tasks and conditions within the dual-mechanistic framework outlined in Pratts et al. (2023).



Note: Aphantasia is hypothesised to have a larger impact in conditions positioned in the blue shaded areas as these place increasing demands on the perceptual simulation mechanism of generation.

Methods

Participants

Based on power analyses conducted using G*Power (Version 3.1.9; http://www.gpower.hhu.de/), we estimated that a total of 58 participants would be sufficient to detect a large effect size (f = 0.4) during ANOVA analysis. The aphantasia group consisted of 49 individuals, recruited from the *r/aphantasia* subreddit as outlined in the

Pre-experiment methodology. Ethical approval for the study was granted by the local ethics board (Ethics Committee Approval Code: 2019-5370-10112), with subreddit moderators allowing the experiment to be posted on the *r/aphantasia* subreddit. Participants agreeing to take part in the study were then able to access the experiment via an attached URL link. The experiment was hosted on Gorilla (www.gorilla.sc), an experiment authoring and deployment platform. No names, e-mail addresses or any other type of self-identifying information were collected throughout the experiment.

A total of 33 participants with aphantasia were included in the final dataset. Participants had a mean age of 33 years (SD = 9.78), with 25 male participants and 8 female participants. Aphantasia was defined as those scoring ≤ 25 out of a possible 80 in the VVIQ questionnaire (*M* = 15, *SD* = 2.7), the cutoff criteria was based on previous aphantasia research (Pounder et al. 2022). Seven participants were excluded as their VVIQ scores (>25) exceeded our cutoff criteria. An additional nine participants were excluded as English was not their primary language, which would potentially impact their performance in subsequent tasks.

A further 33 participants were recruited for the control group using an online research participant provider (Prolific Research). The control group consisted of first language English speakers with a mean age of 29 years (SD = 5.62), with 24 male participants and 9 female participants. All participants in the control group scored above 25 points on the VVIQ scale, the criteria used to determine aphantasia (*M* = 59, *SD* = 11.6). Participants were compensated for their time.

Materials & Procedure

Participants first completed 2 screening measures: a demographics questionnaire which recorded age, gender, level of education and languages spoken, and the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). The VVIQ consists of 16 items divided into 4 different blocks. Each block describes a general scenario which participants should visualise, with each item then asking participants to rate the vividness of a specific visual aspect of that scenario. For example, in one block of items a participant might be told to bring to mind the image of a friend. Then, participants can be asked to specifically rate the vividness of: (1) the contours of the face, (2) poses of head, (3) carriage when walking and (4) clothes worn. Participants rate each item using a 5-item scale as presented in figure 3. Participants rated each item using a 1-5 scale in which 1 corresponded with “No image at all…” and 5 corresponded with “Perfectly clear…”.

Figure 3. 5-point scale used in the VVIQ.

5	Perfectly clear and as vivid as normal vision
4	Clear and reasonably vivid
3	Moderately clear and vivid
2	Vague and dim
1	No image at all, you only “know” that you are thinking of an object

Note: numerical scale may be reversed in some research articles. In this case, a higher number corresponds to more vivid inner speech.

Participants then completed 2 tasks in a counterbalanced order: a scenario rating task which examined high intentionality inner speech subtypes, and a mind wandering task which examined low intentionality inner speech subtypes.

Scenario Rating Task

The scenario rating task contained three conditions:

Own voice: In this condition participants were asked “*Using your own inner voice, say…*” followed by a sentence in prose. Participants would respond using by rating the vividness of their elicited inner speech on a 5-point scale.

Other voice: In this condition, participants were asked ““*In another persons voice, say…*” followed by a sentence in prose. Participants would then generate the inner speech in the voice of someone familiar to them and rate the vividness of the elicited auditory imagery using a 5-point scale. Participants were instructed to choose a single individual whose voice they were familiar with at the beginning of the experiment.

Sound: In this condition, participants were asked ““*Imagine the sound of…*” followed by a description of a commonly heard non-vocal noise. For example, a siren. Participants would then attempt to generate auditory imagery of the given scenario and rate its vividness using a 5-point scale.

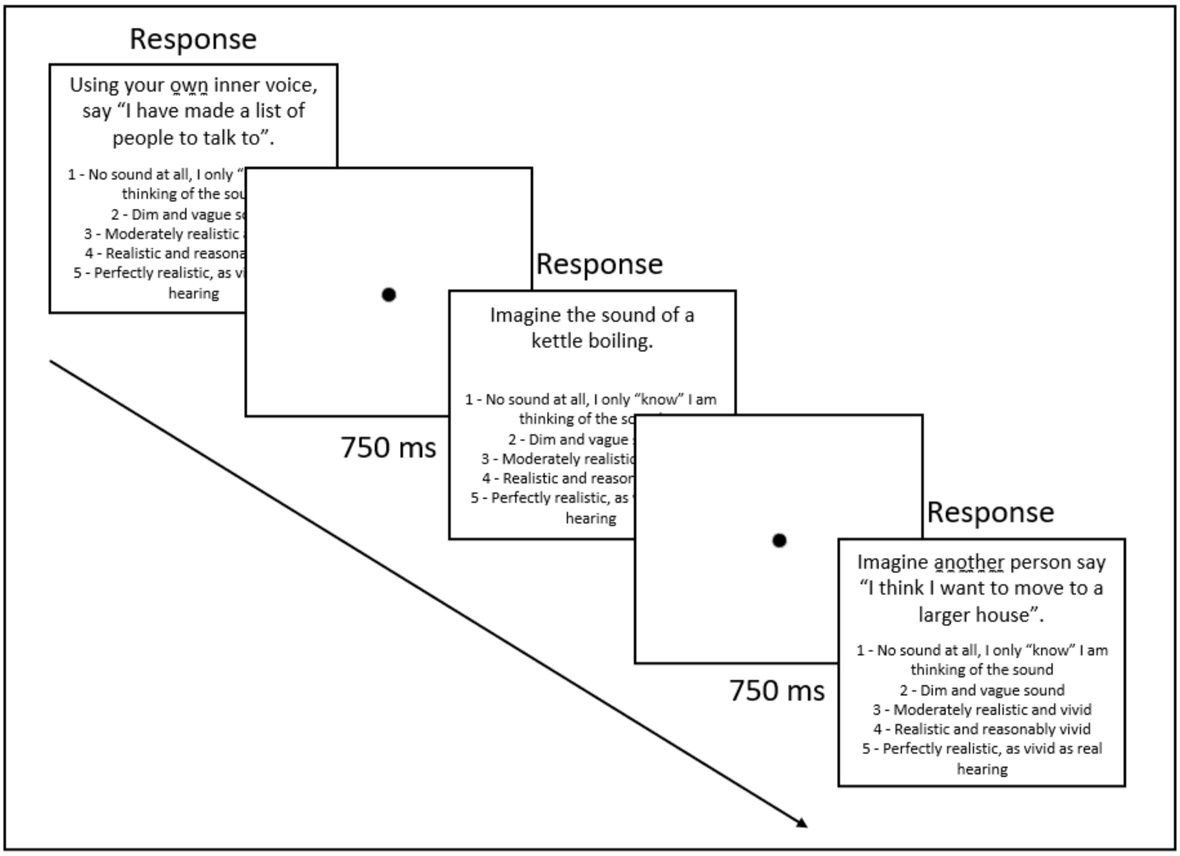
Own voice scenarios represent a subtype of inner speech which places heaviest demands on the motor-route of generation (i.e. corollary discharge), being both highly intentional and egocentric. *Other voice* scenarios are intended to generate inner speech which places demands on both the motor-route of generation and perceptual-route of inner speech generation (i.e. perceptual simulation), given that it is high in intentionality, but low in egocentricity. The *sound* scenarios represent a type of auditory imagery which is maximally low on egocentricity, in that it cannot be accurately reproduced using one's own articulators and must rely exclusively on the perceptual-route.

All conditions used a scale which was adapted from that used in the VVIQ questionnaire and can be seen below in Figure 4, with responses entered on a keyboard using numbers 1 - 5. Participants were presented with 8 scenarios per condition, for a total of 24 scenarios. Conditions and scenarios were interspersed in a random order with a 750ms ISI separating each trial. Scenarios were not time limited in order to prevent ceiling effects. Stimuli sets were rotated across participants such that participants were not exposed to the same set more than once.

Figure 4. 5-point scale used in the scenario rating task.

5	Perfectly realistic, as vivid as real hearing
4	Realistic and reasonably vivid
3	Moderately realistic and vivid
2	Dim and vague sound
1	No sound at all, I only “know” I am thinking of the sound

Figure 5. Procedure for the scenario rating task.



Note: ISI is presented for 750ms, participants then presented with an own voice, other voice or sound scenario. Participants respond using keyboard numbers 1 to 5.

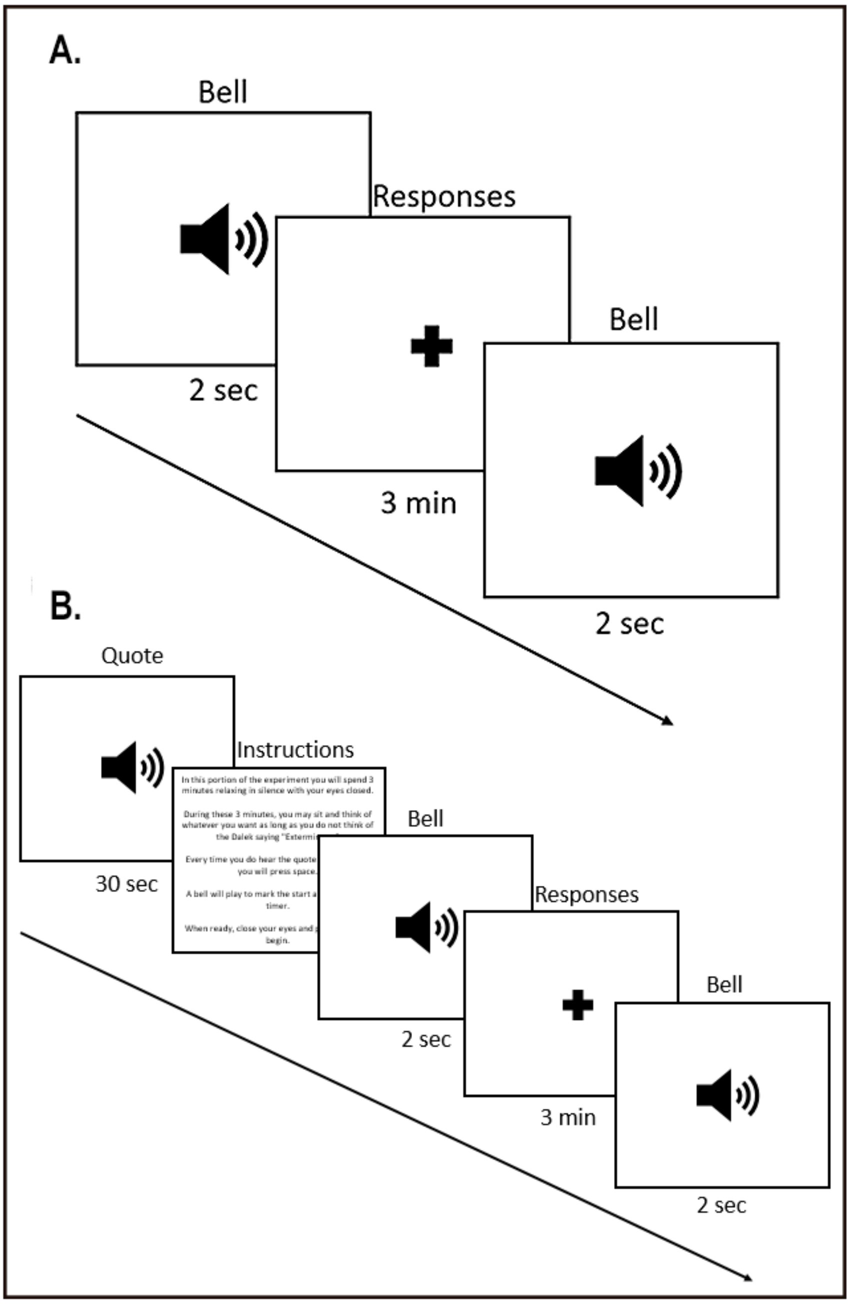
Spontaneous Imagery Task

The spontaneous imagery task examined low intentionality subtypes of inner speech and consisted of two conditions which were completed in a counterbalanced order. In the *own voice* condition, participants were presented with a fixation cross for three minutes. During this time participants were instructed to close their eyes, relax and press the spacebar button if they heard their inner voice. They were permitted to press the button multiple times. An audible bell was used to mark the beginning and end of the three minute period. The instructions made clear that the spacebar should be pressed per occurrence rather than per word. As an example, participants are given the hypothetical inner speech experience "I wonder what the end bell will sound like" and are told that such a thought should result in a single keypress.

In the *other voice* condition participants first heard a 30 second audio clip of a movie quotation being repeated. This quotation was from the television show Doctor Who and consisted of a Dalek repeating the phrase "Exterminate!". Participants were instructed not to think of a Dalek saying "Exterminate!". Participants then underwent 3 minutes of silence marked by an audible bell at the beginning and end of the period, as in the *own voice* condition. During this period participants were tasked with pressing the response key every time they experienced auditory imagery of a Dalek saying “Exterminate!”.

Both conditions were preceded by short examples of the task. In these examples, participants spent 30 seconds instead of 3 minutes in silence. In order to preserve the novelty of the stimuli, the *other voice* practice condition used a quotation from Star Wars, in which Darth Vader says “No, I am your father.”

Figure 6. Own and Other conditions of the spontaneous imagery task.



In this portion of the experiment you will spend 3 minutes relaxing in silence with your eyes closed.

During these 3 minutes, you may sit and think of whatever you want as long as you do not think of the Dalek saying "Extermi

Every time you do hear the quote you will press space.

A bell will play to mark the start a timer.

When ready, close your eyes and begin.

Note: A) Own voice condition. When ready, participants close their eyes and an audible bell marks the beginning of the 3 minute response period. During this period they may press the response key if they hear their inner voice. A final bell marks the end of this period. B) Other voice condition. The movie quotation is presented auditorily for 30 seconds, participants are then presented with instructions. When ready, participants close their eyes and an audible bell marks the beginning of the 3 minute response period. During this period they may press the response key if they "hear" the movie quote internally. A final bell marks the end of this period.

Data Analysis

Analysis of all data was completed using R (version 4.1.2) and the RStudio IDE (Build 443). *Tidyverse* (Version 1.2.1; Wickham et al., 2019) and *Rstatix* (Version 0.3.0; Kassambara, 2019) packages were used for data cleaning and analysis. In the scenario rating task, the data was trimmed by grouping the data by condition (Own, Other, Sound) and participant group (Aphantasia group, Control group), and then excluding any trials whose response time was more than 2.5 times the median absolute deviation (MAD), this excluded 5.9% of trials. This procedure has been used in previous imagery research (Logie et al. 2011) and is outlined in Leys et al. (2013). It exhibits several benefits over competing approaches, including a standardised procedure and parameters in order to avoid subjectivity when selecting degrees of freedom. As the spontaneous imagery task consisted of a single three minute trial per condition, per participant, we adopted a different approach which avoided excessive data loss. This consisted of an outlier threshold of ≥ 180 responses per trial. This is equivalent to participants reporting one spontaneous thought per second for the entire three minute trial. Two trials were excluded based on this criteria.

A comparison of VVIQ scores in the aphantasia group and control group was completed using a t-test on participants' total scores. Statistical analysis of the scenario rating task consisted of an ANOVA with Group (Aphantasia, Control) and Scenario Type (Own, Other, Sound) as fixed factors. Post-hoc tests using the Bonferroni correction would be used to further examine any significant effects. Statistical analysis of the spontaneous imagery task consisted of an ANOVA with Group (Aphantasia, Control) and Inner Speech Type (Own, Other) as main effects, with post-hoc tests using Bonferroni corrections.

Results

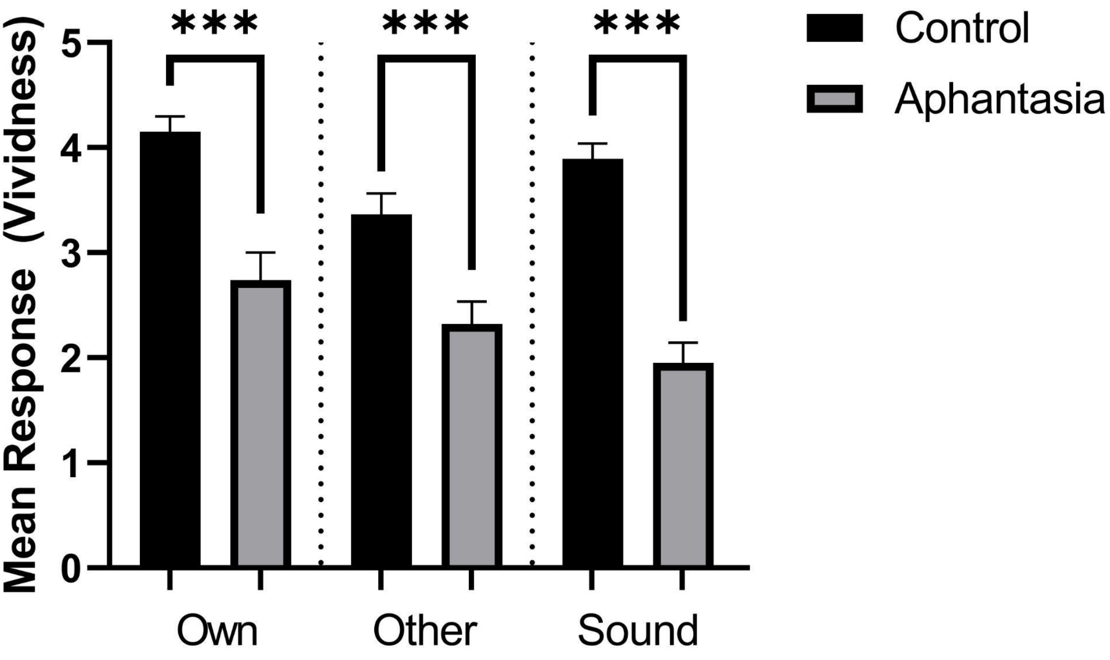
VVIQ Score Comparison

Participants in the aphantasia group exhibited significantly reduced VVIQ scores (M = 15.42; SD = 2.70) when compared to the control group (M = 59, SD = 11.56; t(36) = -21.47, p < .001). The mean VVIQ score in the aphantasia group was below the 25 point VVIQ threshold used to identify aphantasia. The mean score in the control group was above this threshold.

Scenario Rating Task

ANOVA analysis of the scenario rating task yielded a significant main effect of Group, *F*(1, 192) = 82.627, *p* < .001, and Scenario Type, *F*(2, 192) = 5.498, *p* = .005. The interaction between Group and Scenario Type did not reach statistical significance *F*(2, 192) = 2.603, *p* = .077. Post-hoc tests examining differences in response between the aphantasia group and control group per Scenario Type were all significant, with the aphantasia group reporting lower vividness in the Own condition (*t*(50) = -4.72, *p* < .001), Other condition (*t*(64) = -3.60, *p* < .001), and Sound condition (*t*(60) = -7.95, *p* < .001).

Figure 7. Performance of Aphantasia and Control group in the scenario rating task examining high intentionality inner speech.

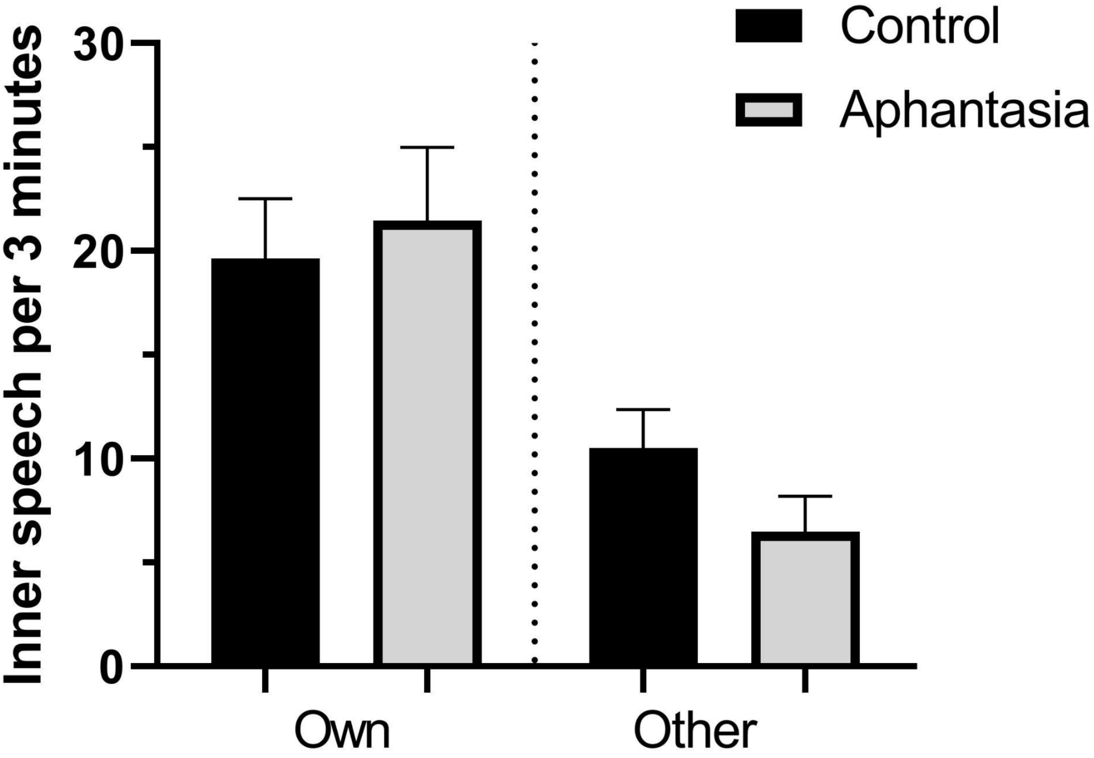


*Note: Mean vividness response per condition (higher = more vivid). Own columns correspond to scenarios in one's own voice, other columns correspond to scenarios using the voice of another person, the sound columns represent scenarios involving non-vocal sounds. Error bars represent standard error of the mean. *** = p <.001.*

Spontaneous Imagery Task

The ANOVA examining the spontaneous imagery task found a significant main effect of Scenario Type on the dependent variable [F(1, 124) = 21.670, p < .001], with fewer responses in the Other condition than the Own condition. However, no significant effects of Group (Control, Aphantasia) [F(1, 124) = 0.181, p = .671] or interaction between Group and Inner Speech Type were observed [F(1, 124) = 1.253, p = .265].

Figure 8. Performance of Aphantasia and Control group in the spontaneous imagery task examining low intentionality inner speech.



Note: Mean inner speech experiences per 3 minutes. Self columns are data collected using the own inner speech paradigm, with the other columns being data collected using the other voice paradigm. Error bars represent standard error of the mean.

Discussion

The aim of the experiment was to determine the extent to which aphantasia affects the various phenomenological subtypes of inner speech outlined in Pratts et al. (2023), and to investigate whether any effects were consistent with the dual-mechanistic model of inner speech. Inner speech which was highly intentional and egocentric inner speech was predicted to primarily utilise the motor-route of inner speech generation. If this is correct, and aphantasia represents dysfunction of perceptual simulation mechanisms, we would expect performance in these tasks to be preserved. Within the high intentionality task (scenario rating), the aphantasia group reported significantly less vivid imagery when compared to the control group for all three types of imagery (own, other, sound). The finding that the other and sound conditions were negatively impacted by aphantasia relative to the control group are in line with our hypotheses, given that they place demands on the perceptual-route of generation. However, the finding that the own condition was also impacted by aphantasia runs contrary to our hypothesis. Given that task elicited inner speech in one's own voice is highly intentional and egocentric, it should place greatest demands on the unaffected motor-route. Multiple interpretations of this finding are possible. First, the observed effect of aphantasia on highly intentional and egocentric inner speech could suggest that all subtypes of inner speech rely on perceptual simulation to some extent. This aligns with Tian et al.'s (2016) view that the motor-route and perceptual-route of imagery generation operate in a complimentary fashion in order to generate imagery. If correct, this indicates that the precise interactions between neurocognitive mechanisms and inner speech phenomenology outlined in Pratts et al. (2023) should be re-evaluated, with consideration given to the possibility that no variety of inner speech is reliant in a single mechanism of generation. It is notable, however, that this alternative hypothesis does not align with the findings of the spontaneous imagery task investigating low intentionality inner speech subtypes.

Both low intentionality inner speech subtypes (i.e. high egocentricity and low egocentricity) were expected to place demands on the perceptual-route of activation. We predicted these subtypes of inner speech would therefore occur at a lower frequency when compared to the control group. However, no differences in the rate of low intentionality inner speech were found between the aphantasia group and control group. This was true when the inner speech was in one's own voice (high egocentricity) as well as when it was in the voice of another individual (low egocentricity). As low intentionality and egocentricity inner speech is predicted to rely on the perceptual-route most heavily, given both its spontaneous nature and the limited ability of the motor-route to assist in recreating the voices of other people, this result ran strongly against our predictions. Although the causes of this cannot be wholly deduced based on the current data, the unexpecting findings further suggest unaccounted for complexity in the underpinnings of aphantasia and interactions with inner speech, justifying further research.

In proposing alternative explanations for the findings of the high intentionality task, we proposed that all subtypes of inner speech might be, to some extent, dual-mechanistic. However, the findings of the low intentionality task provide evidence to the contrary, suggesting that a preserved route of inner speech generation does exist, albeit one which does not interact with phenomenology in the manner outlined by Pratts et al. (2023). In order to further elucidate the disparate evidence, future research could first examine the possibility that aphantasia has a negative impact on the vividness of low intentionality subtypes of inner speech, rather than quantity. In this case, it is possible that the relative effects of aphantasia on inner speech do align with the hypotheses outlined in the dual-mechanistic framework, with divergences being the result of different dependent variables across the various subtypes of inner speech.

General Discussion

The series of experiments served to address two questions: (1) Given the proposal that aphantasia is caused by aberrant perceptual simulation, does aphantasia also affect auditory imagery and inner speech? (2) Do individuals with aphantasia show a preserved ability to generate subtypes of inner speech which rely on the motor-route of generation? We conducted two experiments to address these questions. The results of the first experiment demonstrated that aphantasia is not limited to decriments in the ability to generate visual imagery, but is correlated with broader deficits in other sensory modalities. With an imagery vividness questionnaire finding that auditory imagery, tactile imagery, gustatory imagery and olfactory imagery were also affected when compared to a control group.

Having demonstrated that aphantasia is correlated with deficits in the imagery of other sensory modalities, we carried out a more comprehensive experiment which consisted of a battery of inner speech tasks. These tasks were selected as they generated types of inner speech which varied in egocentricity and intentionality, the two phenomenological dimensions which have been proposed to be of most mechanistic significance according to the dual-mechanistic framework described in Pratts et al. (2023). According to this framework, inner speech places greater demands on the motor-route of generation as it increases in egocentricity and intentionality. Given our prediction that this route is unaffected by aphantasia, we predicted that highly intentional and egocentric inner speech would be relatively preserved and inner speech low in intentionality and egocentricity would be highly affected. Within the high intentionality task, we found that aphantasia had a detrimental impact on the vividness of other and sound conditions, which we predicted to rely on the perceptual route. However, we also observed aphantasia to have a detrimental impact on performance in the own voice condition, contrary to our hypothesis. Whilst this supports the notion that perceptual simulation may play a role in the generation of inner speech, how it interacts with phenomenology remains unclear. The finding that aphantasia does not affect the frequency of low intentionality inner speech subtypes, which we predicted would both place greatest demands on the affected perceptual route, raise further questions as to what role the perceptual-route plays in inner speech, and how it interacts with phenomenology.

The broader implications from this study are twofold. First, the finding that individuals with a self-identified deficit in visual imagery are also highly likely to exhibit deficits in the imagery of other sensory modalities indicates a common underlying mechanism. Aphantasia research has repeatedly examined whether the affected mechanism could relate imagery perception (i.e meta-cognition) rather than imagery production. However, recent research indicates that aphantasia is not a deficit in meta-cognition (Wicken et al. 2021; Kay et al. 2022; Keogh and Pearson 2018). If aphantasia is a result of imagery production issues, the perceptual simulation mechanism neatly explains why aphantasia is not limited to the visual domain and how imagery can be generated in a variety of sensory modalities. Importantly, if aphantasia is a production issue caused by aberrant perceptual simulation, it provides an important tool in understanding how varieties of inner speech are generated. The initial evidence generated by this study does indicate that inner speech is a dynamic phenomenon, with some subtypes being more impacted by aphantasia than others. However, the results did not align with the predictions outlined in the dual-mechanistic model (Pratts et al., 2023). This justifies future research which can further examine the interaction between the neurocognitive mechanisms of inner speech and how these might interact with phenomenology. In order to further triangulate the neurocognitive underpinnings of different types of inner speech, future research could adopt distinct experimental techniques such as dual-task or brain-stimulation paradigms in order to suppress the motor-route and perceptual-route at a cognitive level and a neurobiological level, respectively.

Conclusion

In conclusion, this series of experiments reveals that aphantasia not only impacts visual imagery but also correlates with deficits in other sensory modalities, suggesting a common underlying mechanism. The findings have important implications for understanding inner speech generation and suggest that inner speech is not only generated via motor processes, but also via perceptual simulation. Some evidence for these mechanisms of generation interacting with inner speech phenomenology was observed. Future research should utilise multiple methodologies, such as dual-task or brain-stimulation paradigms, to further explore and elucidate whether inner speech is supported by multiple mechanisms of generation, and what determines which mechanism is used.

Chapter 4 - Investigating a Dual-Mechanistic Framework of Inner Speech Using Dual-Task Interference

Jaydan Pratts[1*], Gorana Pobric[1], Bo Yao[12*]

1 Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, UK
2 Department of Psychology, Fylde College, Lancaster University, UK

*Corresponding Authors:
Jaydan Pratts
Division of Psychology, Communication and Human Neuroscience
Zochonis Building
Brunswick Street
Manchester M13 9PT jaydan.pratts@manchester.ac.uk
United Kingdom

Dr Bo Yao
Department of Psychology
Lancaster University
Lancaster LA1 4YF b.yao1@lancaster.ac.uk
United Kingdom

Abstract

The neurocognitive mechanisms which underpin inner speech generation remain the subject of debate. Some evidence suggests that inner speech emerges from corollary-discharge, a signal used to predict the consequences of motor actions. Alternatively, inner speech has been explained via perceptual simulation, the reactivation of encoded perceptual experiences. In Pratts et al. (2023) we propose a dual-mechanistic framework which argues that both corollary discharge and perceptual simulation are implicated, with the phenomenology of inner speech determining the demands placed on each mechanism. We investigate this framework using a dual-task paradigm which examines the effect of motor interference (targeting corollary discharge), and perceptual interference (targeting perceptual simulation), on inner speech varying along two phenomenological dimensions: intentionality and egocentricity. The effects of motor interference and perceptual interference on a selection of inner speech tasks were broadly in line with the predictions outlined in the framework. Within highly intentional (i.e. task elicited) inner speech, motor interference was more effective than perceptual interference when the inner speech was high in egocentricity (i.e. in one's own voice), but not when it was low in egocentricity (i.e. in someone else's voice). In inner speech which was low in intentionality (i.e. spontaneous), perceptual interference was more effective than motor interference in the high egocentricity condition, with neither interference type affecting low egocentricity inner speech. The results suggest that inner speech stems from more than one generative mechanism and that the dual-mechanistic framework could help determine how these mechanisms interact with phenomenology.

Keywords: inner speech, corollary discharge, perceptual simulation, dual-task

Highlights:

- Effects of motor and perceptual interference are dependent on inner speech phenomenology.
- Perceptual suppression mostly affects less egocentric & intentional inner speech.
- Results at odds with purely corollary discharge interpretation of inner speech.

Introduction

Two Models of Inner Speech Generation

Inner speech is an internal, speech-like experience which can be perceived in the absence of an external stimulus (Alderson-Day & Fernyhough, 2015; Perrone-Bertolotti et al., 2014). Despite its ubiquity and association with a wide variety of cognitions (Baddeley, 1992; D'Esposito, 2007; Emerson & Miyake, 2003; Filik & Barber, 2011; Miyake et al., 2004; Yao & Scheepers, 2011; Yao et al., 2011), the exact cognitive mechanisms and neural substrates which support inner speech generation remain the subject of debate (Hurlburt et al., 2016).

One approach to explaining the cognitive mechanisms underpinning inner speech is based on the corollary discharge model. This model posits that copies of articulatory motor commands are used in order to provide predictions of the perceptible consequences of the individuals own actions. These predictions are generated via forward-models and assist in determining whether a sensory input was caused by one's own actions, or whether they were caused by an external agent (Crapse & Sommer, 2008). By generating predictions of the perceptible consequences of articulatory movements whilst inhibiting the generation of actual speech sounds, it is argued that internally perceptible inner speech can be generated (Jack et al., 2019; Jacobson, 1932; Scott, 2013; Watson, 1913). As this model relies on speech production regions generating motor commands, the model can be said to use a motor-route in order to generate inner speech.

Viewing inner speech as an outcome of speech production processes elegantly explains why tasks which require the generation of inner speech (e.g. syllable counting, rhyme judgement) elicit activity in speech production brain regions such as the left inferior frontal gyrus (Marvel & Desmond, 2012; Shergill et al., 2001, 2002). However, there are several areas where the corollary discharge model of inner speech provides a less parsimonious explanation. For example, inner speech often incorporates the voices of other individuals (McCarthy-Jones & Fernyhough, 2011). These are likely to contain vocal features which are distinct from one's own and therefore not easily replicated by one's own articulators. While few studies have examined these less egocentric varieties of inner speech, some evidence indicates distinct neural activation (Brück et al., 2014). This aligns with a growing neuroimaging literature which associates certain types of inner speech with activation of speech perception regions, rather than speech production regions (Barsalou, 2008; Hurlburt et al., 2016; Yao & Scheepers, 2011; Yao et al., 2011, 2021).

The finding that some occurrences of inner speech are accompanied by activation of speech perception, rather than production, regions suggests a role for perceptual simulation in inner speech generation. Perceptual simulation refers to a mechanism by which patterns of neural activation which occurred when perceiving a sensory stimulus (e.g. when listening to someone speak) can later be reactivated in order to internally simulate the perceptual experience (Barsalou, 1999; Damasio, 1989). Notably, it is this top-down reactivation mechanism which is used to explain imagery in other sensory modalities (Dijkstra et al., 2020; Zatorre & Halpern, 2005). Viewing inner speech as the result of perceptual simulation would help explain some of the questions raised by the corollary discharge interpretation described earlier. For example, inner speech can accurately produce a wide range of voices and vocal features as it is reliant on stored memories of speech rather than the physiological abilities of one's articulatory system. As these perceptual simulations do not depend on speech production regions, it would also explain why some neuroimaging studies do not observe activation in speech production areas during spontaneous inner speech (Yao et al., 2011; Hurlburt et al., 2016).

Although the perceptual simulation model does help explain the results of neuroimaging studies investigating more diverse forms of inner speech, it also fails to provide a comprehensive model which can account for the range of observed neural correlates. For example, there is evidence that inner speech can be generated in the absence of auditory cortex activation (De Nil et al., 2000; Gulyás, 2001). This implies a lack of perceptual simulation. Moreover, repetitive transcranial magnetic stimulation (rTMS) suppression of speech production regions has been found to inhibit inner speech (Aziz-Zadeh et al., 2005). This suggests that at least some types of inner speech are generated via a mechanism reliant on speech production regions.

A Dual-Mechanistic Framework of Inner Speech

In Pratts et al. (2023), we attempted to reconcile the conflicting findings in the inner speech literature, as well as the individual limitations of the two proposed mechanisms. In doing so, we argued that it is essential to view inner speech not as an invariable phenomenon, but a multi-dimensional, flexible phenomenon which manifests in a variety of forms (Hurlburt et al., 2013; McCarthy-Jones & Fernyhough, 2011). We proposed a framework which measures the phenomenological aspects of inner speech along two axes: egocentricity and intentionality. *Egocentricity* is a measure of whether inner speech is a recreation of one’s own voice (high egocentricity), or the voice of another individual (low egocentricity). *Intentionality* represents the extent to which inner speech emerges as the result of explicit task demands (high intentionality) or spontaneously (low intentionality).

The framework also proposes that both corollary discharge *and* perceptual simulation are used to generate inner speech. These are termed the motor-route and perceptual-route of inner speech generation, respectively. The extent to which one mechanism is used over the other varies depending on where the inner speech exists along the egocentricity × intentionality space. As inner speech becomes more egocentric and more intentional, it places greater demands on the motor-route. Conversely, as inner speech becomes less egocentric or less intentional, it draws more heavily on the perceptual-route. For example, inner speech which emerges spontaneously (low intentionality), and inner speech in the voice of another person's voice (low egocentricity), would preferentially utilise the perceptual-route of generation. Inner speech generated in a paradigm asking participants to read a sentence using their inner voice would be expected to recruit the motor-route of generation.

Aim & Hypotheses

The aim of the framework described in Pratts et al. (2023) is to provide a robust model of inner speech which can explain its neural and cognitive underpinnings whilst accounting for its diverse phenomenology. Pratts et al. (2023) provide some initial evidence for this framework in the form of a neuroimaging meta-analysis. This found evidence of distinct patterns of neural activation across varieties of inner speech. However, this data was observational and a particularly limited number of available studies was noted in the low egocentricity and low intentionality domains.

Dual-task interference could provide a more direct approach of testing the hypotheses laid out in the Pratts et al. (2023). Dual-task interference is the finding that executing two tasks concurrently results in larger decrements in performance if the two tasks are reliant on the same cognitive modules or systems (Shallice et al., 1985). By asking participants to generate inner speech as they complete a secondary task designed to preoccupy either the motor-route or the perceptual-route, it should be possible to determine whether different types of inner speech are supported by distinct cognitive mechanisms and whether the framework described in Pratts et al. (2023) provides useful predictions. Similar approaches have previously been used to examine aspects of inner speech and auditory imagery, although in different contexts (Beaman et al., 2015; Chincotta

& Underwood, 1998; Johnston & McDermott, 1986). According to the framework described in Pratts et al. (2023), inner speech should increasingly rely on the motor-route of generation as inner speech becomes increasingly egocentric and intentional, and rely on the perceptual-route of generation when the reverse is true. As a result, we would expect a secondary task which occupies the motor-route to interfere with inner speech which is highly egocentric or intentional. A secondary task which occupies the perceptual-route should interfere with inner speech which is low in intentionality or egocentricity.

In this four part study we presented participants with a scenario rating task which examined highly intentional varieties of inner speech (high and low egocentricity), and a spontaneous imagery task which examined low intentionality varieties of inner speech (high and low egocentricity). During the completion of these inner speech tasks, participants were also asked to complete secondary tasks aimed at interfering with the motor-route and the perceptual-route, respectively. In order to ensure any observed suppression was not the result of the additional attentional resources demanded by a secondary task, control varieties of both motor-route interference and perceptual-route interference were tested. Based on the framework described in Pratts et al. (2023), we predicted the interaction between inner speech type and suppression effectiveness to be as follows:

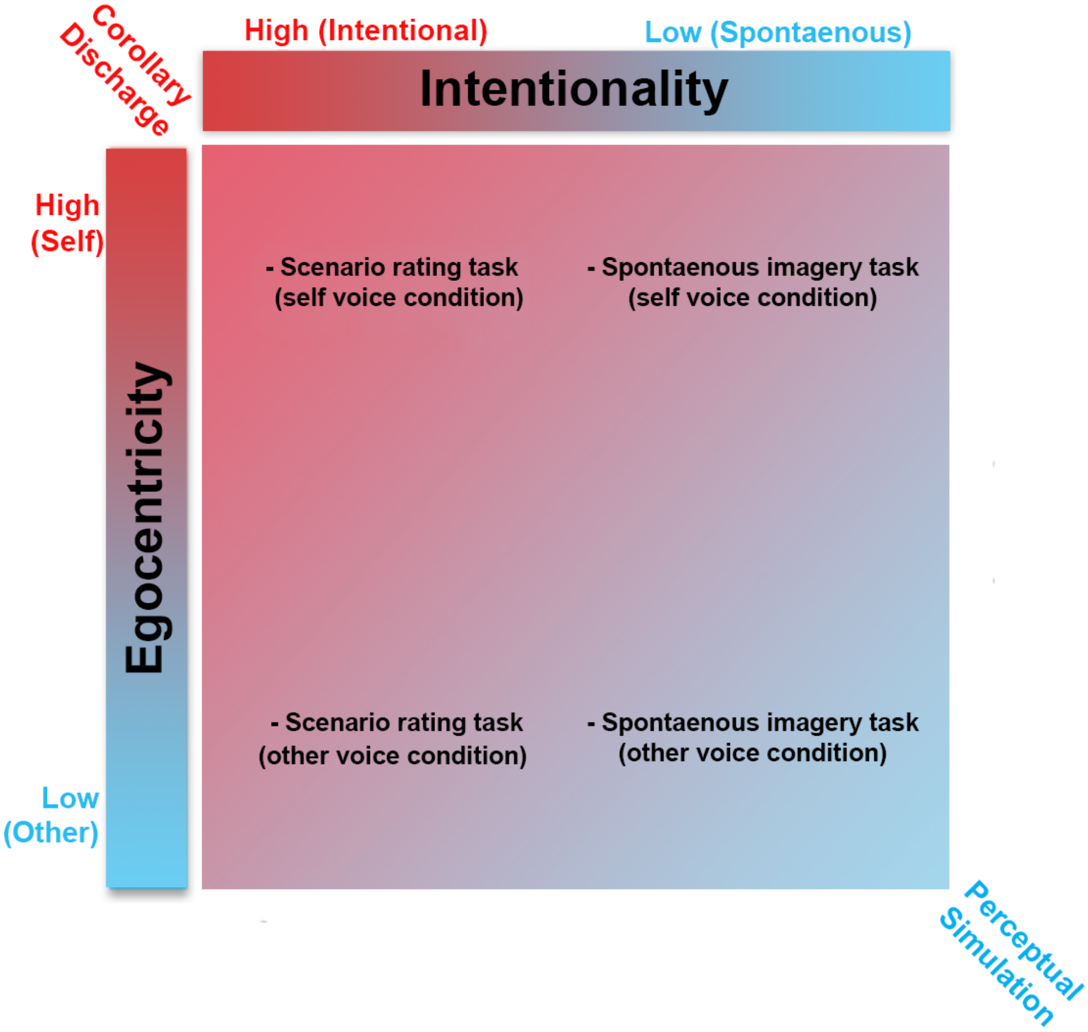
Scenario Rating Task (High Intentionality)

- *Self* Voice Condition (High Egocentricity - Suppressed by motor-route interference)
- *Other* Voice Condition (Low Egocentricity - Suppressed by motor-route & perceptual-route interference)

Spontaneous Imagery Task (Low Intentionality)

- *Self* Voice Condition (High Egocentricity - Suppressed by motor-route & perceptual-route interference)
- *Other* Voice Condition (Low Egocentricity - Suppressed by perceptual-route interference)

Figure 1. Placement of inner speech tasks and conditions used in the dual-task experiments within the dual-mechanistic framework outlined in Pratts et al. (2023).



Note: Motor-route interference is hypothesised to have a larger impact in conditions positioned in the red shaded areas as these place increasing demands on the motor-route mechanism of generation. Perceptual-route interference is predicted to impact inner speech conditions positioned in the blue shaded areas.

Experiment 1 - Testing High Intentionality Inner Speech Using Dual-Task Paradigms

In order to examine the dual-mechanistic framework, we utilised a paradigm which would test the high intentionality varieties of inner speech. The paradigm aimed to elicit inner speech which was either:

(1) High in intentionality and high in egocentricity.
(2) High in intentionality and low in egocentricity.

As participants carried out the respective tasks, they were exposed to conditions aimed at interfering with either the speech motor regions necessary to the motor-route of inner speech generation, or the voice-sensitive regions necessary to the perceptual-route of inner speech generation. We hypothesised that motor-route interference would be most effective in the high egocentricity condition, whereas perceptual-route suppression would be most effective in the low egocentricity condition. Notably, inner speech which is high in intentionality and low in egocentricity resides at a point between the extreme motor-route inner speech variant and the extreme perceptual-route inner speech variant (see Figure 1.). We therefore expect any differences to be relative rather than absolute (i.e. motor-route interference could theoretically be suppressive in the low egocentricity condition).

Methods

Participants

A total of thirty undergraduate students were recruited in exchange for course credits, with twenty-nine being present in the final dataset (23 female, *Mean*age = 20.5, *SD*age = 2.3). One participant was excluded from the final dataset due to response times which consistently exceeded the outlier threshold described in the *Data Analysis* section. Participants were native English speakers and did not report reading or language impairments.

Materials & Procedure

The study followed a repeated-measures design, with each participant completing the scenario rating task three times, each under a different condition: motor-route interference, perceptual-route interference, and no interference. In the motor-route interference condition participants were asked to silently and repeatedly mouth the word “aluminium” for the duration of the block. This was intended to generate irrelevant corollary discharge thus interfering with a principle step in the motor-route of inner speech generation. In the perceptual-route interference condition, participants completed the block whilst an Arabic audiobook played in the background. Arabic was selected in order to avoid non-phonological types of suppression (e.g. semantic interference). This was intended to tax the voice sensitive speech regions which facilitate perceptual simulation. The volume level was set to 60 decibels on Sony WH-1000XM3 stereo headphones. The no interference condition was completed without mouthing words or background audio. The order of suppression was counterbalanced across participants, with three different stimuli sets for the scenario rating task also being counterbalanced.

Scenario Rating Task

The scenario rating task contained three conditions:

Own voice: In this condition participants were given the instructions "*Using your own inner voice, say…*" followed by a sentence in prose. Participants would respond using by generating the necessary inner speech and rating the vividness of their elicited inner speech on a 5-point scale.

Other voice: In this condition, participants were given the instructions "*In another persons voice, say…*" followed by a sentence in prose. Participants would then generate the inner speech in the voice of someone familiar to them and rate the vividness of the elicited auditory imagery using a 5-point scale. Participants were instructed to choose a single individual whose voice they were familiar with at the beginning of the experiment.

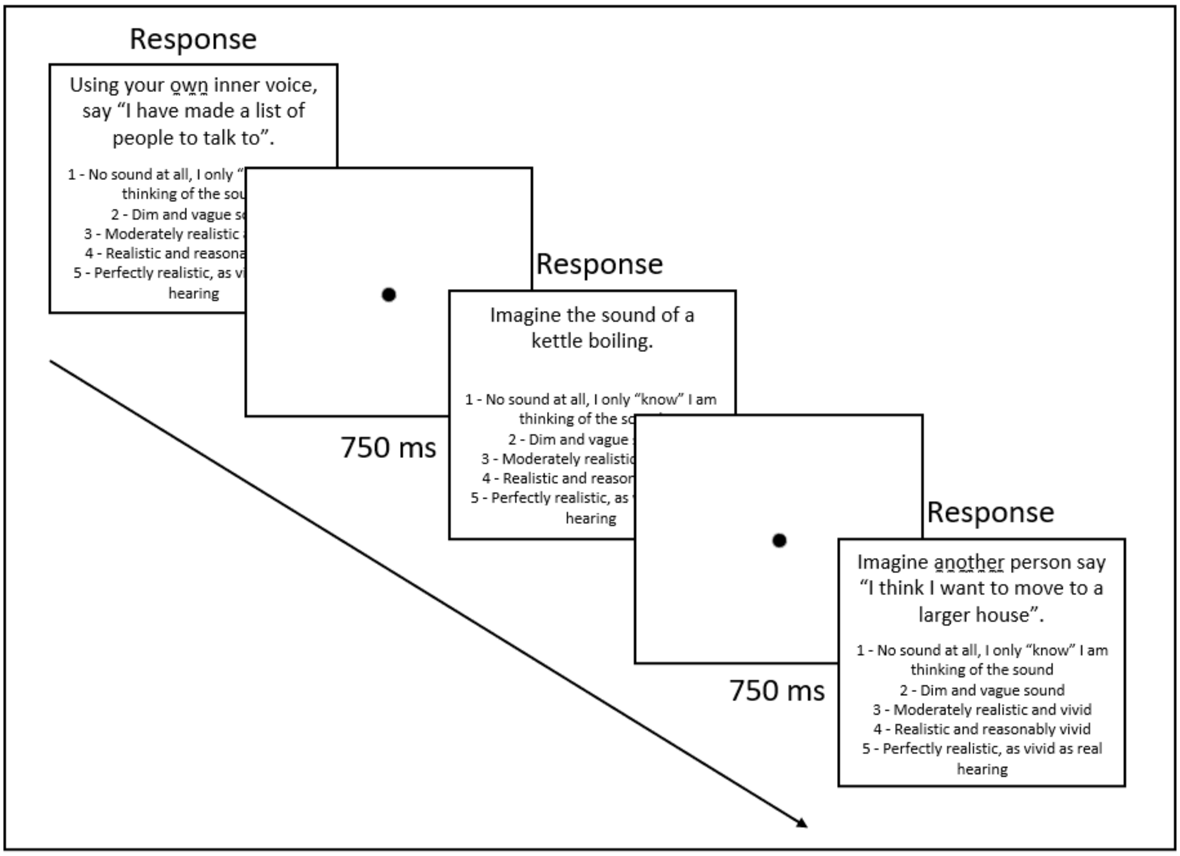
Sound: In this condition, participants were asked "*Imagine the sound of…*" followed by a description of a commonly heard non-vocal sound (e.g. rain on a metal roof). Participants would then attempt to generate auditory imagery of the given scenario and rate its vividness using a 5-point scale.

All conditions used the same scale which can be seen below in *Figure 2,* with responses entered on a keyboard using numbers 1 - 5. Participants were presented with 8 scenarios per condition, for a total of 24 scenarios. Conditions and scenarios were interspersed in a random order with a 750ms ISI separating each trial. Scenarios were not time limited in order to prevent ceiling effects. Multiple stimuli sets were rotated across participants, ensuring that stimuli were not reused across the various interference conditions.

Figure 2. 5-point scale used in the scenario rating task.

5	Perfectly realistic, as vivid as real hearing
4	Realistic and reasonably vivid
3	Moderately realistic and vivid
2	Dim and vague sound
1	No sound at all, I only “know” I am thinking of the sound

Figure 3. Procedure for the scenario rating task.



Using your own inner voice, say “I have made a list of people to talk to”.

Imagine the sound of a kettle boiling.

Imagine another person say “I think I want to move to a larger house”.

1 - No sound at all, I only “know” I am thinking of the sound
2 - Dim and vague sound
3 - Moderately realistic and vivid
4 - Realistic and reasonably vivid
5 - Perfectly realistic, as vivid as real hearing

Note: ISI is presented for 750ms, participants then presented with an own voice, other voice or sound scenario. Participants respond using keyboard numbers 1 to 5.

Data Analysis

Data analysis was completed using R (version 4.1.2) and the RStudio IDE (Build 443). Data cleaning and analysis used the *Tidyverse* (Version 1.2.1; Wickham et al., 2019) and *Rstatix* (Version 0.3.0; Kassambara, 2019) packages. The data was trimmed by grouping the data by condition and suppression type, and then excluding any trials with a response time outside the median plus / minus 2.5 times the median absolute deviation (MAD). This procedure is outlined in Leys et al. (2013) and provides several benefits over competing approaches, including standardised parameters and statistical robustness. The number of excluded trials did not exceed 5% in any condition. One participant was excluded entirely due to response times consistently outside the MAD outlier detection threshold.

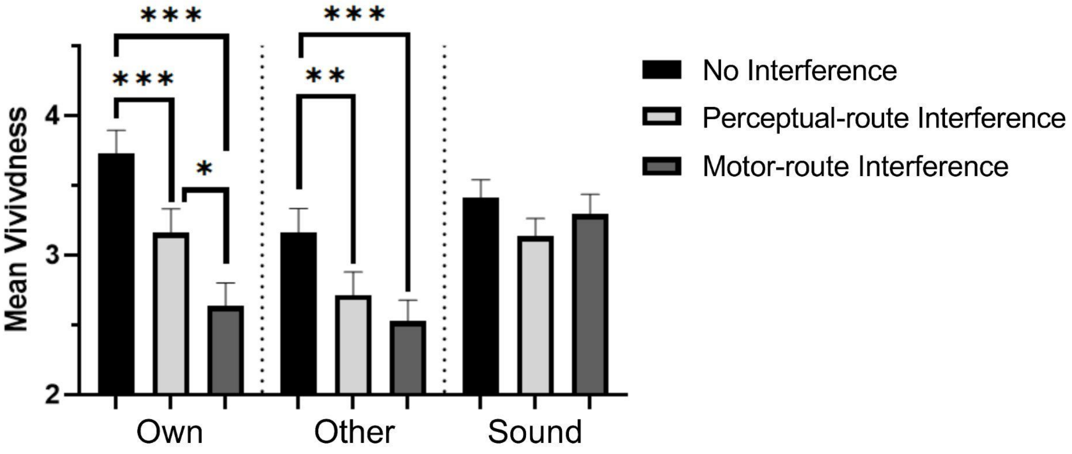
Statistical analyses were conducted using a repeated measures analysis of variance (ANOVA). The within-subject factors were Scenario Type (Own, Other, and Sound) and Interference Type (No Interference, Perceptual-route Interference, and Motor-route Interference) with vividness as the dependent variable. Post-hoc tests were conducted where appropriate using a Bonferroni correction to control for multiple comparisons.

Results

The repeated measures ANOVA revealed significant main effects of Scenario Type ($F(2, 252) = 7.445$, $p < .001$) and Interference Type ($F(2, 252) = 13.435$, $p < .001$). The interaction between Scenario Type and Interference Type was also significant ($F(4, 252) = 3.051$, $p = .018$).

Post-hoc analysis of the Own condition revealed that perceptual-route interference resulted in decreased vividness scores when compared to no interference ($p < .001$). Motor-route interference resulted in decreased vividness scores when compared to both no interference ($p < .001$) and perceptual-route interference ($p = .026$). In the Other condition, perceptual-route interference ($p = .007$) and motor-route interference ($p < .001$) resulted in decreased vividness scores when compared to no interference. There was no significant difference between motor-route interference and perceptual-route interference on vividness scores ($p = .38$). Interference Type did not have a statistically significant impact on vividness scores in the Sound condition ($p > .05$ for all comparisons).

Figure 4. Effect of motor-route and perceptual-route interference on high intentionality inner speech subtypes.



*Note: Own columns correspond to scenarios in one's own voice, Other columns correspond to scenarios using the voice of another person, the Sound columns represent scenarios involving non-vocal sounds. Error bars represent standard error of the mean. * = <.05, ** = p <.01, *** = p <.001.*

Discussion

In line with the hypothesis that different subtypes of inner speech vary in their mechanism of generation, analyses revealed that vividness scores varied as a function of interference and scenario type. Inner speech in participants' own voices was most suppressed by the motor-route interference condition. This aligns with our hypothesis that highly intentional inner speech in one’s own voice (i.e. high egocentricity) preferentially recruits the motor-route of generation. It is also notable that the perceptual-route interference condition negatively impacted vividness scores in the own voice condition. One explanation for this finding is that inner speech places some demands on the perceptual-route even when the speech is highly intentional and highly egocentric. This is in line with Tian et al. (2016) explanation of how the motor-route and perceptual-route interact in the process of imagery generation. In their dual-stream model, they argue that the two mechanisms are complementary and can combine contributions depending on the needs of the task. It is therefore possible that the effectiveness of perceptual-route interference reflects a role in providing supplementary perceptual detail in this particular paradigm. It is also possible that the suppressive effects of perceptual-route interference is caused by the increased attentional demands inherent to dual-task paradigms.

Inner speech which is highly intentional yet low in egocentricity is predicted to rely on both the motor-route and perceptual-route for generation. This hypothesis was supported by our finding that inner speech in the other voice condition was suppressed by perceptual-route interference and motor-route interference to an equal extent. It is notable that despite perceptual simulation theoretically being more suited to generating inner speech in other voices, perceptual-route interference was no more effective than motor-route interference. This could reflect the decision to allow participants to recreate the voice of whomever they wished. The advantage of allowing free choice over the recreated voice is that participants are likely to select the voice of someone they are highly familiar with. This approach was found to produce vividness scores which, in the absence of suppression, were more comparable to the own voice condition during pilot studies. The limitation of this approach is the potential for bias towards choosing persons with easily reproducible voices. In this case, we would expect the motor-route to present a more viable mechanism of generation, thereby increasing the effectiveness of motor-route interference whilst decreasing the effectiveness of perceptual-route interference.

Auditory imagery of non-vocal sounds was not suppressed by motor-route interference or perceptual-route interference. The different effects of perceptual-route interference on inner speech in other voices and imagery of non-vocal sounds is noteworthy. Both conditions were proposed to rely on the perceptual-route of generation yet no suppressive effects were observed in the non-vocal sounds condition. This could indicate that the audio book stimuli used during perceptual-route suppression specifically taxed voice sensitive brain regions and did not suppress the auditory cortex more broadly. This reflects a large body of evidence indicating that the cognitive mechanisms which facilitate speech and non-speech sound processing are distinct and specialised (Zatorre et al., 2002). At a neuroanatomical level this is evidenced by distinct regions of activation when hearing speech sounds and non-speech sounds (Shklovsky et al., 2019). Given that auditory processing might rely on different brain regions and cognitive mechanisms depending on the category of auditory stimuli, it is possible that non-vocal auditory imagery would be better suppressed by presenting participants with irrelevant non-vocal sounds.

Experiment 1b - Testing High Intentionality Inner Speech Using Dual-Task Paradigms - Foot Tapping Motor Control & White Noise Perceptual Control

The results of Experiment 1 indicate that both motor-route interference and perceptual-route interference can reduce the vividness of inner speech in both one's own voice and in the voice of others. In order to exclude the possibility that this finding stems from the increased attentional demands of carrying out a secondary task (i.e. mouthing a word or listening to irrelevant audio), two adapted versions of the experiment were created. These adaptations compared the effects of motor-route interference to a non-articulatory motor movement, and the effects of perceptual-route interference to a non-vocal audio stimulus.

Methods

Participants

Sixteen participants were recruited to complete the motor control study (12 female, *Mean*age = 20.0, *SD*age = 2.0), all participants were undergraduate students who were recruited in exchange for course credits. Participants were native English speakers and did not report reading or language impairments. An additional sixteen participants were recruited in the same manner to complete the perceptual control study (12 female, *Mean*age = 20.2, *SD*age = 1.6).

Materials & Procedure

All participants completed the same scenario rating task used in Experiment 1. Changes to the task were limited to modifications in the interference conditions used.

In the motor control study, three interference conditions were used: motor interference, motor interference control, and no interference. In the motor interference condition participants were asked to silently and repeatedly mouth the word "aluminium" for the duration of the block. In the motor interference control condition participants were asked to repeatedly tap a foot of their choice at a rapid pace for the duration of the condition. Foot tapping was monitored by the experimenter, with all participants complying with the instructions. The no interference condition was completed without mouthing words or foot tapping. The order of interference was counterbalanced across participants.

The perceptual control study used three interference conditions: perceptual interference, perceptual interference control, and no interference. In the perceptual suppression condition participants listened to an Arabic audiobook as they completed the task. In the perceptual interference control condition, participants listened to 60 dB white noise as they completed the task. The no interference condition was completed without any audio playing. The order of suppression was counterbalanced across participants.

Data Analysis

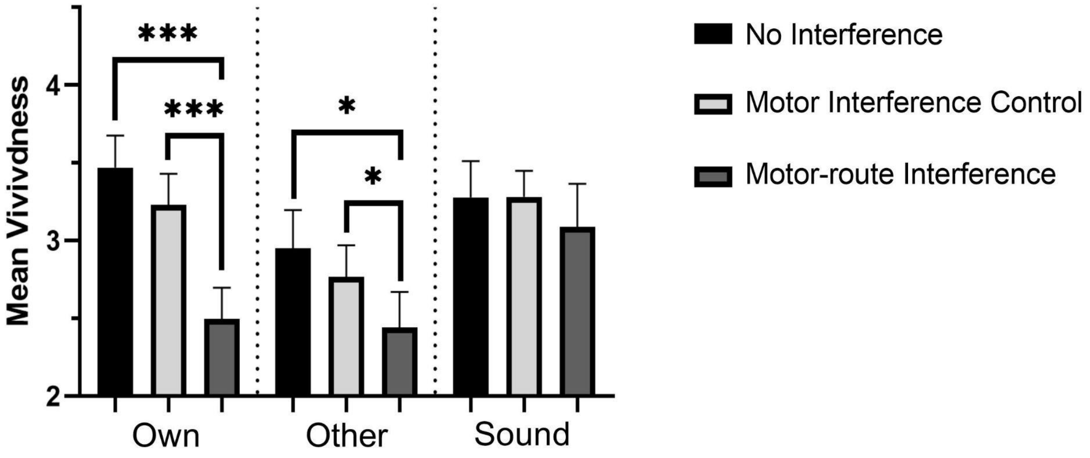
For both the motor control study and perceptual control study, data analysis involved the same steps and parameters as those adopted in Experiment 1. Analysis used R (version 4.1.2) and the RStudio IDE (Build 443). Data cleaning and analysis used the *Tidyverse* (Version 1.2.1; Wickham et al., 2019) and *Rstatix* (Version 0.3.0; Kassambara, 2019) packages. Trials were excluded based on Leys et al.'s (2013) recommended criteria of the median plus / minus 2.5 times the median absolute deviation. This was done on a by-condition and by-suppression basis, with neither study experiencing data loss of over 5% of trials. For each task, the resulting data was tested using a repeated-measures analysis of variance (ANOVA) tests on vividness scores with Scenario Type (Own, Other, Sound) and Interference Type (Motor-route / Perceptual-route Interference, Motor / Perceptual Interference Control, and No Interference) as within-subject factors. Post-hoc tests were conducted using a Bonferroni correction where appropriate.

Results

Motor Interference Control

The repeated measures ANOVA revealed significant main effects of Scenario Type ($F(2, 135) = 4.012$, $p = .020$) and Interference Type ($F(2, 135) = 5.205$, $p = .007$). The interaction between Scenario Type and Interference Type was not significant ($F(4, 135) = 0.858$, $p = .491$). Post-hoc analysis of the Own condition found reductions in vividness in the motor interference condition compared to both no interference ($p < .001$) and motor interference control ($p < .001$) conditions. No differences were observed between motor interference control and no inference conditions ($p = .146$). Analysis of the Other condition observed similar results, with reduced vividness in the motor interference condition when compared to both the motor interference control ($p = .028$) and no interference ($p = .022$) conditions, and no difference between motor interference control and no interference conditions (p = .423). No significant comparisons were observed in the Sound condition (p > .05 for all comparisons).

Figure 5. Effect of motor-route interference and motor interference control on high intentionality inner speech subtypes.



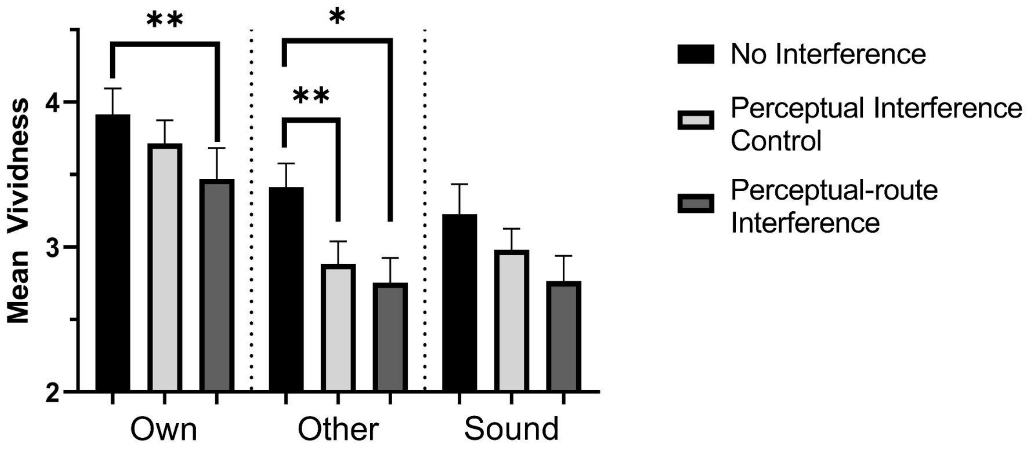
*Note: Mean response per condition. Own columns correspond to scenarios in one's own voice, Other columns correspond to scenarios using the voice of another person, the Sound columns represent scenarios involving non-vocal sounds. Error bars represent standard error of the mean. * = <.05, *** = p <.001.*

Perceptual Interference Control

The ANOVA revealed significant main effects of Scenario Type (F(2, 135) = 15.575, p < .001) and Interference Type (F(2, 135) = 6.640, p = .002). The interaction between Scenario Type and Interference Type was not significant (F(4, 135) = 0.265, p = .900).

Post-hoc analysis of the Self condition found reductions in vividness in the perceptual interference condition compared to the no interference condition (p = .009). No differences between perceptual interference control and no interference (p = .461), or perceptual interference (p = .543), were found. Analysis of the Other condition found both perceptual interference (p = .017) and perceptual interference control (p = .008) to reduce vividness when compared to no interference. No significant difference between perceptual interference and perceptual interference control were observed (p = 1.00). No significant comparisons were observed in the Sound condition (p > .05 for all comparisons).

Figure 6. Effect of perceptual-route interference and perceptual interference control on high intentionality inner speech subtypes.



*Note: Mean vividness per condition. Own columns correspond to scenarios in one's own voice, Other columns correspond to scenarios using the voice of another person, the Sound columns represent scenarios involving non-vocal sounds. Error bars represent standard error of the mean. * = <.05, ** = p <.01.*

Discussion

The results of the motor interference control study indicate that the suppressive effects of motor interference on high intentionality inner speech subtypes cannot be attributed to the increased attentional or cognitive demands placed on the participants. In line with our proposal, the effect of motor-route interference likely stems from the preoccupation of the orofacial muscles used in speech generation. At a neural mechanism level, the suppressive effect of these orofacial movements can be attributed to the generation of irrelevant corollary discharge whilst inhibiting the generation of relevant corollary discharge.

The results of the perceptual interference control study replicated the finding that perceptual interference leads to reduced vividness in own voice and other voice conditions when compared to no interference. However, no significant difference was observed between perceptual interference, and the perceptual interference control condition which consisted of white noise rather than vocalisations. This could suggest that the effects of perceptual interference relate to the increased attentional demands of having irrelevant audio play during an inner speech task. It is also possible that white noise is able to place demands on the auditory regions needed for perceptual simulation, despite lacking a vocal component. This presents an interesting topic for future research.

Experiment 2 - Testing Low Intentionality Inner Speech Using Dual-Task Paradigms

In order to examine the two low intentional inner speech subtypes within the dual-mechanistic framework, we utilised two further paradigms. These two paradigms aimed to elicit inner speech which was either:

(1) Low in intentionality and high in egocentricity.
(2) Low in intentionality and low in egocentricity.

As with Experiment 1, participants carried out the tasks as they were exposed to secondary task conditions aimed at interfering with either the motor-route of inner speech generation or the perceptual-route of inner speech generation. Given that the framework posits that all low intentionality inner speech subtypes place demands on the perceptual route of generation, we hypothesised that perceptual-route interference would significantly attenuate both low egocentricity and high egocentricity inner speech conditions.

Methods

Participants

A total of thirty undergraduate students were recruited in exchange for course credits. Participants were native English speakers and did not report reading or language impairments. Four participants were excluded as they did not complete the motor suppression procedure (gum chewing) for the full duration of the task, leaving twenty-six participants in the final analysis pool (21 female, *Mean*age = 20.7, *SD*age = 1.8).

Materials & Procedure

The experiment followed a repeated-measures design, with each participant completing the spontaneous imagery task three times, once per interference condition: motor-route interference, perceptual-route interference, and no interference. In the motor-route interference condition participants were asked to vigorously chew gum for the duration of the block. This followed the procedure set out in Beaman et al. (2015), including the use of ELMA sugar-free mastic gum. Chewing gum was chosen over the mouthing of a word like “aluminium” as spontaneous inner speech is particularly sensitive to the effects of task demands (Jin et al., 2020; Rummel & Boywitt, 2014). It is therefore likely that the silent mouthing of a word would have negatively affected spontaneous inner speech for reasons unrelated to the interruption of its generative mechanism.

In the perceptual-route interference condition participants completed the block whilst an Arabic audiobook played in the background. As in Experiment 1, Arabic was selected in order to avoid non-phonological types of suppression (e.g. semantic interference). The volume level was set to 60 decibels on Sony WH-1000XM3 stereo headphones. The no interference condition was completed without chewing gum or background speech. The order of suppression and order of tasks was counterbalanced across participants.

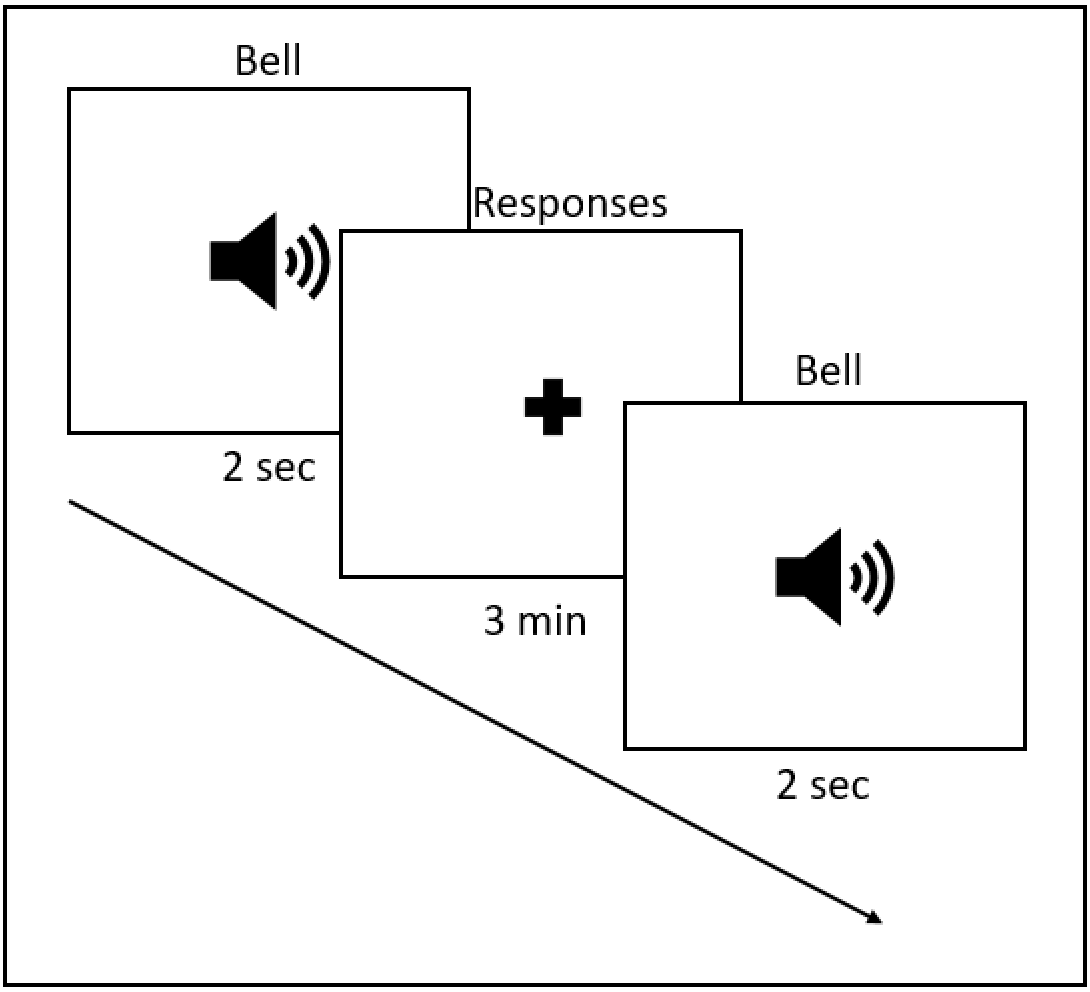
The spontaneous imagery task consisted of two conditions which were completed in a counterbalanced order, an *own voice* condition which examined the high egocentricity variant of inner speech, and an *other voice* condition which examined the low egocentricity variant of inner speech:

In the *own voice* condition, participants were presented with a fixation cross for three minutes. During this time participants were instructed to close their eyes, relax and press the spacebar button if they heard their inner voice. They were permitted to press the button multiple times. An audible bell was used to mark the beginning and end of the three minute period. The instructions made clear that the spacebar should be pressed per occurrence rather than per word. As an example, participants are given the hypothetical inner speech experience “I wonder what the end bell will sound like” and are told that such a thought should result in a single keypress.

In the *other voice* condition participants first heard a 30 second audio clip of a movie quotation being repeated. This quotation was from the television show Doctor Who and consisted of a Dalek repeating the phrase “Exterminate!”. Participants were instructed *not* to think of a Dalek saying “Exterminate!”. Participants then underwent 3 minutes of silence marked by an audible bell at the beginning and end of the period, as in the *own voice* condition. During this period participants were tasked with pressing the response key every time they experienced auditory imagery of a Dalek saying “Exterminate!”.

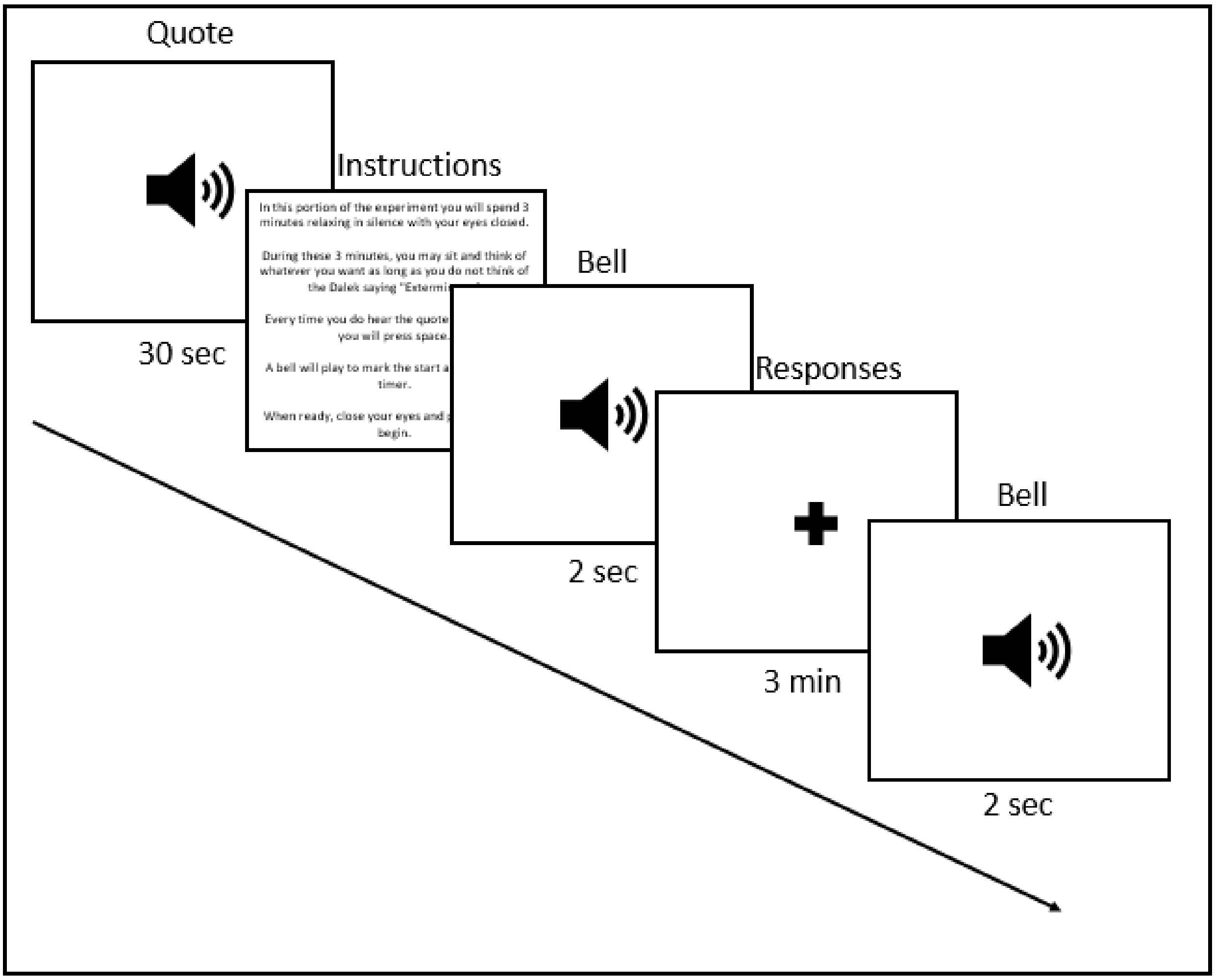
Both conditions were preceded by short examples of the task. In these examples, participants spent 30 seconds instead of 3 minutes in silence. In order to preserve the novelty of the stimuli, the *other voice* practice condition used a quotation from Star Wars, in which Darth Vader says “No, I am your father.”

Figure 7. Spontaneous imagery task - Own voice condition.



Note: *When ready, participants close their eyes and an audible bell marks the beginning of the 3 minute response period. During this period they may press the response key if they hear their inner voice. A final bell marks the end of this period.*

Figure 8. Spontaneous imagery task - Other voice condition.



Note: The movie quotation is presented auditorily for 30 seconds, participants are then presented with instructions. When ready, participants close their eyes and an audible bell marks the beginning of the 3 minute response period. During this period they may press the response key if they "hear" the movie quote internally. A final bell marks the end of this period.

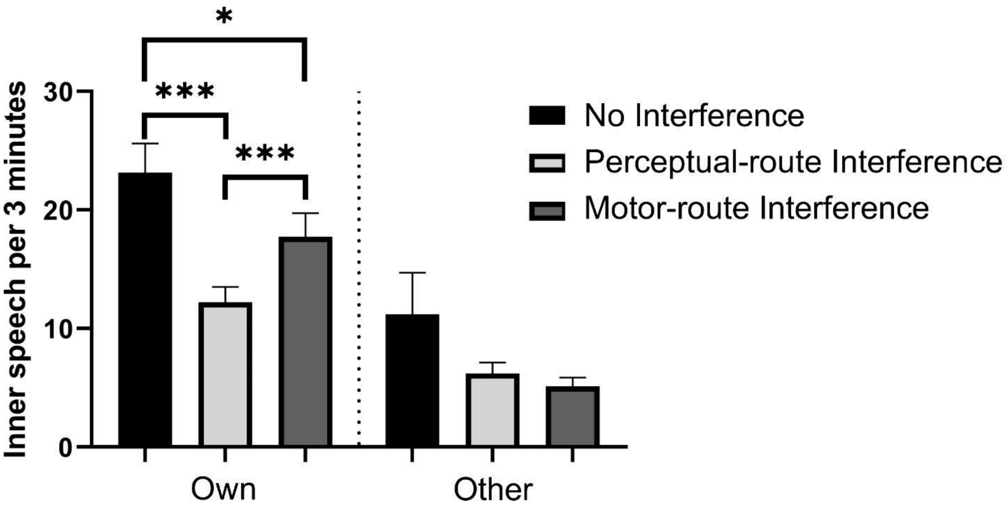
Data Analysis

Analysis of all data was completed using R (version 4.1.2) and the RStudio IDE (Build 443). Data cleaning and analysis used the *Tidyverse* (Version 1.2.1; Wickham et al., 2019) and *Rstatix* (Version 0.3.0; Kassambara, 2019) packages. Outlier detection was based on a predetermined threshold of ≥ 180 responses per trial. This is equivalent to participants reporting one spontaneous thought per second for the entire three minute trial. No trials were excluded based on this criteria. Statistical analysis consisted of a repeated-measures two-way ANOVA analysing spontaneous thoughts per three minutes, with Scenario Type (Own, Other) and Interference Type (Motor-route Interference, Perceptual-route Interference, No Interference) as within-subject factors. Post-hoc tests utilised a Bonferroni correction.

Results

The ANOVA revealed a significant main effect of Scenario Type (F(1, 156) = 36.715, p < .001), with more spontaneous thoughts generated in the Own voice condition compared to the Other voice condition. A significant main effect of Suppression Type (F(2, 156) = 7.919, p < .001) was also found. However, the interaction between Scenario Type and Suppression Type was not significant (F(2, 156) = 1.568, p = .212). Post-hoc analyses examined the differences between interference types within the Own and Other condition. In the Own condition, there were fewer spontaneous thoughts in the perceptual-route interference condition than the motor-route interference condition (p < .001) and no interference condition (p < .001). The motor-route interference condition also demonstrated a reduction in spontaneous thoughts when compared to no interference (p = .050). No significant differences were found between suppression types within the Other condition (all comparisons p > .05).

Figure 9. Effect of motor-route interference and perceptual-route interference on low intentionality inner speech subtypes.



*Note: Mean inner speech experiences per 3 minutes. Own columns are data collected using the own inner speech paradigm, with the other columns being data collected using the other voice paradigm. Error bars represent standard error of the mean. * = <.05, *** = p <.001.*

Discussion

The findings of Experiment 2 are broadly consistent with the proposal that less intentional (i.e. more spontaneous) forms of inner speech predominantly rely on a perceptual-route of generation rather than a motor-route of generation. Low intentionality inner speech in one's own voice was proposed to rely on the perceptual-route of inner speech generation, with some involvement of the motor-route possible given its high egocentricity. In line with this, perceptual-route interference was significantly more suppressive than both motor-route interference and no interference. This suggests a vital role for speech perceptual regions in the generation of at least some varieties of spontaneous inner speech, in line with proposal of a perceptual-route of inner speech generation, as well as the findings of Pratts et al. (2023) and Hurlburt et al. (2016).

Motor-route interference was also found to result in statistically significant attenuation of spontaneous inner speech in one’s own voice, albeit to a lesser degree that perceptual-route interference. This aligns with our prediction of some involvement of the motor-route given that the spontaneous inner speech was of high egocentricity. However, it does not align with the findings of Hurlburt et al. (2016), who reported significantly decreased activity in regions associated with the motor-route during spontaneous / low intentionality inner speech. Several explanations for this divergence are worth exploring. First, our findings could represent a genuine role for the motor-route in low intentionality inner speech production which was not observed in Hurlburt et al. (2016) for methodological reasons. This would align with Grandchamp et al. (2019) observations. Second, our findings could reflect differences in the elicitation paradigms adopted. Hurlburt et al. (2016) adopted a direct experience sampling approach which involves participants recording their internal state at random moments in time. This arguably generates less intentional inner speech as a greater degree of participant blinding can be maintained. This would align with the notion that intentionality is a continuous rather than discrete dimension, with our paradigm representing inner speech which is low in intentionality, but not maximally so. Finally, the observed effect of motor interference could reflect increased attentional demands caused by the requirement to carry out task-irrelevant motor movements, rather than suppression derived from the preoccupation of orofacial muscles - this possibility is explored in Experiment 2b.

We also hypothesised that inner speech in other voices would exhibit similar characteristics, with perceptual-route interference being more effective than motor-route interference and no interference . According to the dual-mechanistic model, inner speech which is low in intentionality and egocentricity should place minimal demands on the motor-route, thereby negating any effects of motor-route interference. However, analysis of the *other* condition did not yield statistically significant effects for either motor-route interference or perceptual-route interference when compared to no interference. Given a visual trend towards an effect of both motor-route and perceptual-route interference, it is possible that the lack of statistically significant effects was caused by insufficient statistical power. This might be compounded by the choice of low egocentricity elicitation paradigm which relies on the paradoxical effects of thought suppression instructions (Abramowitz et al., 2001) rather than a more traditional sampling approach. This choice reflects the difficulties of experimentally examining inner speech which is low in both egocentricity and intentionality, with alternative elicitation paradigms failing to generate sufficient data during pilot experiments. Therefore, whilst we cannot discount the possibility that the lack of statistically significant results stems from mechanistic differences in inner speech generated via the particular elicitation paradigm, methodological challenges preclude the use of more traditional approaches.

In sum, the results of Experiment 2 provide some evidence indicating that a perceptual-route of generation exists and subsumes at least some subtypes of low intentionality inner speech. The results are, however, open to the criticism that the mechanisms driving the perceptual-route suppression are unclear. In turn, this could impact our interpretation of the observed effects. We argue that the suppressive effect of auditory stimuli on inner speech is caused by interfering with the reactivation of encoded neural firing patterns, a principal step in the generation of inner speech via the perceptual-route. This limitation on the perceptual simulation mechanism has not been extensively examined in the context of inner speech, but evidence can be drawn from research on other imagery domains. This research is typically associated with the Perky effect (Perky, 1910), the observation that visual imagery results in degraded visual perception (Ishai & Sagi, 1995). Suppression has also been observed in the opposite direction, with perception degrading the ability to generate visual imagery, analogous to our own experiments (Pearson et al., 2008). This effect is not limited to the visual domain (Djordjevic et al., 2004; Segal & Fusella, 1969), suggesting that the effect emerges from general imagery generation mechanisms rather than processes specific to visual imagery. We adopt the view which has been used to explain the Perky effect, in which the interaction between perception and imagery arises from their perception and perceptual simulation (i.e. imagery) relying on the same neural substrates, with only the manner of activation differing (i.e. top-down vs. bottom-up) (Ishai & Sagi, 1995; Palmiero et al., 2019). However, we accept that further research is needed to investigate the generative mechanism in and of itself - a shortcoming also applicable to the motor-route approach. Until that point, it is not entirely possible to rule out alternative interpretations. For example, it could be argued that the auditory stimuli in perceptual-route interference impedes the perception, rather than generation, of inner speech.

Experiment 2b - Testing Low Intentionality Inner Speech Using Dual-Task Paradigms - Foot Tapping Motor Control & White Noise Perceptual Control

The findings of Experiment 2 indicate that both motor-route and perceptual-route interference significantly impact the generation of spontaneous inner speech in one's own voice. As both types of interference involve the addition of a secondary stimulus which imparts additional attentional demands, it is necessary to determine whether the effect of interference were due to increased cognitive demands, or whether it reflects genuine suppression of the mechanism of generation. To do this, two variants of Experiment 2 were created and tested. One variant replicated Experiment 2 with the addition of a motor control condition involving foot tapping. This was intended to induce a non-orofacial motor movement. A second task replicated Experiment 2 whilst comparing the effects of perceptual-route suppression to a non-vocal perceptual control condition, consisting of white noise.

Methods

Participants

A total of sixteen undergraduate students were recruited for the motor control experiment. Participants were native English speakers with no reported language or reading impairments (14 female, *Mean*age = 19.6, *SD*age = 1.3). An additional sixteen undergraduate students were recruited for the perceptual control experiment. Participants were also native English speakers with no reported language or reading impairments (16 female, *Mean*age = 19.5, *SD*age = 0.9). All participants were compensated for their time using course credits.

Materials & Procedure

Participants completed adapted versions of the spontaneous imagery task used in Experiment 2, with alterations being made to the suppression conditions. The order of suppression conditions were counterbalanced across participants in both experiments.

The motor control experiment consisted of three interference conditions: motor-route interference, motor interference control, and no interference . No interference and motor-interference conditions were replicated from Experiment 2 and consisted of either no secondary task, or the chewing of gum. The motor interference control condition required participants to repeatedly tap a foot of their choice at a rapid pace for the duration of the trial. Foot tapping was monitored by the experimenter, with all participants following the instructions correctly.

The perceptual control experiment consisted of three interference conditions: perceptual interference, perceptual interference control, and no interference. As in Experiment 2, no interference and perceptual interference consisted of either no secondary task, or the playing of an Arabic audiobook via headphones, respectively. In the perceptual interference control condition, participants listened to 60 dB white noise via headphones as they completed the task.

Data Analysis

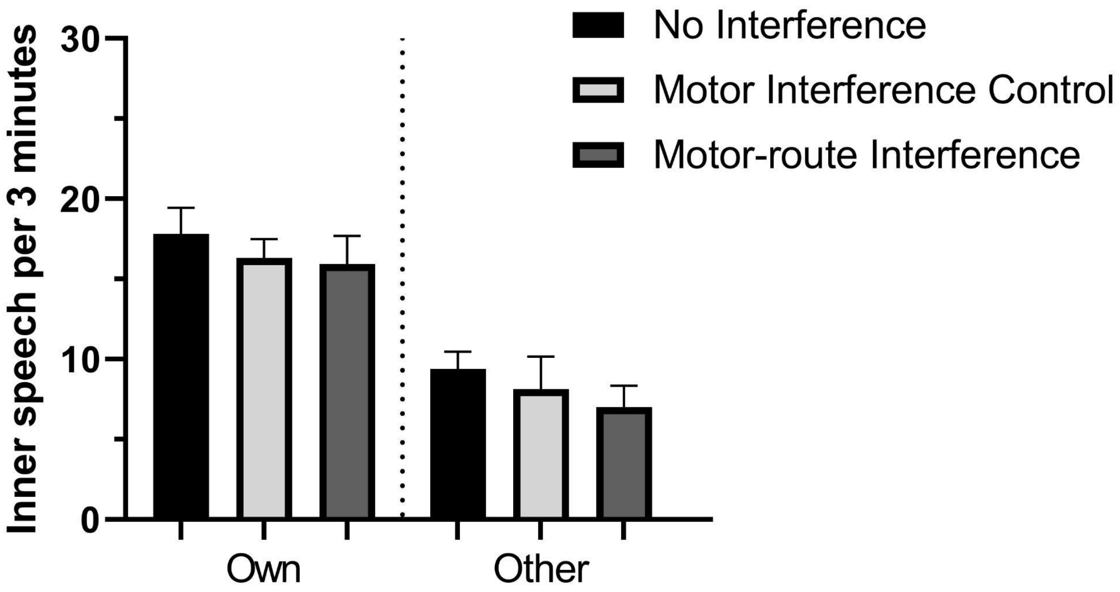
For both experiments, data analysis was completed using the same procedure and parameters used in Experiment 2. This involved the use of R (version 4.1.2) and RStudio IDE (Build 443), along with *Tidyverse* and *Rstatix* packages. As in Experiment 2, we established an outlier detection protocol in which trials with ≥ 180 responses were excluded. This is equivalent to a participant recording an inner speech experience at least once per second for the entire three minute trial. No trials were excluded as no participants reached or exceeded this threshold. Statistical analysis used a repeated-measures two-way ANOVA with Scenario Type (Own, Other) and Suppression Type (Motor-route / Perceptual-route interference, Motor / Perceptual interference control, No Interference) as within-subject factors, and spontaneous thoughts per three minutes as the dependent variable. Post-hoc tests using Bonferonni's correction were conducted when appropriate.

Results

Motor Interference Control

The ANOVA revealed a significant main effect of Scenario Type (F(1, 90) = 45.825, p < .001), indicating a substantial difference between the Own and Other scenario types. However, there was no significant main effect of the Interference Type (F(2, 90) = 0.977, p = .38), on the number of inner speech occurrences. Moreover, the interaction between Scenario Type and Interference Type was not significant (F(2, 90) = 0.031, p = .97).

Figure 10. Effect of motor-route interference and motor interference control on low intentionality inner speech subtypes.

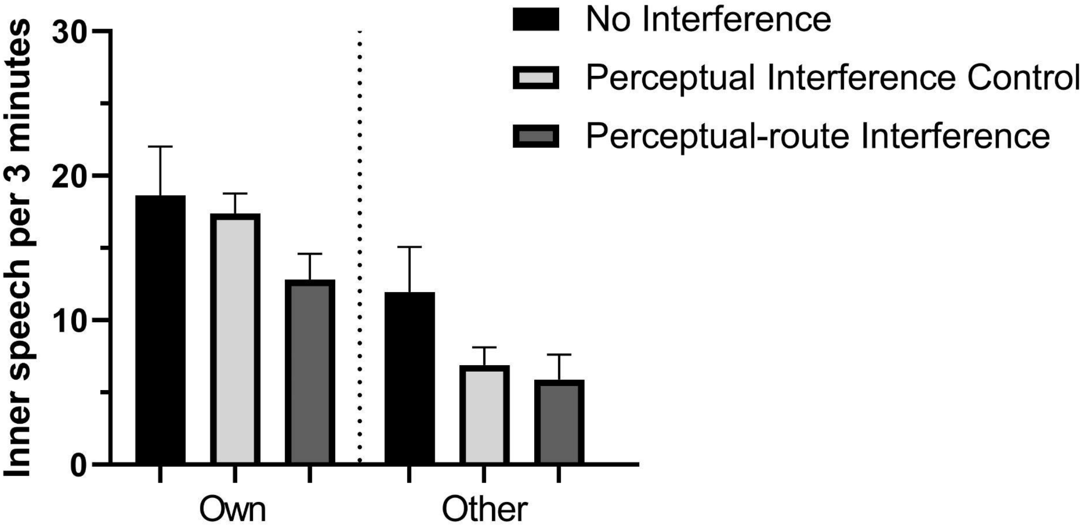


Note: Mean inner speech experiences per 3 minutes. Own columns are data collected using the own voice inner speech paradigm, with the Other columns being data collected using the other voice paradigm. Error bars represent standard error of the mean.

Perceptual Interference Control

The ANOVA revealed a significant main effect of Scenario Type (F(1, 90) = 12.824, p < .001), with fewer inner speech occurrences in the Other condition compared to the Own condition. However, there was no significant main effect of the Interference Type (F(2, 90) = 1.937, p = 0.15), suggesting that the type of interference did not have an impact on the number of inner speech occurrences. The interaction between Scenario Type and Interference Type did not reach statistical significance (F(2, 90) = 0.098, p = 0.906).

Figure 11. Effect of perceptual-route interference and perceptual interference control on low intentionality inner speech subtypes.



Note: Mean inner speech experiences per 3 minutes. Own columns are data collected using the own voice inner speech paradigm, with the Other columns being data collected using the other voice paradigm. Error bars represent standard error of the mean.

Discussion

Both the motor interference control experiment and perceptual interference control experiment failed to replicate the findings of Experiment 2, where both interference types were found to reduce rates of inner speech in one's own voice. Whilst statistically significant suppression of auditory imagery using similar techniques has previously been observed with similar sample sizes (Beaman et al., 2015), it is possible that this represents a Type II error due to insufficient statistical power. In light of this limitation, further research is required in order to determine whether the suppressive effects of perceptual-route and motor-route interference on spontaneous inner speech in one’s own voice represents genuine suppression of the generative mechanisms underpinning inner speech, or the effects of greater attentional demands in a dual-task environment.

General Discussion

The present study demonstrated that the suppressive effects of dual-task interference on inner speech varies depending on inner speech type and whether interference targets the motor-route or the perceptual-route. In highly intentional inner speech in one's own voice, we found that motor-route interference is most effective (i.e.

attenuates to the greatest degree) at suppression. However, when asked to generate inner speech in the voice of another individual, perceptual-route interference becomes as effective as motor-route interference (Experiment 1). Interference of the motor-route had no effect on the generation of non-vocal sounds, highlighting the limitations of the motor-route in generating auditory imagery which is not easily articulable (Experiment 1). In low intentionality (i.e. spontaneous) inner speech in one's own voice, we found that perceptual-route interference is most effective at reducing rates of inner speech. Interestingly, neither motor-route or perceptual-route interference significantly affected rates of spontaneous inner speech in other voices (Experiment 2).

The current results are at odds with the view that inner speech generation invariably relies on speech production regions (Scott, 2013). The results suggest that while some types of inner speech do rely heavily on speech production regions and speech articulators, others do not rely on this system to an equal degree. This can be interpreted as support for a more flexible interpretation of inner speech, such as that outlined in Pratts et al.'s (2023) framework as well as Yao et al., (2011), Hurlburt et al. (2016) and Tian et al. (2016). In addition to proposing that two mechanisms of generation exist, Pratts et al. (2023) attempts to link these two generative mechanisms to inner speech phenomenology. According to this model, the involvement of the motor-route is linked to the egocentricity and intentionality of the inner speech. As inner speech becomes less egocentric and less intentional, the demands on the motor-route are reduced and the demands on the perceptual-route increase. This prediction is generally supported by the data. In the high intentionality domain, there was a decrease in the effectiveness of motor-route interference as the inner speech decreased in egocentricity. Motor-route interference also showed diminished effectiveness in the low intentionality subtypes, with no effect on spontaneous inner speech in other voices and inferior suppression of spontaneous inner speech in one's own voice when compared to perceptual-route interference.

The framework outlined in Pratts et al. (2023) is an early attempt at explaining the full spectrum of inner speech experiences and their underlying neurocognitive mechanisms. However, we argue that this initial framework provides a more cohesive and parsimonious explanation of the results than those provided by competing models. For example, if adopting a more traditional view in which inner speech is invariably the result of corollary-discharge, it is not clear why the effectiveness of motor-route interference would vary across inner speech subtypes. Moreover, the inability of motor-route interference to inhibit imagery of non-vocal sounds provides strong evidence in favour of an alternative route of auditory imagery generation. The existence of a perceptual-route mechanism for generating auditory imagery has also been theorised and evidenced in the field of musical imagery research (Zatorre & Halpern, 2005). A question requiring further investigation therefore relates to whether this separate mechanism which supports non-vocal auditory imagery and musical imagery also supports inner speech which cannot be easily subvocalised by the individual.

Several research questions remain unaddressed due to limitations on the explanatory power of dual-task paradigms. One concern is whether the two mechanisms we have identified at a cognitive level reside within the neuroanatomical locations outlined in Pratts et al. (2023) and Tian et al. (2016). For the motor-route, this location corresponds to a pathway originating in the left inferior frontal gyrus, the location wherein speech-derived corollary discharge originates (Tourville & Guenther, 2011). For the perceptual-route, this location is centred around the left superior temporal sulcus / gyrus, regions involved in speech perception (Alderson-Day et al., 2020; Yao et al., 2011, 2012). Given that dual-task paradigms cannot map behaviours to specific brain regions, future studies should adopt neuroimaging methodologies in order to determine whether neural correlates differ across varieties of inner speech.

In sum, the study provides evidence indicating that inner speech can vary in its generative mechanisms. These generative mechanisms are likely linked to the phenomenology of the generated inner speech. Whilst this remains a relatively unexplored area, the phenomenological dimensions of intentionality and egocentricity appear to be cognitively significant and could serve as useful indicators of which generative mechanism will be used.

Conclusion

In conclusion, the present study highlights the varying effectiveness of dual-task interference on inner speech depending on its type and the suppression target, either the motor-route or the perceptual-route. The results challenge the notion that inner speech generation solely relies on speech production regions and corollary discharge, instead offering support for a more flexible interpretation of inner speech as proposed by Pratts et al. (2023) and other researchers (Tian et al., 2016). The findings emphasise a role for the phenomenological dimensions of intentionality and egocentricity as significant factors which influence generative mechanisms employed in inner speech. These findings serve as a basis for future research investigating the interaction between inner speech phenomenology and underlying neurocognitive mechanisms.

Chapter 5 - Inner Speech Suppression Using Transcranial Magnetic Stimulation: An Investigation of a Dual-Mechanistic Framework of Inner Speech

Jaydan Pratts[1*], Bo Yao[12*], Gorana Pobric[1]

1 Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, UK
2 Department of Psychology, Fylde College, Lancaster University, UK

*Corresponding Authors:
Jaydan Pratts
Division of Psychology, Communication and Human Neuroscience
Zochonis Building
Brunswick Street
Manchester M13 9PT jaydan.pratts@manchester.ac.uk
United Kingdom

Dr Bo Yao
Department of Psychology
Lancaster University
Lancaster LA1 4YF b.yao1@lancaster.ac.uk
United Kingdom

Abstract

The mechanisms by which inner speech is generated remain unclear. One proposal is that inner speech is a sensory prediction of intended speech using corollary discharge. This mechanism implicates the left inferior frontal gyrus. An alternative explanation is that inner speech is derived from the reactivation of neurons in perceptual regions, in order to simulate perception. This model implicates the left superior temporal gyrus / sulcus. Here, we examine a dual-mechanistic framework which argues that both of these mechanisms exist, with their contribution varying depending on the phenomenology of the inner speech. We demonstrate evidence inconsistent with this hypothesis by applying rTMS to the L-IFG and L-STG and monitoring performance on inner speech subtypes varying in phenomenology. Stimulation of the L-IFG or L-STG did not lead to impairments of any subtype of inner speech. Implications for a dual-mechanistic framework are discussed.

Keywords: inner speech, corollary discharge, perceptual simulation, transcranial magnetic stimulation

Highlights:

- We test the involvement of speech production and perceptual regions in inner speech.
- No effect of L-IFG or L-STG stimulation on inner speech performance was observed.
- Findings suggest the need for further research on the neural underpinnings of inner speech.

Introduction

Neurocognitive Models of Inner Speech Generation

Inner speech is an internal, speech-like experience which can be perceived in the absence of an external stimulus (Alderson-Day & Fernyhough, 2015; Perrone-Bertolotti et al., 2014). It is ubiquitous in daily-life where individuals often associate it with the 'little voice' in their head (Perrone-Bertolotti et al., 2014). It also subserves a range of cognitions (Baddeley, 1992; D'Esposito, 2007; Emerson & Miyake, 2003; Filik & Barber, 2011; Miyake et al., 2004; Yao & Scheepers, 2011; Yao et al., 2011) and could serve as a marker of developing psychopathy (Alderson-Day et al., 2018). Despite these significant roles, the neural substrates and cognitive mechanisms which support inner speech generation remain the subject of debate (Hurlburt et al., 2016).

One attempt at explaining the mechanisms driving inner speech can be found in the corollary discharge model. The corollary discharge model of inner speech posits that the motor movement-commands which generate overt speech can also be used to generate inner speech via a system of forward-models (Scott, 2013). These forward-models use a copy of the motor command to simulate its execution and predict its perceptual consequences. Within the broader literature of corollary discharge, these predictions of perceptual outcomes are used to distinguish whether a sensory input was caused by one's own movements, or whether it was externally generated (Crapse & Sommer, 2008). Within inner speech, it is argued that these predictions are internally perceptible and provide the sensory component of inner speech (Scott, 2013). Given that the model relies on the generation of motor signals in order to generate inner speech, it can be said to use a motor-route in order to generate inner speech.

The corollary discharge model of inner speech provides a clear explanation for the observed link between tasks which elicit inner speech (e.g. phonological judgement tasks) and increased increased activation of speech production regions (Marvel & Desmond, 2012; Shergill et al., 2001; 2002), such as the left inferior frontal gyrus (L-IFG). When viewing this link from the perspective of the corollary discharge model, activation of the L-IFG would represent the generation of motor movement-commands and their projection to forward-models in order to generate corollary discharge (Tourville & Guenther, 2011). This explanation provides a tidy account of some of the inner speech neuroimaging literature, but it lacks explanatory power in other areas. For example, an individual's inner speech is not limited to their own voice and often incorporates the voices of other individuals (McCarthy-Jones & Fernyhough, 2011). These voices are likely to contain distinct vocal features (e.g. pitch, timbre) which are distinct from one's own and therefore challenging to replicate using a physiologically-constrained system based on one's own articulators (Giovanni & Duflo, 2010). Although relatively under-investigated, some initial evidence does indicate distinct neural activation during the generation of less egocentric varieties of inner speech (Brück et al., 2014). This overlaps with recent neuroimaging studies which find that certain types of inner speech are correlated with the activation of speech perception regions, rather than speech production regions (Barsalou, 2008; Hurlburt et al., 2016; Yao & Scheepers, 2011).

The involvement of speech perceptual regions, rather than speech production regions, in some types of inner speech suggests the existence of an alternative mechanism of generation. One model which could describe how inner speech (Yao et al., 2011), and imagery more broadly (Dijkstra et al., 2020; Zatorre & Halpern, 2005), can emerge from perceptual regions is the perceptual simulation model. The perceptual simulation model argues that inner speech, and imagery more broadly, can be generated via a mechanism in which the neuronal firing patterns which activate during the perception of a sensory stimulus can be stored and later reactivated. By reactivating the encoded firing patterns, it is possible to recreate a perceptual experience in the absence of a matching external stimulus. Importantly, encoded firing patterns can be mixed and extrapolated upon, allowing for a finite number of encoded patterns to generate a large number of perceptual experiences (Barsalou, 1999; Damasio, 1989).

Interpreting inner speech as the outcome of perceptual simulation addresses a number of the issues discussed in relation to the corollary-discharge model. Inner speech can incorporate a range of acoustically-distinct voices as it relies on an extrapolatable inventory of stored speech memories, rather than a system dependent on one's articulators. Given that perceptual simulation depends on speech perceptual regions, rather than production regions, it would also provide a plausible explanation as to why some neuroimaging research has observed spontaneous inner speech in the absence of activation in speech production regions (Yao et al., 2011; Hurlburt et al. 2016).

Although the perceptual simulation model provides a more flexible approach which is better able to account for the diverse nature of inner speech phenomenology, it cannot explain the full range of neuroimaging and neuropsychological findings. For example, while the perceptual simulation model helps explain the findings of studies investigating more diverse forms of inner speech (Yao et al., 2011; Hurlburt et al. 2016), there remains a large body of literature which does implicate speech production regions - with some studies observing activation of speech production, but not perception regions (De Nil et al., 2000; Gulyás, 2001). This suggests a lack of perceptual simulation in at least some cases of inner speech. It is also notable that the use of transcranial magnetic stimulation (TMS) to induce a ‘virtual lesion’ over speech production regions has been found to inhibit inner speech (Aziz-Zadeh et al., 2005). This suggests a causal role of speech production regions in some cases of inner speech.

A Dual-Mechanistic Framework of Inner Speech

Recently, Pratts et al. (2023) proposed a dual-mechanistic framework of inner speech which aimed to reconcile the disparate theoretical models and their associated evidence. The dual-mechanistic framework proposes that inner speech can be generated via a motor-route based on corollary discharge and a perceptual-route based on perceptual simulation, with two phenomenological dimensions determining which mechanism is utilised. The dimension of *egocentricity* measures whether the inner speech is in one’s own voice (high egocentricity) or whether it attempts to recreate the voice of another individual (low egocentricity). The dimension of *intentionality* measures whether the inner speech emerged as the result of explicit task demands / instructions (high intentionality) or whether it emerged spontaneously (low intentionality). As inner speech becomes increasingly egocentric and intentional, it places higher demands on the motor-route of generation. As inner speech becomes less egocentric or intentional, it places greater demands on the perceptual-route of activation.

The dual-mechanistic framework provides a more parsimonious explanation for the observed link between highly intentional and egocentric inner speech studies and activation of speech production regions, and the link between less intentional and egocentric inner speech studies with speech perceptual region activation. Moreover, by segmenting inner speech into gradations based on phenomenology, the framework acknowledges and aims to explain the wide range of inner speech experiences which are encountered yet often overlooked. Pratts et al. (2023) provide initial evidence for the framework in an activation-likelihood estimation (ALE) meta-analysis which observed differential activation of speech production and speech perceptual regions, depending on phenomenology. One cluster was observed in the L-IFG, the region associated with the motor-route of generation. An additional cluster was also observed in the left superior temporal gyrus (L-STG), a speech perceptual region implicated in the perceptual-route of activation. Given these observations, as well as theoretical rationale, it was argued that the L-IFG and L-STG support the motor-route of activation and perceptual-route of activation, respectively.

Aim & Hypotheses

Whilst the ALE analyses provide useful initial evidence, it is an observational approach which contended with biases in the neuroimaging literature towards highly intentional and egocentric paradigms. Therefore, we investigated the dual-mechanistic model using transcranial magnetic stimulation (TMS). TMS uses electromagnetic induction to induce a small current within the targeted brain region. This current is sufficient to depolarize adjacent neurons and cause temporary reductions in cortical excitability, reducing performance in tasks which are dependent on those regions (Bohning et al., 1997; Klomjai et al., 2015). TMS therefore provides a non-invasive means of mapping cognitive processes to brain regions. By examining the effect of TMS suppression over the L-IFG and L-STG on performance in tasks which vary across the egocentricity x intentionality dimension, we tested the hypotheses laid out in Pratts et al. (2023).

If the dual-mechanistic framework is correct, inhibiting the motor-route via TMS stimulation of the L-IFG should suppress inner speech subtypes which are high in intentionality and egocentricity. Inhibiting the perceptual-route via TMS stimulation of the L-STG should suppress inner speech subtypes which are low in intentionality and egocentricity. Inner speech subtypes which place demands on both the motor-route and perceptual route (e.g. high in intentionality and low in egocentricity) should be suppressed by both L-IFG and L-STG suppression.

Methods

Participants

Based on statistical power analyses conducted using G*Power (Faul, 2007), we estimated that 18 participants would be sufficient to detect a large effect ($f = 0.40$) within a repeated measures ANOVA. Thirty right-handed participants were initially recruited to take part in the experiment, of which ten completed all three sessions (8 female, 2 male; mean age = 20.1 years, SD = 1.97). Participant dropout was predominantly due to safety precautions during the COVID-19 pandemic, with one participant withdrawing due to discomfort during TMS stimulation. All participants were native English speakers with normal or correct-to-normal vision. Participants were free from any history of neurological disease and other TMS contraindications outlined in international safety guidelines (Rossi et al., 2009). All participants gave informed consent and were recompensed for their participation using course credits. Ethical approval for the study was granted by the local ethics board (Ethics Committee Approval Code: 2019-5370-10112).

Design

The study employed a repeated-measures design, with stimulation site (L-STG vs. L-IFG vs. occipital pole / Oz), inner speech task (scenario rating task vs. spontaneous imagery task) and TMS (no rTMS vs. rTMS stimulation) as the 3 within-participant factors. rTMS was applied using an offline approach in which participants underwent stimulation without a concurrent inner speech task. Inner speech performance under the rTMS refractory period was compared to performance in the same tasks prior rTMS stimulation.

Stimuli

Participants completed a scenario rating task which tested high intentionality subtypes of inner speech and a spontaneous imagery task which tested low intentionality subtypes of inner speech. Each task contained high egocentricity and low egocentricity conditions, allowing each quadrant across the two phenomenological dimensions to be tested. The exact pairing of task conditions to hypotheses are as follows:

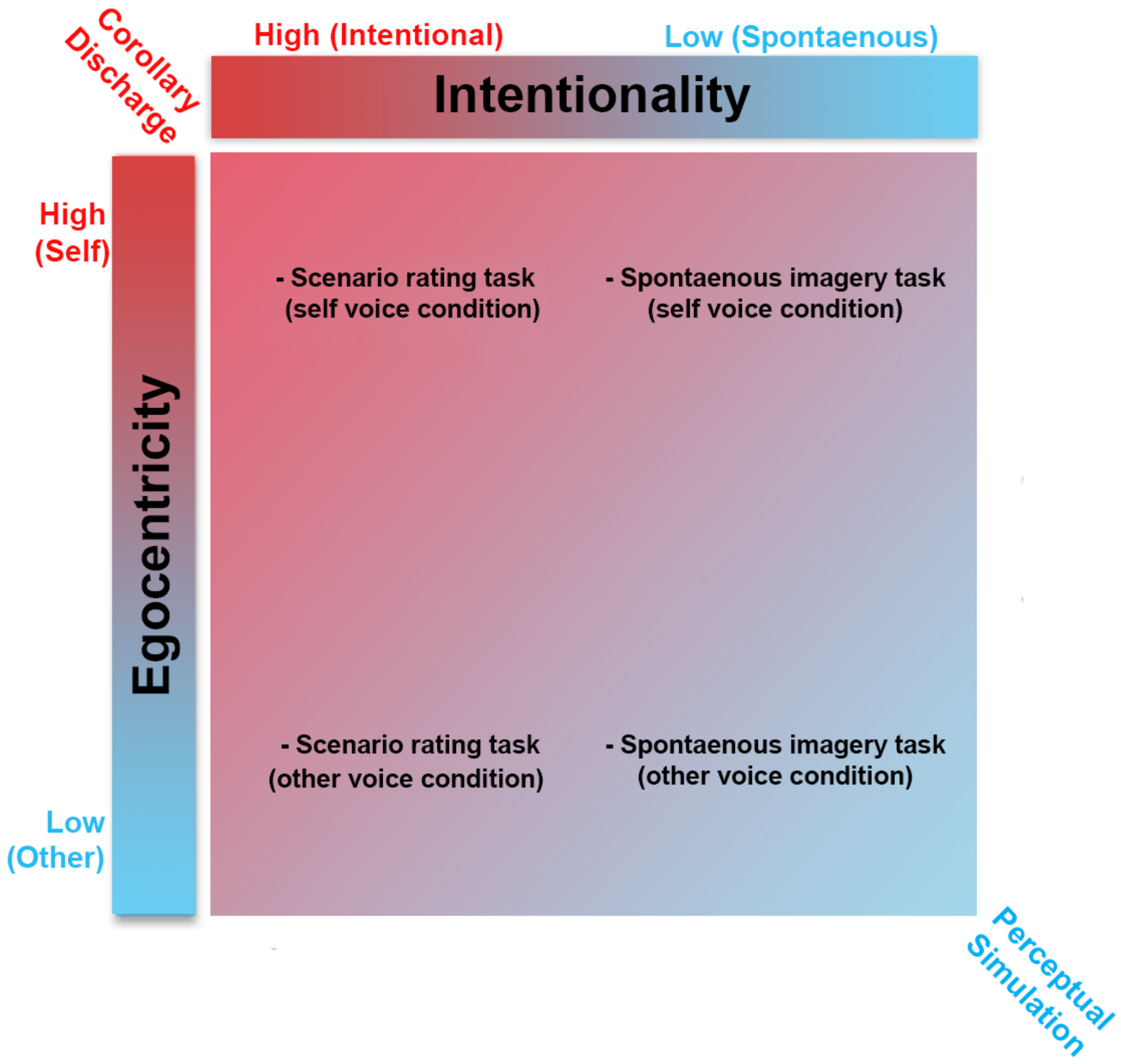
Scenario Rating Task (High Intentionality)

- Own Voice Condition (High Egocentricity - Suppressed by L-IFG stimulation)
- Other Voice Condition (Low Egocentricity - Suppressed by both L-IFG & L-STG stimulation)
- Visual Imagery Condition (Control - No significant suppression compared to control site)

Spontaneous Imagery Task (Low Intentionality)

- Own Voice Condition (High Egocentricity - Suppressed by both L-IFG & L-STG stimulation)
- Other Voice Condition (Low Egocentricity - Suppressed by L-STG stimulation)
- Visual Imagery Condition (Control - No significant attenuation compared to control stimulation sites)

Figure 1. Placement of inner speech tasks and conditions used in the battery of tasks within the dual-mechanistic framework outlined in Pratts et al. (2023).



Note: L-IFG stimulation is hypothesised to have a larger impact in conditions positioned in the red shaded areas as these place increasing demands on the motor-route mechanism of generation. L-STG stimulation is predicted to impact inner speech conditions positioned in the blue shaded areas.

Scenario Rating Task

The scenario rating task tested high intentionality varieties of inner speech and contained three conditions: *own*, *other* and *visual*. All conditions followed a similar format, with participants being asked to imagine a scenario and rate the vividness of the resulting imagery:

Own: The *own* condition elicited inner speech high in intentionality and high in egocentricity. Participants were given a scenario in the format “*Using your own inner voice, say…*”, followed by a sentence in prose. Participants would respond using by rating the vividness of their elicited inner speech on a 5-point scale.

Other: The *other* condition elicited inner speech high in intentionality and low in egocentricity. Participants were instructed ““*Imagine another person saying…*”, followed by a sentence in prose. Participants would then generate the inner speech in the voice of someone familiar to them and rate the vividness of the elicited speech using a 5-point scale. Participants were instructed to choose a single individual whose voice they were familiar with at the beginning of the experiment.

Visual: The *visual* condition elicited visual imagery and served as a control condition. Participants were asked ““*Imagine the appearance of…*” followed by the description of a visual scene (e.g. clouds in the sky). Participants then rated the vividness of the elicited visual imagery using a 5-point scale.

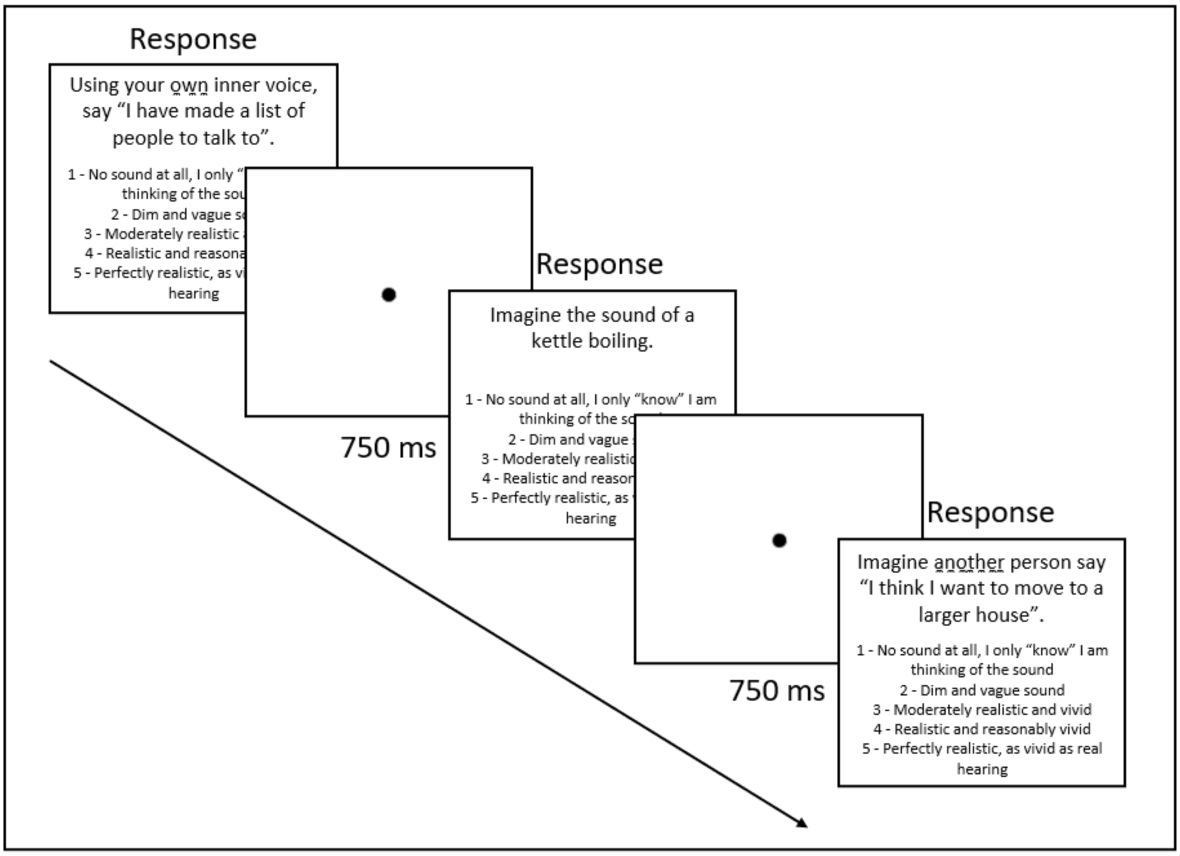
Own scenarios represent a subtype of inner speech which heavily utilises the motor-route of inner speech generation, being both highly intentional and highly egocentric. *Other* scenarios are intended to generate inner speech which places greater demands on the perceptual-route relative to *own*, given that it is low in egocentricity. The *visual* condition should not place heavy demands on either the motor-route or the same perceptual-route which subserves inner speech.

All conditions used the same rating scale with responses entered on a keyboard using numbers 1 - 5 (as illustrated in *Figure 4)*. Participants were presented with 8 scenarios per condition, for a total of 24 scenarios. Conditions and scenarios were interspersed in a random order with a 750ms interstimulus interval (ISI) separating each trial. Scenarios were not time limited in order to prevent floor effects. Six stimuli sets were rotated across participants in a counterbalanced manner.

Figure 2. 5-point scale used in the scenario rating task.

5	Perfectly realistic, as vivid as real hearing
4	Realistic and reasonably vivid
3	Moderately realistic and vivid
2	Dim and vague sound
1	No sound at all, I only “know” I am thinking of the sound

Figure 3. Trial procedure in scenario rating task.



Using your own inner voice, say “I have made a list of people to talk to”.

Imagine the sound of a kettle boiling.

Imagine another person say “I think I want to move to a larger house”.

1 - No sound at all, I only “know” I am thinking of the sound
2 - Dim and vague sound
3 - Moderately realistic and vivid
4 - Realistic and reasonably vivid
5 - Perfectly realistic, as vivid as real hearing

Note: ISI is presented for 750ms, participants then presented with an own, other or visual scenario. Participants respond using keyboard numbers 1 to 5.

Spontaneous Imagery Task

The spontaneous imagery task tested low intentionality varieties of inner speech and was composed of three conditions: *own*, *other* and *visual control*.

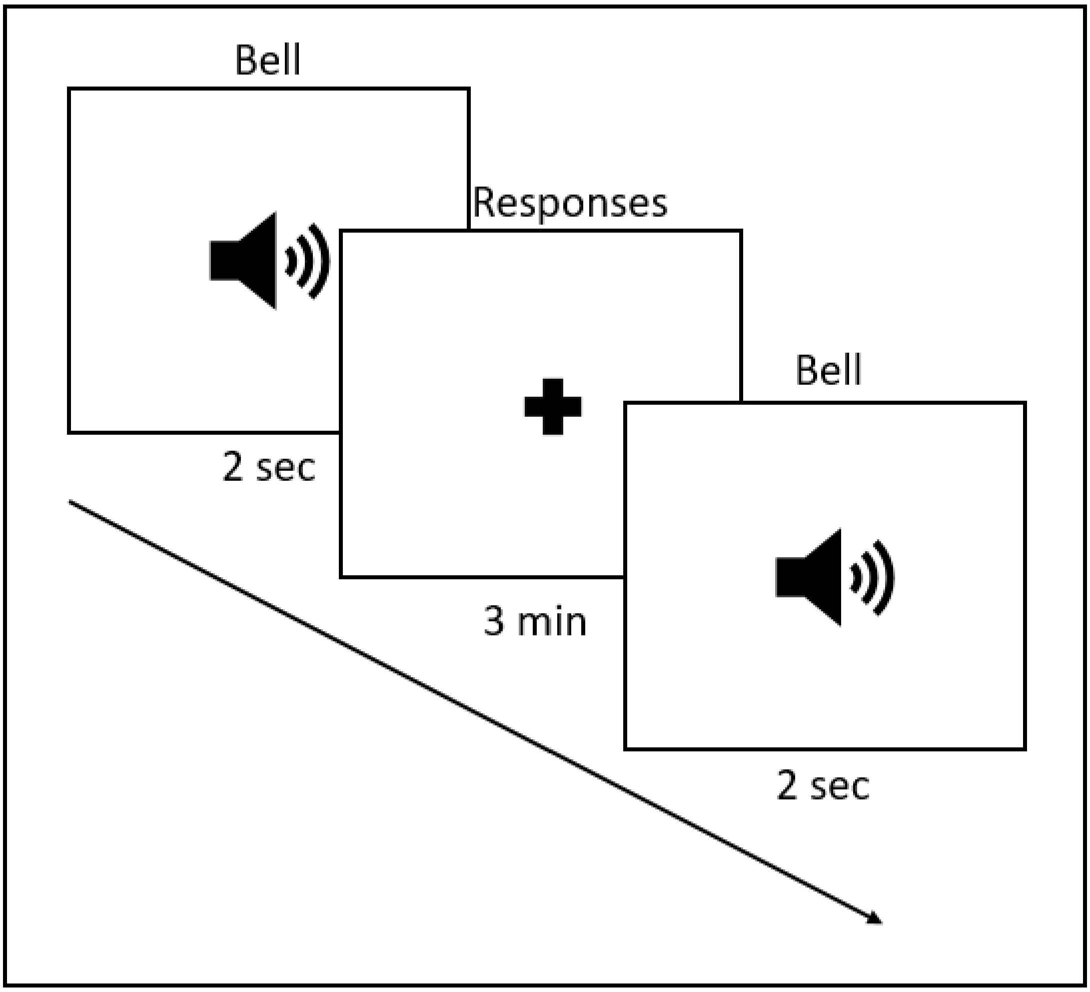
The *own* condition was designed to elicit inner speech low in intentionality and high in egocentricity. In this condition participants were presented with a fixation cross for three minutes. During the three minutes participants were instructed to close their eyes, relax and press the keyboard response key each time they heard their inner voice. An audible bell was used to indicate the beginning and end of the three minute period.

The *other* condition elicited inner seech low in intentionality and low in egocentricity. In the *other* condition, participants were first presented with a 30-second audio clip of a movie quotation being repeated. The audio clip was extracted from the television show Doctor Who and contained a Dalek repeating the word “Exterminate!”. Participants were then presented with a fixation cross for three minutes. During this time, participants were asked to close their eyes, relax, and to not think of the Dalek saying “Exterminate!”. Participants were asked to press the keyboard response key each time they internally ‘heard’ the Dalek saying “Exterminate!”. An audible bell was used to indicate the beginning and end of the three minute period.

The *visual control* condition was designed to elicit visual imagery, which would serve as a non-auditory control. Participants were presented with a fixation cross for three minutes. During the three minutes participants were instructed to close their eyes, relax, and to not think of a white bear (Wegner, 1994). Participants were asked to press the keyboard response key if a white bear came into the mind's eye. An audible bell was used to indicate the beginning and end of the three minute period.

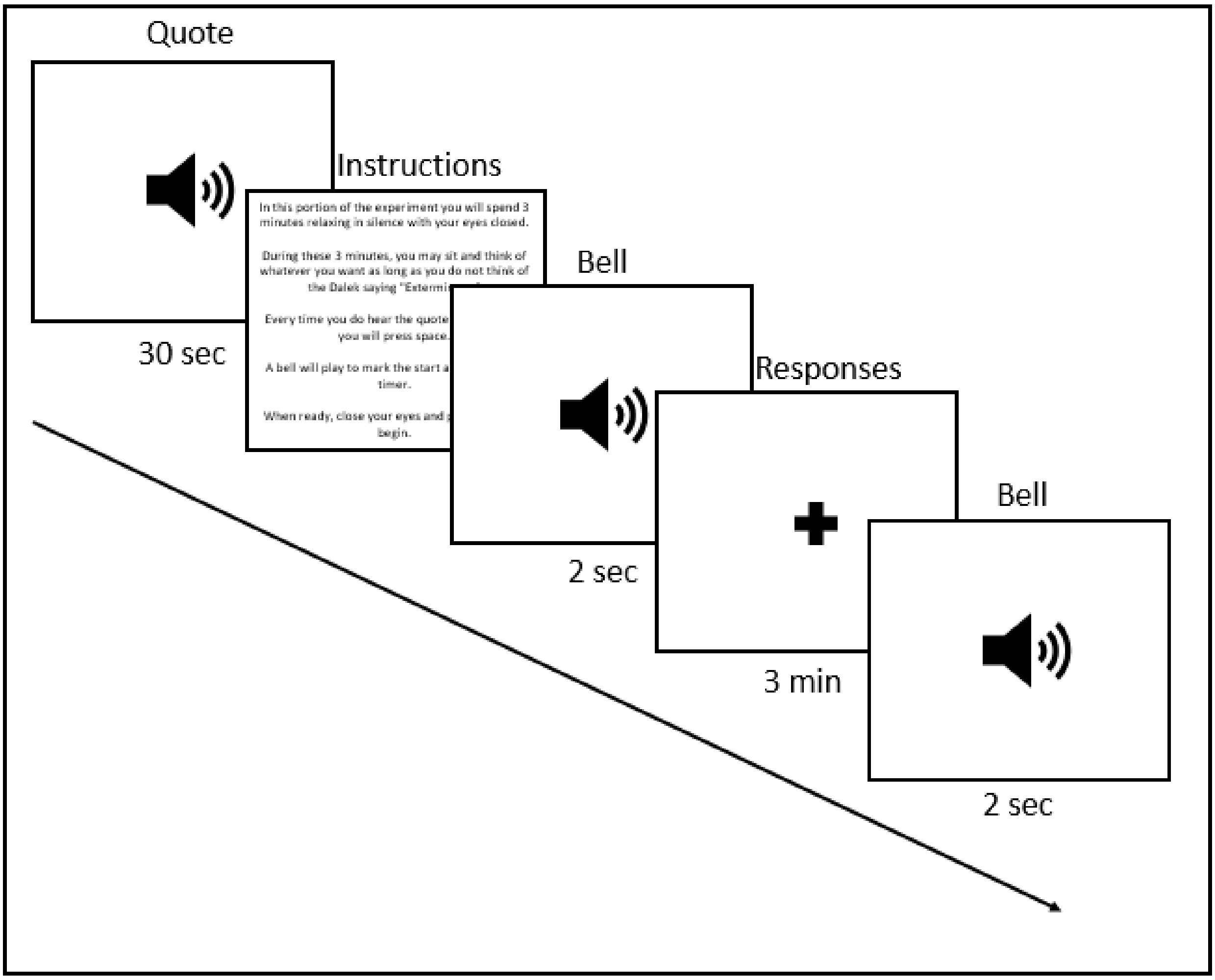
Conditions were completed in a counterbalanced order, with all conditions were preceded by short practice trials. In order to prevent participant fatigue, practice trials lasted 30-seconds as opposed to three minutes. The *other* condition and *visual control* condition used specific stimuli allocated for practice blocks (i.e. novel movie quote & visual imagery target).

Figure 4. Trial procedure in spontaneous imagery task - *own* and *visual control* conditions.



Note: Participants were asked to close their eyes for three minutes, with the end of the period marked by an audible bell. During this period participants would press the response key each time they either hear their inner voice (own condition) or thought of a white bear (visual control condition).

Figure 5. Trial procedure in spontaneous imagery task - other condition.



Note: Participants first heard a movie quote repeated for a period of thirty seconds. After briefly being reminded of the instructions, participants underwent a three minute period of silence in which they were asked to press the response key if they "heard" the movie quote internally.

TMS Procedure

TMS Hardware

TMS stimulation was delivered using a MagStim Rapid 2 stimulator (Magstim Co., Whitland, UK) with two external boosters (2.2 T maximum output). The coil was a 70-mm figure-of-eight coil (D70 Alpha Coil; Magstim Co., Whitland, UK) with an affixed neuronavigation tracker.

Anatomical MRI Image Generation

3D T1-weighted anatomical MRI images were generated for all participants by spatially transforming an MRI template image to the individual head shape of each participant. We followed the procedure outlined in Fleischmann et al. (2020), which was found to generate images which were topographically accurate and functionally suitable for neuronavigated TMS stimulation. Participant head shapes were recorded using an infrared-based Polaris Vicra 3D tracking camera (Northern Digital, Waterloo, Canada) in combination with Brainsight's 3D-tracked pointer and subject tracker (Rogue Research, Montreal, Canada). These were used to mark the three-dimensional locations of three fiducial points (nasion, left and right preauricular) and 33 points around their scalp. The 33 points were evenly distributed across the scalp following the pattern used in the 10-20 system of electrode placement. A high-resolution T1-weighted anatomical template (Lalys et al., 2010) was then warped to the participants registered head coordinates in Matlab (version 2016a, Mathworks) using the Statistical Parametric Mapping toolbox and the Fieldtrip toolbox (read Fleischmann et al., 2020 for a detailed account of this process).

Selection of TMS Site

Coordinates for the left-STG and left-IFG were drawn from Pratts et al. (2023) neuroimaging meta-analysis. The Talairach coordinates for the left-STG were [-57, -45, 11]. Taliarach coordinates for the left-IFG were [-43, 20, 2]. The coordinates for the occipital pole (Oz) were [2, -92, 10] in Talairach space. Additionally, the motor hand area was visually identified within the MRI images and marked for motor thresholding purposes. This was identified by locating the characteristic omega sign on the central sulcus. A Polaris Vicra 3D tracking camera was used in conjunction with the Brainsight system in order to carry out real-time neuronavigation during stimulation. The MRI image was registered to the participants head using six anatomical landmarks (nasion, tip of nose, left/right tragus, left/right eye outer canti) and nine point surface registration (Fleischmann et al., 2020).

Stimulation Parameters

Stimulator output was determined on a per-participant basis using their motor threshold (MT). Motor threshold was defined as the minimum stimulator output required to generate a reliable visible twitch (≥ 3/6) in the relaxed abductor pollicis brevis muscle of the contralateral hand. The rTMS was applied in the form of continuous theta burst stimulation which consists of three high-frequency (50 Hz) TMS pulses delivered at a 5Hz rhythm for a total of 600 pulses over 40 seconds. Continuous theta burst stimulation is an inhibitory protocol with a refractory period characterised by reduced cortical excitability and BOLD responses lasting up to 50 minutes after stimulation (Wischnewski and Schutter, 2015). The coil was held firmly against their scalp, centred over the site to be stimulation. When stimulating the L-IFG and L-STG, the coil was oriented at 45° relative to the interhemispheric fissure. During stimulation of the occipital pole, the coil was held with the handle facing upwards. Stimulation intensity was set to 80% of each participant's resting motor threshold. The average MT was 59% of the maximal stimulator output and the average stimulation intensity during rTMS was 47.2%. Due to the proximity of the L-STG and L-IFG to the ear, all participants were given earplugs to attenuate the sound produced by the discharging coil.

Procedure

All behavioural tasks were completed using a PC running Opensesame experimental software (Mathôt et al., 2012). Peripherals consisted of a 24” LCD monitor (60Hz; 1920 x 1080 resolution) and Sony WH-1000M3 headphones.

Participants completed three experimental sessions, each applying rTMS to one of three target sites (L-STG, L-IFG, Occipital pole) and taking place at least 72 hours from the previous session. Prior to behavioural testing, individualised MRI templates were generated (first experimental session only) and resting motor thresholds were determined (each experimental session). The main experiment then commenced in the following order: pre-rTMS behavioural testing, offline rTMS (L-STG, L-IFG or Oz), post-rTMS behavioural testing. Post-rTMS behaviour testing was initiated after a 5 minute rest period which allowed the cortical suppression induced by rTMS to take effect. Pre-rTMS and post-rTMS behavioural testing were identical, each taking approximately 12 minutes to complete and consisting of the scenario rating task and spontaneous imagery task. The order in which tasks and their conditions were completed was counterbalanced across participants.

Data Analysis

Data analysis was completed using R (version 4.1.2) and the RStudio IDE (Build 443). In the scenario rating task the data was trimmed by grouping the data by condition and stimulation site, and then excluding any trials whose dependent variable was more than 2.5 times the median absolute deviation (MAD). This procedure is outlined in Leys et al. (2013) and has several benefits over competing approaches, including a standardised procedure. Data loss did not exceed 5% of the total number of trials. TMS effect scores were then calculated by subtracting pre-TMS behavioural scores from post-TMS behavioural scores. This yields a dependent variable in which a negative score represents decreased performance following rTMS, a score of zero represents no change following rTMS, and a positive score represents increased performance following rTMS (Esterman et al. 2006). In the spontaneous imagery task, trials in which participants provided ≥ 180 responses were deemed outliers, this was equivalent to participants reporting an instance of spontaneous thought every second for the whole three minute trial - no trials reached this threshold.

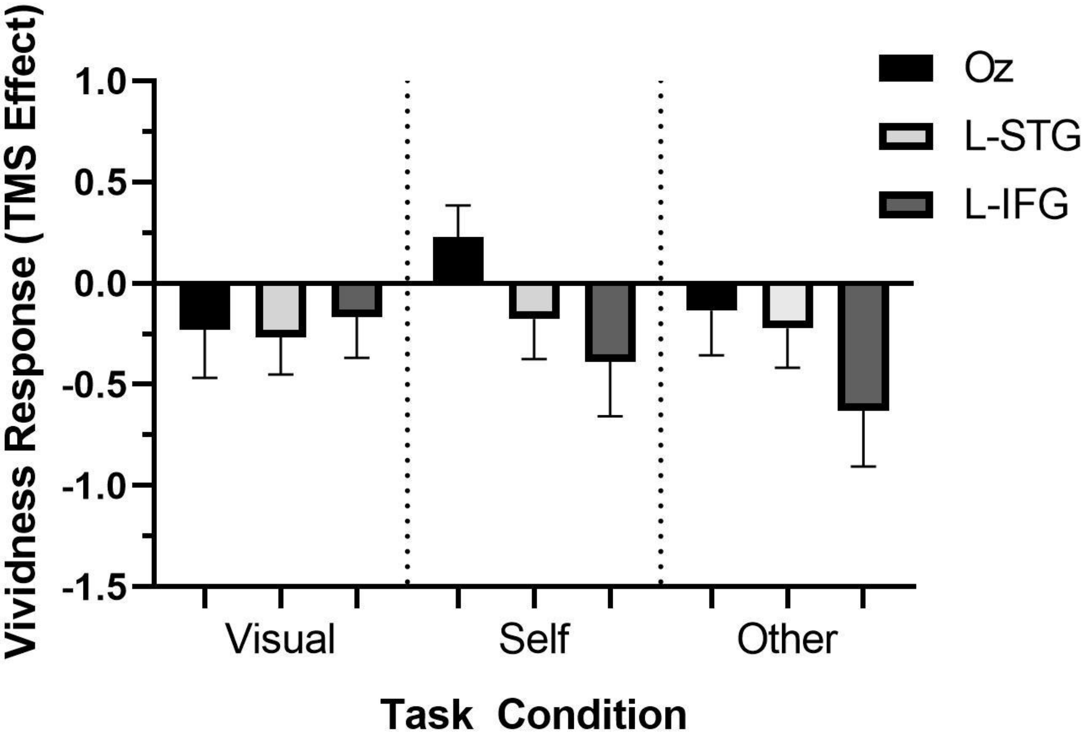
To assess the hypothesis that the neural underpinnings of inner speech vary by phenomenology, we conducted a two-way ANOVA for the scenario rating task and a two-way ANOVA for the spontaneous imagery task, respectively. The ANOVAs contained TMS Site (Oz, L-IFG, L-STG) and Task (Own, Other, Visual) as the main factors. In the scenario rating task, the two-way ANOVA compared the vividness scores (TMS effect; post-TMS scores minus pre-TMS scores) across the main factors. In the spontaneous imagery task, the two-way ANOVA compared the number of spontaneous thoughts per three minutes (TMS effect; post-TMS scores minus pre-TMS scores) across the main factors. Post-hoc comparisons were conducted, when appropriate, using the Bonferroni method for controlling the multiple comparisons.

Results

Scenario Rating Task

The results of the ANOVA showed no significant main effect of Task (F(2, 81) = 0.735, p = .483), indicating that there was no significant difference in the outcome measure across the three tasks. Similarly, there was no significant main effect of TMS site (F(2, 81) = 1.912, p = .154), suggesting that the outcome measure did not significantly differ between the three stimulation sites. Moreover, the interaction effect between task and site was not significant (F(4, 81) = 0.816, p = .518), indicating that the relationship between task and the outcome measure did not differ significantly across the stimulation sites.

Figure 6. Effect of TMS stimulation across conditions in the high intentionality inner speech subtypes.

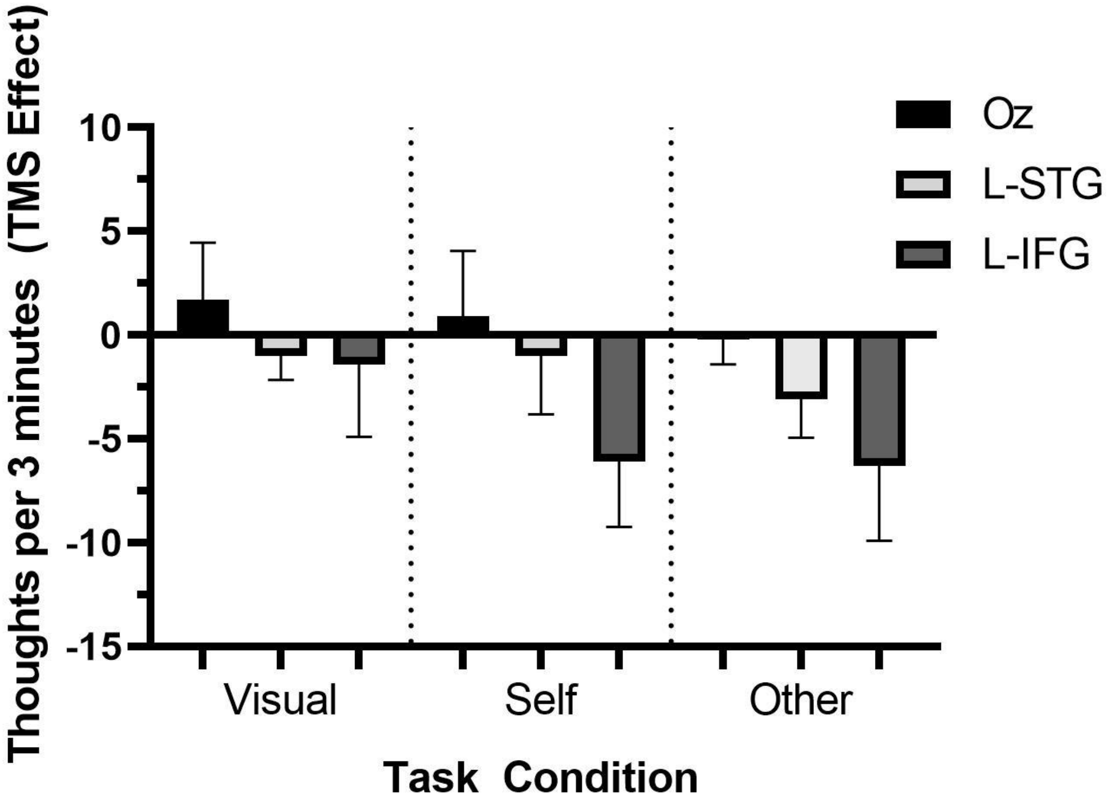


Note: Mean vividness responses (TMS effect) in the scenario rating task across task conditions and TMS sites. Error bars represent standard error of the mean.

Spontaneous Imagery Task

ANOVA analysis of the spontaneous imagery task revealed no significant main effect of Task, F(2, 81) = 0.902, p = .4099, suggesting that the differences in means across the three task conditions were not statistically significant. There was no significant interaction between Task and TMS site, F(4, 81) = 0.224, p = .9240, indicating that the effect of site did not differ significantly between the task conditions. However, the main effect of TMS site did approach statistical significance , F(2, 81) = 2.939, p = .0586, suggesting that the differences in means across the three site conditions might be meaningful, although not reaching the conventional level of significance (p < .05).

Figure 7. Effect of TMS stimulation across conditions in the low intentionality inner speech subtypes.



Note: Mean spontaneous thought occurrences per 3 minute trial across task conditions and TMS sites. Error bars represent standard error of the mean.

Discussion

The present study used rTMS to investigate the roles of the L-IFG and L-STG in a variety of inner speech subtypes. High intentionality varieties of inner speech were examined using a scenario rating task which contained elicited inner speech of varying egocentricity. We predicted that inner speech in one's own voice (high egocentricity) would be primarily suppressed by stimulation of the L-IFG, with inner speech in the voice of another person (low egocentricity) condition being suppressed by stimulation of both the L-IFG and L-STG. ANOVA analysis did not reveal any significant main effects of either TMS stimulation on the vividness of inner speech. However, a visual trend towards L-IFG stimulation suppressing scores in both the *own* and *other* conditions was visible and could justify further research.

Low intentionality varieties of inner speech were examined using a spontaneous inner speech task. Here, we predicted that spontaneous inner speech in one's own voice (high egocentricity) and inner speech in the voice of others (low egocentricity) would place demands on the perceptual-route of activation - the route principally associated with the L-STG. While a main effect of TMS site did approach significance, the trend was towards L-IFG stimulation suppressing *own* and *other* conditions, rather than an effect of L-STG stimulation.

Implications for The Dual-Route Model

The current results are contrary to the hypothesis that the neural mechanisms which underpin inner speech vary depending on its phenomenological attributes. While the results of the statistical analyses did not reach significance, likely due to substantial reductions in statistical power following participant dropout, a trend towards L-IFG suppression is visible across all experimental conditions with the exception of control conditions, suggesting a more ubiquitous role for the L-IFG than we previously expected. One interpretation of these findings is that inner speech is primarily the outcome of motor-route activation irrespective of its diverse phenomenological attributes. This aligns with more traditional explanations of inner speech generation (Hoffman & Kravitz, 1987; Scott, 2013) as well as more recent work by Grandchamp et al. (2019).

On the one hand, a purely motor-route interpretation of inner speech neatly accounts for results of both the spontaneous imagery task and the scenario rating task. On the other hand, while we observed a trend towards stimulation of the L-IFG resulting in decreases in inner speech occurrences (spontaneous imagery task) and vividness (scenario rating task), the exact role of the L-IFG cannot be discerned in this study. According to previous work on both inner speech (Tian & Poeppel, 2013) and overt speech (Tourville & Guenther, 2011), the L-IFG serves a role in the generation of speech motor signals and efference copies. This would suggest that the suppressive effect of L-IFG stimulation on inner speech stems from interference with the generation of the motor signals and efference copies used in forward-models. However, activation of the L-IFG is ubiquitous across imagery of different sensory modalities, being observed in studies investigating olfactory imagery (Ripp et al., 2018), somatic imagery (Olivetti Belardinelli et al., 2009; Tomasino et al., 2022) and tactile imagery (Schmidt & Blankenburg, 2019). This suggests a more general, modality-independent role for the L-IFG rather than one limited to the generation of motor signals. Viewing the L-IFG as having a more general role in imagery generation also explains a trend towards stimulation of the L-IFG suppressing imagery of non-vocal sounds. Discussing similar findings in which subthreshold activation of supplementary motor areas was observed during the imagery of musical instrument timbre, Halpern et al. (2004) similarly suggests that the L-IFG could reflect either more general imagery processes or participants attempts at coarse recreations of the musical instrument sounds (e.g. subvocalizing the relevant pitch).

The broad literature implicating the L-IFG in imagery of various sensory domains justifies a degree of caution when interpreting the results via the lens of the corollary-discharge / motor-route interpretation. However, the finding that TMS stimulation did not attenuate performance in the visual control conditions of either task could represent initial evidence contrary to a modality-general interpretation of the L-IFG's role, with further research justified given the limited scope of the two control conditions.

Methodological Considerations

While these questions remain, it is notable that stimulation of the L-STG did not yield clear suppressive effects in any of the conditions, in either task. One interpretation of this finding is that this further evidences the reliance of inner speech on motor, rather than perceptual, processes. However, as the L-STG/STS are also hypothesised to represent the regions in which corollary discharge is audibly perceived (Tian et al., 2016), some effect of L-STG stimulation could be expected even if inner speech was produced via the motor-route. The lack of effects following L-STG stimulation could therefore represent methodological limitations in the study. One limitation encountered was the substantial amount of participant withdrawals which occurred across the three sessions. A further factor which should be considered is the high degree of focality inherent to figure-of-eight TMS coils (Thielscher & Kammer, 2004), with simulations predicting that the volume of the brain area exposed to the majority of the magnetic pulse is as low as 5 cm^2 (Deng et al., 2013). This focality does not represent a challenge when targeting a well-defined neuroanatomical area such as the L-IFG, but does lead to potential limitations given that the L-STG target was the maxima of a large fMRI meta-analysis cluster extending across a volume of 14 cm^2 (Pratts et al., 2022). It is therefore possible that the stimulation site did not sufficiently overlap with the specific subregion which underpins perceptual simulation, resulting in suboptimal stimulation. It is also notable that while we used the results of an fMRI meta-analysis to inform our target location, alternative candidate sites do exist and could prove to be effective stimulation targets, with both Heschl's Gyrus and the Left Superior Temporal Sulcus having been identified in previous research on spontaneous forms of inner speech (Hurlburt et al., 2016; Yao et al., 2011).

Conclusion

The study did not observe a statistically significant effect of TMS stimulation on any variety of inner speech. However, some initial evidence was observed indicating that the left inferior frontal gyrus might play a causal role in the generation of inner speech. This role does not seem to be limited to specific subtypes of inner speech, but rather a more general role in mediating inner speech of varying phenomenology. Our study therefore justifies further research into the role of the L-IFG in generating inner speech. A role for the left superior temporal gyrus in inner speech was not established with several explanations being put forward.

Chapter 6 - General Discussion

This final discussion Chapter is divided into three sections. First, we will summarise the key findings from each of the four empirical Chapters. The second section will discuss whether these findings are best explained by our proposed framework, or more traditional models of inner speech. In the final section, we will consider possible avenues for further research.

An overarching aim of this thesis was to further our understanding of the neurocognitive mechanisms underpinning inner speech by reconciling conflicting models and evidence. We proposed that the conflicting evidence within the research literature was driven by the tendency to interpret inner speech as a monolithic and invariant experience, rather than a dynamic behaviour with a diverse phenomenology. Our view was driven by recent research which suggests that the neural correlates of inner speech vary in line with phenomenology (Hurlburt et al., 2016; Yao et al., 2011). We created and tested a dual-mechanistic framework which aimed to bridge the diverse phenomenology of inner speech and its underlying neurocognitive mechanisms. This framework consisted of two phenomenological dimensions by which inner speech could vary: intentionality and egocentricity. Intentionality represented the extent to which inner speech emerged spontaneously, or as the result of task-demands. Egocentricity represented the extent to which inner speech reflected the voice of the speaker, or the voice of another person. These phenomenological dimensions determined whether inner speech was recruited via a motor-route (i.e. based on corollary discharge) or a perceptual-route - two mechanisms which have previously been argued to facilitate inner speech production. As inner speech increases in intentionality and egocentricity, it places greater demands on the motor-route of inner speech generation. As it decreases in intentionality and egocentricity, it is argued to place greater demands on the perceptual-route of inner speech generation.

A cross-cutting theme throughout the thesis is the belief that our understanding of inner speech should be informed by converging evidence from studies utilising distinct and complementary methodologies. This theme reflects the fact that inner speech is a challenging phenomenon to investigate and that each method of observing or manipulating inner speech exhibits weaknesses which are best overcome via the use of a complimentary methodology (e.g. fMRI and TMS allow for both observational and causal inferences). This approach underpinned the majority of the thesis, with the dual-mechanistic framework being examined using four distinct approaches across the four empirical chapters. As defined in the abstract, the thesis aims were as follows:

1. To develop an integrated model of inner speech which explains heterogeneous findings and accounts for the diverse phenomenology of inner speech.
2. To determine the validity of this model by testing its predictive validity using a range of experimental methodologies.

Summary of Findings

In Chapter 2, our first empirical Chapter, we carried out a review of the inner speech neuroimaging literature and proposed a predictive framework which argued that inner speech could be generated via two mechanisms: one centred on speech production regions (motor-route) and another centred on speech perceptual regions (perceptual-route). Whether inner speech would be generated via one mechanism or another was determined by its phenomenology, with two phenomenological dimensions being proposed: intentionality and egocentricity. In order to examine this framework, we performed a series of ALE fMRI meta-analyses on 22 eligible studies. As hypothesised, the results of the meta-analyses indicated that distinct neural mechanisms are differentially engaged as inner speech varies across the phenomenological dimensions of intentionality and egocentricity. We observed speech production brain regions associated with the motor-route of generation to be engaged when the inner speech was highly intentional and egocentric, but not when inner speech was low in intentionality and egocentricity. In studies eliciting low intentionality inner speech, we observed convergence over the speech perceptual regions implicated in the perceptual-route of generation. The results provided initial evidence that varieties of inner speech are supported by more than one neural mechanism, and that a framework bridging phenomenology and neurocognitive mechanisms could present a useful tool in future research.

In Chapter 3, we carried out further analysis of the framework by examining whether aphantasia, which could represent dysfunction of perceptual simulation mechanisms, affects the subtypes of inner speech which our framework predicts are perceptually-generated. Given the lack of previous research investigating the relationship between aphantasia and inner speech, the Chapter first established that aphantasia is not limited to decrements in the ability to generate visual imagery, but is correlated with broader deficits in other sensory modalities. Individuals with aphantasia and a control group then completed a battery of inner speech tasks which elicited inner speech of varying intentionality and egocentricity. With aphantasia predicted to have an increasingly detrimental impact on inner speech as it decreased in intentionality and egocentricity, as this is argued to place greater demands on the perceptual-route of inner speech generation. Examination of inner speech which was highly intentional found that aphantasia resulted in a widespread reduction in vividness, regardless of the egocentricity of the inner speech or its predicted reliance on the unaffected motor-route. Examination of low intentionality inner speech found both high and low egocentricity subtypes to be unaffected by aphantasia. Interestingly, these were both predicted to place demands on the affected perceptual-route of generation. In observing a varying impact of aphantasia across different subtypes of inner speech, the experiment supported the idea of inner speech being a varying, rather than invariant, phenomena. However, the precise predictions outlined in the dual-mechanistic framework did not align with the observed data.

In Chapter 4 we examined the validity of the dual-mechanistic framework by using a dual-task interfere paradigm to suppress the motor-route and perceptual-route, respectively. The study provided evidence that the effectiveness of motor-route interference and perceptual-route interference was dependent on the phenomenology of the inner speech. In highly intentional inner speech, motor-route interference was found to be most effective at suppressing inner speech which was highly egocentric (i.e. in one's own voice). However, in the low egocentricity condition where participants were asked to generate inner speech in the voice of another individual, perceptual-route interference was found to be as effective as motor-route interference. Motor-route interference had no effect on the generation of non-vocal sounds, indicating that the motor-route has a limited role in generating auditory imagery which is not easily articulable. In low intentionality inner speech, it was found that perceptual-route interference was most effective at suppressing highly egocentric inner speech (i.e. in one’s own voice). In the low egocentricity condition, neither motor-route or perceptual-route interference suppressed rates of inner speech at statistically significant levels. Comparisons of motor-route interference to a control task which involved foot tapping confirmed that any effects of motor-route interference could not be attributed to the additional attentional demands imposed by the secondary task.

Chapter 5 further tested the framework by using transcranial magnetic stimulation to suppress the brain regions implicated in the motor-route (L-IFG) and perceptual-route (L-STG) of generation. The effect of a virtual lesion over each brain region on inner speech of varying intentionality and egocentricity was then examined. As in previous chapters, highly intentional and egocentric inner speech was predicted to place most demands on the motor-route of generation, and therefore the L-IFG. As inner speech decreased in intentionality and egocentricity, it was predicted to place increasing demands on the perceptual-route and L-STG. In both the scenario rating task which examined high intentionality inner speech, and in the spontaneous imagery task examining low intentionality inner speech, we observed a trend (albeit not statistically) towards L-IFG stimulation suppressing inner speech regardless of egocentricity. Stimulation of the L-STG did not suppress any subtype of inner speech. While requiring further examination, the results of the experiment suggest that the L-IFG plays a role in inner speech generation regardless of phenomenology.

The Dual-Mechanistic Framework of Inner Speech

Comparing the initial predictions and justifications of the dual-mechanistic framework to the findings of the empirical Chapters of this thesis reveals several points for discussion. First, it can be argued that the initial basis of the framework was justified, with chapters generally demonstrating that results (e.g. neural correlates, effect of dual-task suppression) varied depending on the phenomenology of inner speech. In the fMRI meta-analysis described in Chapter 2, convergence of brain regions varied widely depending on the phenomenology of the inner speech. This ranged from convergence over speech production regions in tasks which were highly intentional and highly egocentric, to convergence over speech perceptual regions in tasks which were low in intentionality, replicating the findings of Hurlburt et al. (2016). This provided initial evidence in favour of the existence of two cognitive mechanisms, as well as their ability to interact with inner speech phenomenology. Later chapters further replicated this finding, with Chapter 3 demonstrating that some subtypes of inner speech were impacted to a greater degree than others by aphantasia. Chapter 4 demonstrated that different secondary tasks in a dual-task paradigm can differentially impact different subtypes of inner speech.

The finding that inner speech phenomenology interacts with its neurocognitive mechanisms is in line with one of the fundamental assertions of the dual-mechanistic framework, and supports the concerns indicated in previous works (Hurlburt et al., 2016; Yao et al., 2011, 2012). However, a more detailed comparison of the predictions made by our dual-mechanistic framework and the results observed within the empirical chapters reveals some divergence. First, while demonstrating that varieties of inner speech do exhibit distinct neural activation, the results of the ALE meta-analyses (Chapter 2) were not completely in line with the activation assumed by the model. The dual-mechanistic framework argues that inner speech which is high in both intentionality and egocentricity should place the greatest demands on the motor-route of generation and the L-IFG. However, after balancing the meta-analysis dataset to reduce the over-representation of certain paradigms, analysis of high intentionality studies did not yield significant convergence over the L-IFG. The framework also argues that inner speech which is low in egocentricity should place demands on the perceptual-route of generation and speech perceptual regions. However, analysis of low egocentricity studies did not converge over speech perceptual regions. It is notable that while these findings were not predicted by our dual-mechanistic framework, they are not readily explained by models based on a single mechanism of generation. If the corollary-discharge model of inner speech were correct, the meta-analyses should have yielded reliable convergence over the L-IFG. If the perceptual-simulation model of inner speech were correct, more reliable convergence over speech-perceptual regions should have been observed.

Chapters 3 and 4 presented similar cases in which results were broadly in line with our proposed framework, but contained several unanticipated findings which could not be readily explained by alternative models. Chapter 3 analysed the impact of aphantasia on inner speech which varied along intentionality and egocentricity dimensions. As aphantasia could stem from dysfunction of perceptual simulation mechanisms, we posited that it could be used to determine which, if any, types of inner speech were generated via the perceptual-route. While this produced results inconsistent with our dual-mechanistic framework, the chapter did demonstrate that inner speech phenomenology interacts with cognitive mechanisms and should therefore be accounted for in future research.

Chapter 4 analysed the effectiveness of two types of dual-task suppression on inner speech varying along intentionality and egocentricity. Here, results aligned strongly with the predictions made by the dual-mechanistic framework and were at odds with an explanation based purely on corollary discharge or perceptual simulation. Highly intentional and egocentric inner speech was best suppressed by motor-route interference, with motor-route interference decreasing in effectiveness in line with decreases in egocentricity. Within the low intentionality conditions it was found that perceptual-route interference is most effective at suppressing high egocentricity inner speech, but neither motor-route or perceptual-route interference suppressed rates of low egocentricity inner speech. This diverges from the predictions laid out in the dual-mechanistic framework, which predicts that inner speech low in intentionality and egocentricity should place greatest demands on the perceptual-route of suppression. Given the lack of an effect from both motor-route interference and perceptual-route interference, the results do not lend evidence to either a purely corollary discharge or perceptual simulation interpretation of inner speech. Results instead suggest the need for refinement in the choice of suppression technique or the manner by which the inner speech is elicited and measured.

Chapters 2 - 4 present data which broadly supports the dual-mechanistic framework of inner speech, but are also characterised by divergences which indicate room for further refinement in both experimental methodologies, as well as the framework itself. Chapter 5 represents the only chapter which was wholly inconsistent with the idea of a perceptual-route of generation centred on speech perceptual regions. Chapter 5 observed a trend towards TMS stimulation of the L-IFG suppressing subtypes of inner speech within both high intentionality and low intentionality conditions. Stimulation of speech perceptual regions, which we predicted would have a greater impact on inner speech which was low in egocentricity and intentionality, had no observable effect on any subtype of inner speech.

Taken together, we argue that the empirical chapters generally demonstrate an interaction between inner speech phenomenology and underlying neurocognitive correlates which is better explained via our dual-mechanistic framework than by corollary discharge or perceptual simulation models on their own. By recognizing the role of both motor-route and perceptual-route mechanisms, the dual-mechanistic framework accommodates the variety of inner speech experiences and the differing patterns of neural activation observed across studies (e.g., McGuire et al., 1996; Jones & Fernyhough, 2007). The interaction between phenomenology and the cognitive mechanisms of inner speech was strongly demonstrated in the empirical chapters, which lend support to the dual-mechanistic framework, even though results diverged from the initial predictions when examining the neural correlates (Chapter 2 & 5).

One of the main advantages of our more flexible approach lies in its ability to account for the diversity of inner speech experiences described in Alderson-Day (2018) and the different contexts in which they occur (Fernyhough, 2016). The dual-mechanistic framework also enables researchers to generate predictions regarding the relative contributions of the motor-route and perceptual-route mechanisms for a given phenomenological subtype of inner speech. This predictive capacity allows for more targeted empirical investigations and hypothesis-driven research, which can contribute to a deeper understanding of inner speech and its various subprocesses. Even if future researchers do not adopt the dual-mechanistic framework in its entirety, we argue that the frameworks can serve as a broader reminder that inner speech phenomenology can have implications at the neurocognitive level, and that careful selection of elicitation paradigms is required in order to make accurate inferences. For example, if examining the neural correlates of inner speech in individuals prone to rumination, eliciting inner speech via a highly-intentional sentence reading task could yield different results compared to a more spontaneous approach using direct experience sampling.

Directions for Future Research

Given the utility of the dual-mechanistic framework in its current state, we argue that our understanding of inner speech has been significantly advanced through research exploring the interaction between phenomenological attributes and cognitive processes. Yet, further research is required. Research can be divided into two avenues, one avenue further investigating the neural underpinnings of inner speech, on which we observed mixed findings, and another investigating the phenomenological attributes of inner speech. Regarding alternative neural underpinnings, the thesis focused on the L-STG being the hub of perceptual simulation. This was based on the fMRI meta-analyses of Chapter 2, but did not produce clear suppressive effects on inner speech according to the TMS study in Chapter 5. As discussed in Chapter 5, this may be due to methodological limitations or the high focality of the TMS coil, which could have led to suboptimal stimulation of the target region. However, it is crucial for future research to explore other candidate neural targets that have been implicated in spontaneous and less egocentric forms of inner speech. Heschl's Gyrus and the Left Superior Temporal Sulcus, for example, have been identified in previous research (Hurlburt et al., 2016; Yao et al., 2011) as potential regions of interest.

In our examination of inner speech generated via the motor-route, we focused on the L-IFG given the range of studies evidencing its role in generating speech motor commands and efference copies (see Chapter 2). The results of the TMS study described in Chapter 5 support the notion that the L-IFG plays a causal role in the generation of inner speech. However, whether other brain regions also contribute to this mechanism remains unclear. The cerebellum is one potential region which could play a role, with its involvement in inner speech being demonstrated in our meta-analysis of fMRI studies, as well as being associated with motor functions and error prediction (Wolpert et al., 1998). This predictive capability may extend to inner speech, potentially allowing the cerebellum to modulate the sensory predictions associated with inner speech. However, the exact relationship between the cerebellum, error prediction, and inner speech remains ambiguous. The current evidence does not definitively establish whether the cerebellum's involvement is solely related to error prediction and monitoring, or if it plays a more causal role in generating inner speech through the planning of motor commands and production of corollary discharge (Marvel & Desmond, 2010). To clarify this connection, future research may explore the specific pathways and interactions between the cerebellum and other neural regions implicated in inner speech.

Beyond examining individual brain regions, investigating the contributions of brain networks to inner speech is key to furthering our understanding of inner speech production. While the corollary discharge network has been explored in both the context of inner speech and other sensory modalities, the involvement of the fronto-parieto-temporal memory network in inner speech is less understood. As explored in Tian & Poeppel (2010), such a network may underpin inner speech by retrieving episodic memories of speech from neuroanatomically distributed brain regions and propagating that information to speech perceptual regions where they can be simulated. However, our current understanding of its contribution to inner speech is limited to correlational studies, such as the fMRI meta-analyses presented in Chapter 2. This presents challenges in ascertaining whether the network plays a causal role in inner speech production or whether its involvement can be attributed to an indirect role, such as the lexical and semantic processing of inner speech. It also remains unclear precisely how episodic memories contribute to the generation of inner speech, with the perceptual simulation hypothesis proposing the reactivation of neuronal firing patterns stored within the relevant sensory cortex, rather than the retrieval of entire episodic memories from a distributed memory network (Barsalou, 2008). The functional relevance of each region within the network could be disentangled by using non-invasive brain stimulation techniques to inhibit each region within the brain network and examining the corresponding effect on inner speech.

At a network-wide scale, the importance of the fronto-parieto-temporal regions in inner speech could be examined via transcranial direct current stimulation (tDCS). This uses an anode and a cathode placed at any two points across a participant's scalp to deliver a continuous current through the brain, thereby modulating cortical excitability. Usefully these effects, which may be either excitatory or inhibitory, can extend across neuroanatomically distributed brain networks, allowing researchers to investigate the causal relationship between the whole brain network and inner speech (Peña-Gómez et al., 2012). The use of brain stimulation techniques to causally examine brain-behaviour relationships may also be used to examine alternative interpretations of the role the fronto-parieto-temporal network plays in inner speech. For example, Alderson-Day & Fernyhough (2015) describe an alternative explanation whereby the network is used to maintain internally generated representations, rather than to generate them. In such a scenario, it would be hypothesised that reducing excitability across the fronto-parieto-temporal network should not interfere with the number of inner speech occurrences reported over a period of time, but it should interfere with phonological judgement tasks which require the elicited inner speech to be internally maintained for a period of time.

In addition to exploring alternative brain regions or brain networks supporting inner speech, future research should delve into phenomenological attributes of inner speech beyond intentionality and egocentricity. While the current study proposed that these dimensions were critical in determining the neural mechanisms underlying inner speech, the findings do not exclude other possibilities. One dimension that deserves additional exploration is the 'condensed/abstract-expanded' axis. This dimension ranges from a truncated and condensed form of inner speech which lacks the syntactic qualities of overt speech, to a fully expanded form of inner speech with rich acoustic and syntactic qualities. Grandchamp et al. (2019) provided initial evidence in support of such an axis in their research. However, we excluded it from our framework given the partial collinearity with existing axes in the dual-mechanistic model. Specifically, we argued that condensed inner speech only emerges when inner speech is low in intentionality and therefore represents an attribute of low intentionality inner speech, rather than a dimension in and of itself. Given the failure of our dual-mechanistic model to fully account for the empirical findings in this thesis, it would be insightful to further examine the usefulness of this dimension in understanding the neurocognitive underpinnings of inner speech. While difficulties remain in integrating a phenomenological attribute which bears significant overlap with an existing dimension, it may be that a condensation/expanded axis is better represented at a neurocognitive level than our dimension of intentionality, and that substituting the two dimensions results in a framework which better explains our empirical findings without introducing collinearity to our dimensions. An initial step in this process might be the development of experimental paradigms which can measure and manipulate levels of condensation in inner speech, a challenge acknowledged in Grandchamp et al. (2019) and one which limited our ability to empirically examine it.

Grandchamp et al. (2019) also emphasised motor inhibition as a component in inner speech generation. At the core of this argument lies the hypothesis that while overt speech necessitates activation of the motor cortex for articulation, inner speech is characterised by the active inhibition of these motor processes. This reasoning suggests that when we "speak" in our minds, there's a concurrent suppression of the motor commands that would otherwise result in audible speech (Scott, 2013). The work carried out in this thesis does not exclude a role for motor inhibition in inner speech. Indeed, understanding motor inhibition can provide valuable insights into the boundary conditions of when inner speech transitions to overt speech. One implementation of motor inhibition in our framework is that it plays a functional role when inner speech is generated via the motor-route, but not the perceptual-route which does not rely on motor planning. In inner speech generated via the motor-route, motor inhibition may be the mechanism by which the phonological detail of inner speech is modulated. As explored in Oppenheim and Dell's (2010) concept of flexible abstraction, situations where inner speech is more abstract and does not require high levels of phonological detail may allow for motor inhibition to be more pronounced. Conversely, when inner speech requires additional phonological detail, the additional involvement of articulators is beneficial and the degree of motor inhibition might be lessened. Given this potential interaction between intentionality, motor inhibition and phonological detail, future research could examine the hypothesis that subjective increases in the vividness of inner speech are correlated with decreased motor inhibition as determined by electromyography (EMG) or electroencephalography (EEG).

One novel direction by which our understanding of inner speech may be furthered is to examine the extent to which different attributes of inner speech are similarly represented at a neurocognitive level. For example, the concept of agency in inner speech is one which has been explored in previous literature (Hurlburt et al., 2013). This refers to the individual's sense of authorship or ownership of their inner voice. While the notion of agency shares similarities with our dimension of egocentricity, it emphasises the perception of self versus others within inner dialogues, rather than whether the elicited inner voice 'sounds' like that of the speaker (Perrone-Bertolotti et al., 2013). Prima facie, a dimension of agency which focuses on the perceived ownership of inner speech is likely to recruit different cognitive resources compared to a dimension of egocentricity which focuses on acoustic qualities. However, the extent to which egocentricity and agency are meaningfully different at a neurocognitive level, and how agency might interact with other phenomenological dimensions is unknown and presents an interesting avenue for future research.

Another relationship which justifies further research is that of non-verbal thought and inner speech. Inner speech, as a form of self-talk and self-representation, is usually conceptualised within the domain of linguistic thought processes. However, cognition is multifaceted, and non-verbal thoughts represent a significant aspect of the day-to-day human experience (Alderson-Day et al., 2015). Understanding the interplay between non-verbal thought and inner speech opens intriguing avenues of investigation. For instance, non-verbal thought may provide a foundation for the generation of inner speech, serving as a precursor or trigger for the conversion of abstract thoughts into articulated inner dialogue. The nature of non-verbal thought and its implementation at a neurocognitive level is therefore of interest within the context of inner speech research, yet one which is unexplored. Beyond a role as a precursor to inner speech, research might find sufficient similarities between non-verbal thought and inner speech at a neurocognitive level to suggest that they are variants of the same phenomenon. In this case, investigation of non-verbal thoughts and comparisons with its verbal counterpart could help determine which brain regions contribute to the perceptible components of inner speech, and which contribute to the semantic components. Such an investigation requires the further development of experimental paradigms and the use of neuroimaging techniques which elucidate the distinct and shared neural networks involved in verbal and non-verbal cognition.

While continued work on the basic science surrounding inner speech is vital given our limited understanding of how it functions at a cognitive and neural level, avenues for further applied research are numerous, particularly in the contexts of auditory verbal hallucinations (AVH) in schizophrenia and Brain-Computer Interfaces (BCIs). Within the context of auditory verbal hallucinations, our proposal that the neural underpinnings of inner speech vary depending on inner speech phenomenology could have implications for future attempts to reduce hallucination rates via targeted electrical stimulation of brain regions. This subfield of research has yielded mixed results thus far (Koops et al., 2016), but may benefit from a more nuanced approach to selecting neuroanatomical targets which considers the phenomenological characteristics of the auditory hallucination. Similarly, attempts to develop Brain-Computer Interfaces (BCI) which can decode inner speech using electrical activity recorded by intracranial and extracranial electrodes have found limited success (Martin et al., 2018). Future research could attempt to individualise the layout and density of electroencephalogram / electrocorticographic electrodes based on the phenomenology of inner speech they expect to generate and the brain regions which likely support that variety of inner speech.

Conclusion

The overarching finding of this thesis was that inner speech phenomenology can affect underlying neurocognitive mechanisms. However, the exact nature of this relationship is less clear or reliable than was initially anticipated. This suggests the need for further refinement in our proposed framework. An fMRI meta-analysis and series of dual-task experiments both provided evidence that the involvement of motor regions and perceptual regions in inner speech were a function of inner speech phenomenology. The effect of aphantasia, which has been argued to represent a dysfunction of perceptual simulation systems, on inner speech, was also found to vary across phenomenological dimensions. However, the interaction between inner speech phenomenology and the effects of aphantasia on inner speech did not align with the predictions outlined in the dual-mechanistic model. The final TMS study provided the only evidence strongly contrary to the hypothesis of this thesis, with TMS stimulation of the L-IFG producing a trend towards suppression of all inner speech subtypes, regardless of phenomenology.

References

Abramowitz, J. S., Tolin, D. F., & Street, G. P. (2001). Paradoxical effects of thought suppression: a meta-analysis of controlled studies. *Clinical Psychology Review*, *21*(5), 683–703. https://doi.org/10.1016/S0272-7358(00)00057-X

Acar, F., Seurinck, R., Eickhoff, S. B., & Moerkerke, B. (2018). Assessing robustness against potential publication bias in Activation Likelihood Estimation (ALE) meta-analyses for fMRI. *Plos One*, *13*(11), e0208177. https://doi.org/10.1371/journal.pone.0208177

Alderson-Day, B., & Fernyhough, C. (2014). More than one voice: investigating the phenomenological properties of inner speech requires a variety of methods. *Consciousness and Cognition*, *24*, 113–114. https://doi.org/10.1016/j.concog.2013.12.012

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire - Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48–58. https://doi.org/10.1016/j.concog.2018.07.001

Alderson-Day, B., Moffatt, J., Bernini, M., Mitrenga, K., Yao, B., & Fernyhough, C. (2020). Processing Speech and Thoughts during Silent Reading: Direct Reference Effects for Speech by Fictional Characters in Voice-Selective Auditory Cortex and a Theory-of-Mind Network. *Journal of Cognitive Neuroscience*, *32*(9), 1637–1653. https://doi.org/10.1162/jocn_a_01571

Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., & Fernyhough, C. (2016). The brain’s conversation with itself: neural substrates of dialogic inner speech. *Social Cognitive and Affective Neuroscience*, *11*(1), 110–120. https://doi.org/10.1093/scan/nsv094

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*(2), 221–228. https://doi.org/10.1093/cercor/bhh124

Amanzio, M., Benedetti, F., Porro, C. A., Palermo, S., & Cauda, F. (2013). Activation likelihood estimation meta-analysis of brain correlates of placebo analgesia in human experimental pain. *Human Brain Mapping*, *34*(3), 738–752. https://doi.org/10.1002/hbm.21471

Aparicio, M., Gounot, D., Demont, E., & Metz-Lutz, M.-N. (2007). Phonological processing in relation to reading: an fMRI study in deaf readers. *Neuroimage*, *35*(3), 1303–1316. https://doi.org/10.1016/j.neuroimage.2006.12.046

Ashtari, M., Lencz, T., Zuffante, P., Bilder, R., Clarke, T., Diamond, A., Kane, J., & Szeszko, P. (2004). Left middle temporal gyrus activation during a phonemic discrimination task. *Neuroreport*, *15*(3), 389–393. https://doi.org/10.1097/00001756-200403010-00001

Aziz-Zadeh, L., Cattaneo, L., Rochat, M., & Rizzolatti, G. (2005). Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience*, *17*(6), 928–938. https://doi.org/10.1162/0898929054021157

Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, *24*(4), 490–502. https://doi.org/10.1016/0749-596X(85)90041-5

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bain, A. (1859). *The Emotions and the Will by Alexander Bain*. Parker and son.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, Lawrence W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/s1364-6613(02)00029-3

Barsalou, Lawrence W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Bates, D., Mächler, 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

Beaman, C. P., Powell, K., & Rapley, E. (2015). Want to block earworms from conscious awareness?

B(u)y gum! *Quarterly Journal of Experimental Psychology*, *68*(6), 1049–1057. https://doi.org/10.1080/17470218.2015.1034142

Bishop, D. V., & Robson, J. (1989). Accurate non-word spelling despite congenital inability to speak: phoneme-grapheme conversion does not require subvocal articulation. *British Journal of Psychology*, *80 (Pt 1)*, 1–13. https://doi.org/10.1111/j.2044-8295.1989.tb02300.x

Bohning, D. E., Pecheny, A. P., Epstein, C. M., Speer, A. M., Vincent, D. J., Dannels, W., & George, M. S. (1997). Mapping transcranial magnetic stimulation (TMS) fields in vivo with MRI. *Neuroreport*, *8*(11), 2535–2538. https://doi.org/10.1097/00001756-199707280-00023

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Relation between brain activation and lexical performance. *Human Brain Mapping*, *19*(3), 155–169. https://doi.org/10.1002/hbm.10111

Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., & Wildgruber, D. (2014). “Inner voices”: the cerebral representation of emotional voice cues described in literary texts. *Social Cognitive and Affective Neuroscience*, *9*(11), 1819–1827. https://doi.org/10.1093/scan/nst180

Burton, M. W., Locasto, P. C., Krebs-Noble, D., & Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage*, *26*(3), 647–661. https://doi.org/10.1016/j.neuroimage.2005.02.024

Burton, M. W. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science*, *25*(5), 695–709. https://doi.org/10.1207/s15516709cog2505_4

Cereda, C., Ghika, J., Maeder, P., & Bogousslavsky, J. (2002). Strokes restricted to the insular cortex. *Neurology*, *59*(12), 1950–1955. https://doi.org/10.1212/01.wnl.0000038905.75660.bd

Chen, C.-M. A., Mathalon, D. H., Roach, B. J., Cavus, I., Spencer, D. D., & Ford, J. M. (2011). The corollary discharge in humans is related to synchronous neural oscillations. *Journal of Cognitive Neuroscience*, *23*(10), 2892–2904. https://doi.org/10.1162/jocn.2010.21589

Chincotta, D., & Underwood, G. (1998). Simultaneous interpreters and the effect of concurrent articulation on immediate memory. *Interpreting*, *3*(1), 1–20. https://doi.org/10.1075/intp.3.1.01chi

Clowes, R. (2007). A self-regulation model of inner speech and its role in the organisation of human conscious experience. *Journal of Consciousness Studies*.

Cohen, B. H. (1986). The motor theory of voluntary thinking. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 19–54). Springer US.

https://doi.org/10.1007/978-1-4757-0629-1_2

Craig, A. D. B. (2009). How do you feel--now? The anterior insula and human awareness. *Nature Reviews. Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews. Neuroscience*, *9*(8), 587–600. https://doi.org/10.1038/nrn2457

Cui, Y., Liu, B., Song, M., Lipnicki, D. M., Li, J., Xie, S., Chen, Y., Li, P., Lu, L., Lv, L., Wang, H., Yan, H., Yan, J., Zhang, H., Zhang, D., & Jiang, T. (2018). Auditory verbal hallucinations are related to cortical thinning in the left middle temporal gyrus of patients with schizophrenia. *Psychological Medicine*, *48*(1), 115–122. https://doi.org/10.1017/S0033291717001520

Curcic-Blake, B., Liemburg, E., Vercammen, A., Swart, M., Knegtering, H., Bruggeman, R., & Aleman, A. (2013). When Broca goes uninformed: reduced information flow to Broca’s area in schizophrenia patients with auditory hallucinations. *Schizophrenia Bulletin*, *39*(5), 1087–1095. https://doi.org/10.1093/schbul/sbs107

Curtis, H. S. (1900). Automatic movements of the larynx. *The American Journal of Psychology*, *11*(2), 237. https://doi.org/10.2307/1412271

D’Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 761–772. https://doi.org/10.1098/rstb.2007.2086

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-x

Dawes, A. J., Keogh, R., Andrillon, T., & Pearson, J. (2020). A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports*, *10*(1), 10022. https://doi.org/10.1038/s41598-020-65705-7

Deng, Z.-D., Lisanby, S. H., & Peterchev, A. V. (2013). Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. *Brain Stimulation*, *6*(1), 1–13. https://doi.org/10.1016/j.brs.2012.02.005

De Nil, L. F., Kroll, R. M., Kapur, S., & Houle, S. (2000). A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research*, *43*(4), 1038–1053. https://doi.org/10.1044/jslhr.4304.1038

Diaz, R. M., Berk, L. E., & Diaz, R. (Eds.). (2014). *Private Speech: From Social Interaction To Self-regulation*. Psychology Press.

Dijkstra, N, Bosch, S. E., & van Gerven, M. A. J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, Nadine, Ambrogioni, L., Vidaurre, D., & van Gerven, M. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *ELife*, *9*. https://doi.org/10.7554/eLife.53588

Di, X., Huang, J., & Biswal, B. B. (2017). Task modulated brain connectivity of the amygdala: a meta-analysis of psychophysiological interactions. *Brain Structure & Function*, *222*(1), 619–634. https://doi.org/10.1007/s00429-016-1239-4

Djordjevic, J., Zatorre, R. J., Petrides, M., & Jones-Gotman, M. (2004). The mind’s nose: Effects of odor and visual imagery on odor detection. *Psychological Science*, *15*(3), 143–148. https://doi.org/10.1111/j.0956-7976.2004.01503001.x

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159–161. https://doi.org/10.1038/384159a0

Duffau, H., Bauchet, L., Lehéricy, S., & Capelle, L. (2001). Functional compensation of the left dominant insula for language. *Neuroreport*, *12*(10), 2159–2163. https://doi.org/10.1097/00001756-200107200-00023

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*(1), 148–168. https://doi.org/10.1016/S0749-596X(02)00511-9

Falcone, G., & Jerram, M. (2018). Brain Activity in Mindfulness Depends on Experience: a Meta-Analysis of fMRI Studies. *Mindfulness*, *9*(5), 1–11. https://doi.org/10.1007/s12671-018-0884-5

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, *22*(1), 49–68. https://doi.org/10.1016/j.newideapsych.2004.09.001

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader’s regional accent. *Plos One*, *6*(10), e25782. https://doi.org/10.1371/journal.pone.0025782

Fleischmann, R., Köhn, A., Tränkner, S., Brandt, S. A., & Schmidt, S. (2020). Individualized template MRI is a valid and reliable alternative to individual MRI for spatial tracking in navigated TMS studies in healthy subjects. *Frontiers in Human Neuroscience*, *14*, 174. https://doi.org/10.3389/fnhum.2020.00174

Ford, J. M., Roach, B. J., & Mathalon, D. H. (2021). Vocalizing and singing reveal complex patterns of corollary discharge function in schizophrenia. *International Journal of Psychophysiology*, *164*, 30–40. https://doi.org/10.1016/j.ijpsycho.2021.02.013

Frankford, S. A., Nieto-Castañón, A., Tourville, J. A., & Guenther, F. H. (2019). Reliability of single-subject neural activation patterns in speech production tasks. *BioRxiv*. https://doi.org/10.1101/807925

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, *17*(3), 631–648. https://doi.org/10.1017/s0033291700025873

Giovanni, A., & Duflo, S. (2010). Physiology of voice production. In M. Remacle & H. E. Eckel (Eds.), *Surgery of larynx and trachea* (pp. 1–9). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79136-2_1

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.-P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & Lœvenbruck, H. (2019). The condialint model: condensation, dialogality, and intentionality dimensions of inner speech within a hierarchical predictive control framework. *Frontiers in Psychology*, *10*, 2019. https://doi.org/10.3389/fpsyg.2019.02019

Gulyás, B. (2001). Neural networks for internal reading and visual imagery of reading: a PET study. *Brain Research Bulletin*, *54*(3), 319–328. https://doi.org/10.1016/S0361-9230(00)00441-X

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*(9), 1281–1292. https://doi.org/10.1016/j.neuropsychologia.2003.12.017

Hernandez, N., Andersson, F., Edjlali, M., Hommet, C., Cottier, J. P., Destrieux, C., & Bonnet-Brilhault, F. (2013). Cerebral functional asymmetry and phonological performance in dyslexic adults. *Psychophysiology*, *50*(12), 1226–1238. https://doi.org/10.1111/psyp.12141

Hoffman, R. E., & Kravitz, R. E. (1987). Feedforward action regulation and the experience of will.

Behavioral and Brain Sciences, *10*(04), 782. https://doi.org/10.1017/S0140525X00055837

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the ecological validity of thinking on demand: neural correlates of elicited vs. spontaneously occurring inner speech. *Plos One*, *11*(2), e0147932. https://doi.org/10.1371/journal.pone.0147932

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Ishai, A., & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, *268*(5218), 1772–1774. https://doi.org/10.1126/science.7792605

Jack, B. N., Le Pelley, M. E., Han, N., Harris, A. W. F., Spencer, K. M., & Whitford, T. J. (2019). Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *Neuroimage*, *198*, 170–180. https://doi.org/10.1016/j.neuroimage.2019.04.038

Jacobson, E. (1932). Electrophysiology of mental activities. *The American Journal of Psychology*, *44*(4), 677. https://doi.org/10.2307/1414531

Jezzini, A., Caruana, F., Stoianov, I., Gallese, V., & Rizzolatti, G. (2012). Functional organization of the insula and inner perisylvian regions. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(25), 10077–10082. https://doi.org/10.1073/pnas.1200143109

Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2020). Distinguishing vigilance decrement and low task demands from mind-wandering: A machine learning analysis of EEG. *The European Journal of Neuroscience*, *52*(9), 4147–4164. https://doi.org/10.1111/ejn.14863

Johnston, R. S., & McDermott, E. A. (1986). Suppression effects in rhyme judgement tasks. *The Quarterly Journal of Experimental Psychology Section A*, *38*(1), 111–124. https://doi.org/10.1080/14640748608401587

Jones, S. R., & Fernyhough, C. (2007). Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. *Clinical Psychology Review*, *27*(2), 140–154. https://doi.org/10.1016/j.cpr.2006.10.001

Kay, L., Keogh, R., Andrillon, T., & Pearson, J. (2022). The pupillary light response as a physiological index of aphantasia, sensory and phenomenological imagery strength. *ELife*, *11*.

https://doi.org/10.7554/eLife.72484

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60. https://doi.org/10.1016/j.cortex.2017.10.012

Kleider-Offutt, H. M., Grant, A., & Turner, J. A. (2019). Common cortical areas involved in both auditory and visual imageries for novel stimuli. *Experimental Brain Research*, *237*(5), 1279–1287. https://doi.org/10.1007/s00221-019-05492-4

Klomjai, W., Katz, R., & Lackmy-Vallée, A. (2015). Basic principles of transcranial magnetic stimulation (TMS) and repetitive TMS (rTMS). *Annals of Physical and Rehabilitation Medicine*, *58*(4), 208–213. https://doi.org/10.1016/j.rehab.2015.05.005

Koops, S., van Dellen, E., Schutte, M. J., Nieuwdorp, W., Neggers, S. F., & Sommer, I. E. (2016). Theta Burst Transcranial Magnetic Stimulation for Auditory Verbal Hallucinations: Negative Findings From a Double-Blind-Randomized Trial. Schizophrenia bulletin, 42(1), 250–257. https://doi.org/10.1093/schbul/sbv100

Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural Foundations of Imagery. *Nature Reviews Neuroscience*, *2(9)*, 635–642. https://doi.org/10.1038/35090055

Krieger-Redwood, K., & Jefferies, E. (2014). TMS interferes with lexical-semantic retrieval in left inferior frontal gyrus and posterior middle temporal gyrus: Evidence from cyclical picture naming. *Neuropsychologia*, *64*, 24–32. https://doi.org/10.1016/j.neuropsychologia.2014.09.014

Kuhl, B., Rissman, J., Chun, M., & Wagner, A. (2010). Selective Remembering: Multivoxel Pattern Analysis of Cortical Reactivation During Retrieval of Visual Images. *Journal of Vision*, *10*(7), 739–739. https://doi.org/10.1167/10.7.739

Lalys, F., Haegelen, C., Ferre, J.-C., El-Ganaoui, O., & Jannin, P. (2010). Construction and assessment of a 3-T MRI brain template. *Neuroimage*, *49*(1), 345–354. https://doi.org/10.1016/j.neuroimage.2009.08.007

Langland-Hassan, P., Gauker, C., Richardson, M. J., Dietz, A., & Faries, F. R. (2017). Metacognitive deficits in categorization tasks in a population with impaired inner speech. *Acta Psychologica*, *181*, 62–74. https://doi.org/10.1016/j.actpsy.2017.10.004

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling Visual Imagery and Perception of Real-World Objects. *Neuroimage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Leube, D. T., Knoblich, G., Erb, M., Schlotterbeck, P., & Kircher, T. T. J. (2010). The neural basis of disturbed efference copy mechanism in patients with schizophrenia. *Cognitive Neuroscience*, *1*(2), 111–117. https://doi.org/10.1080/17588921003646156

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, *21*(2), 330–337. https://doi.org/10.1093/cercor/bhq097

Lurito, J. T., Kareken, D. A., Lowe, M. J., Chen, S. H., & Mathews, V. P. (2000). Comparison of rhyming and word generation with FMRI. *Human Brain Mapping*, *10*(3), 99–106. https://doi.org/10.1002/1097-0193(200007)10:3<99::AID-HBM10>3.0.CO;2-Q

Mackay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior*, *13*(4), 274–285. https://doi.org/10.1080/00222895.1981.10735253

MacSweeney, M., Brammer, M. J., Waters, D., & Goswami, U. (2009). Enhanced activation of the left inferior frontal gyrus in deaf and dyslexic adults during rhyming. *Brain: A Journal of Neurology*, *132*(Pt 7), 1928–1940. https://doi.org/10.1093/brain/awp129

Martin, S., Iturrate, I., Millán, J. D. R., Knight, R. T., & Pasley, B. N. (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. Frontiers in neuroscience, 12, 422. https://doi.org/10.3389/fnins.2018.00422

Marvel, C. L., & Desmond, J. E. (2012). From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain and Language*, *120*(1), 42–51. https://doi.org/10.1016/j.bandl.2011.08.005

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Max, L. W. (1935). An experimental study of the motor theory of consciousness. III. Action-current responses in deaf-mutes during sleep, sensory stimulation and dreams. *Journal of Comparative Psychology*, *19*(3), 469–486. https://doi.org/10.1037/h0061830

McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition*, *20*(4), 1586–1593. https://doi.org/10.1016/j.concog.2011.08.005

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*, *6*, 285. https://doi.org/10.3389/fnhum.2012.00285

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: the effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, *115*(2–3), 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004

Modinos, G., Ormel, J., & Aleman, A. (2009). Activation of anterior insula during self-reflection. *Plos One*, *4*(2), e4618. https://doi.org/10.1371/journal.pone.0004618

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Clinical Research Ed.)*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Molnar-Szakacs, I., Iacoboni, M., Koski, L., & Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cerebral Cortex*, *15*(7), 986–994. https://doi.org/10.1093/cercor/bhh199

Morin, A., & Hamper, B. (2012). Self-reflection and the inner voice: activation of the left inferior frontal gyrus during perceptual and conceptual self-referential thinking. *The Open Neuroimaging Journal*, *6*, 78–89. https://doi.org/10.2174/1874440001206010078

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: inner speech use during self-related processing. *Brain Research Bulletin*, *74*(6), 387–396. https://doi.org/10.1016/j.brainresbull.2007.06.013

Morin, A. (2009). Self-awareness deficits following loss of inner speech: Dr. Jill Bolte Taylor's case study. *Consciousness and Cognition*, *18*(2), 524–529. https://doi.org/10.1016/j.concog.2008.09.008

Moseley, P., Alderson-Day, B., Ellison, A., Jardri, R., & Fernyhough, C. (2015). Non-invasive Brain Stimulation and Auditory Verbal Hallucinations: New Techniques and Future Directions. *Frontiers in Neuroscience*, *9*, 515. https://doi.org/10.3389/fnins.2015.00515

Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., Tench, C. R., Yarkoni, T., Nichols, T. E., Turkeltaub, P. E., Wager, T. D., & Eickhoff, S. B. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *84*, 151–161. https://doi.org/10.1016/j.neubiorev.2017.11.012

Netsell, R., Kleinsasser, S., & Daniel, T. (2016). The rate of expanded inner speech during spontaneous sentence productions. *Perceptual and Motor Skills*, *123*(2), 383–393. https://doi.org/10.1177/0031512516664992

Oh, A., Duerden, E. G., & Pang, E. W. (2014). The role of the insula in speech and language processing. *Brain and Language*, *135*, 96–103. https://doi.org/10.1016/j.bandl.2014.06.003

Olivetti Belardinelli, M., Palmiero, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., D’Ausilio, A., Ferretti, A., Del Gratta, C., & Romani, G. L. (2009). An fMRI investigation on image generation in different sensory modalities: the influence of vividness. *Acta Psychologica*, *132*(2), 190–200. https://doi.org/10.1016/j.actpsy.2009.06.009

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160. https://doi.org/10.3758/MC.38.8.1147

Palmiero, M., Piccardi, L., Giancola, M., Nori, R., D’Amico, S., & Olivetti Belardinelli, M. (2019). The format of mental imagery: from a critical review to an integrated embodied representation approach. *Cognitive Processing*, *20*(3), 277–289. https://doi.org/10.1007/s10339-019-00908-z

Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982–986. https://doi.org/10.1016/j.cub.2008.05.048

Peña-Gómez, C., Sala-Lonch, R., Junqué, C., Clemente, I. C., Vidal, D., Bargalló, N., Falcón, C., Valls-Solé, J., Pascual-Leone, Á., & Bartrés-Faz, D. (2012). Modulation of large-scale brain networks by transcranial direct current stimulation evidenced by resting-state functional MRI. Brain stimulation, 5(3), 252–263. https://doi.org/10.1016/j.brs.2011.08.006

Perky, C. W. (1910). An Experimental Study of Imagination. *The American Journal of Psychology*, *21*(3), 422. https://doi.org/10.2307/1413350

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Poulton, E. C. (1977). Continuous intense noise masks auditory feedback and inner speech.

Psychological Bulletin, *84*(5), 977–1001. https://doi.org/10.1037/0033-2909.84.5.977

Pounder, Z., Jacob, J., Evans, S., Loveday, C., Eardley, A. F., & Silvanto, J. (2022). Only minimal differences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. *Cortex*, *148*, 180–192. https://doi.org/10.1016/j.cortex.2021.12.010

Pratts, J., Pobric, G., & Yao, B. (2023). Bridging Phenomenology and Neural Mechanisms of Inner Speech: ALE Meta-Analysis on Egocentricity and Intentionality in a Dual-Mechanistic Framework. Retrieved from osf.io/gaum4

Raij, T. T., & Riekki, T. J. J. (2017). Dorsomedial prefontal cortex supports spontaneous thinking per se. *Human Brain Mapping*, *38*(6), 3277–3288. https://doi.org/10.1002/hbm.23589

Ranganath, C., & D’Esposito, M. (2005). Directing the mind’s eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Current Opinion in Neurobiology*, *15*(2), 175–182. https://doi.org/10.1016/j.conb.2005.03.017

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind’s eye: decoding category information during mental imagery. *Neuroimage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Ripp, I., Zur Nieden, A.-N., Blankenagel, S., Franzmeier, N., Lundström, J. N., & Freiherr, J. (2018). Multisensory integration processing during olfactory-visual stimulation-An fMRI graph theoretical network analysis. *Human Brain Mapping*, *39*(9), 3713–3727. https://doi.org/10.1002/hbm.24206

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039. https://doi.org/10.1016/j.clinph.2009.08.016

Ruiz Vargas, E., Sörös, P., Shoemaker, J. K., & Hachinski, V. (2016). Human cerebral circuitry related to cardiac control: A neuroimaging meta-analysis. *Annals of Neurology*, *79*(5), 709–716. https://doi.org/10.1002/ana.24642

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, *21*(5), 1309–1315. https://doi.org/10.3758/s13423-013-0580-3

Samartsidis, P., Montagna, S., Laird, A. R., Fox, P. T., Johnson, T. D., & Nichols, T. E. (2020).

Estimating the prevalence of missing experiments in a neuroimaging meta-analysis. *Research Synthesis Methods*, *11*(6), 866–883. https://doi.org/10.1002/jrsm.1448

Schmidt, T. T., & Blankenburg, F. (2019). The somatotopy of mental tactile imagery. *Frontiers in Human Neuroscience*, *13*, 10. https://doi.org/10.3389/fnhum.2019.00010

Schmidt, T. T., Ostwald, D., & Blankenburg, F. (2014). Imaging tactile imagery: changes in brain connectivity support perceptual grounding of mental images in primary sensory cortices. *Neuroimage*, *98*, 216–224. https://doi.org/10.1016/j.neuroimage.2014.05.014

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, *133*(4), EL286-92. https://doi.org/10.1121/1.4794932

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, *24*(9), 1824–1830. https://doi.org/10.1177/0956797613478614

Segal, S. J., & Fusella, V. (1969). Effects of imaging of signal-to-noise ratio, with varying signal conditions. *British Journal of Psychology*, *60*(4), 459–464. https://doi.org/10.1111/j.2044-8295.1969.tb01219.x

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: are speech perception and production separate processes? *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *37*(4), 507–532. https://doi.org/10.1080/14640748508400917

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. Springer New York. https://doi.org/10.1007/978-1-4612-0795-5

Shergill, Sukhwinder S, Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, *16*(4), 219–227. https://doi.org/10.1002/hbm.10046

Shergill, Sukhwinder S, White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2013). Modulation of somatosensory processing by action. *Neuroimage*, *70*, 356–362. https://doi.org/10.1016/j.neuroimage.2012.12.043

Shergill, S S, Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*(2), 241–253. https://doi.org/10.1017/s003329170100335x

Shklovsky, V. M., Varlamov, S. A., Petrushevsky, A. G., & Mayorova, L. A. (2019). Speech and Non-Speech Sound Categorization in Auditory Cortex: fMRI Correlates. *Human Physiology*, *45*(6), 577–586. https://doi.org/10.1134/S0362119719060124

Smith, M., Sandberg, A. D., & Larsson, M. (2009). Reading and spelling in children with severe speech and physical impairments: a comparative study. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *44*(6), 864–882. https://doi.org/10.1080/13682820802389873

Starkstein, S. E., Berthier, M., & Leiguarda, R. (1988). Bilateral opercular syndrome and crossed aphemia due to a right insular lesion: a clinicopathological study. *Brain and Language*, *34*(2), 253–261. https://doi.org/10.1016/0093-934x(88)90137-x

Thielscher, A., & Kammer, T. (2004). Electric field properties of two commercial figure-8 coils in TMS: calculation of focality and efficiency. *Clinical Neurophysiology*, *115*(7), 1697–1708. https://doi.org/10.1016/j.clinph.2004.02.019

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

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020–1036. https://doi.org/10.1162/jocn_a_00381

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12. https://doi.org/10.1016/j.cortex.2016.01.002

Tomasino, B., Del Negro, I., Garbo, R., Gigli, G. L., D'Agostini, S., & Valente, M. R. (2022). Multisensory mental imagery of fatigue: Evidence from an fMRI study. *Human Brain Mapping*, *43*(10), 3143–3152. https://doi.org/10.1002/hbm.25839

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

Vallar, G., & Cappa, S. F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology*, *4*(1), 55–77. https://doi.org/10.1080/02643298708252035

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *Neuroimage*, *54*(1), 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036

Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*(6), 1083–1094. https://doi.org/10.1162/jocn.2009.21309

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky, Vol. 1: Problems of general psychology.* https://psycnet.apa.org/record/1988-97267-000

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, *53*(1), 5–13. https://doi.org/10.1037/0022-3514.53.1.5

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

Wheeler, M. E., Shulman, G. L., Buckner, R. L., Miezin, F. M., Velanova, K., & Petersen, S. E. (2006). Evidence for separate perceptual reactivation and search processes during remembering. *Cerebral Cortex*, *16*(7), 949–959. https://doi.org/10.1093/cercor/bhj037

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Wicken, M., Keogh, R., & Pearson, J. (2021). The critical role of mental imagery in human emotion: insights from fear-based imagery and aphantasia. *Proceedings. Biological Sciences / the Royal Society*, *288*(1946), 20210267. https://doi.org/10.1098/rspb.2021.0267

Wilding, J., & Mohindra, N. (1980). Effects of subvocal suppression, articulating aloud and noise on sequence recall. *British Journal of Psychology*, *71*(2), 247–261. https://doi.org/10.1111/j.2044-8295.1980.tb01742.x

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. Trends in cognitive sciences, 2(9), 338–347. https://doi.org/10.1016/s1364-6613(98)01221-2

Yao, B, & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*(3), 447–453. https://doi.org/10.1016/j.cognition.2011.08.007

Yao, Bo, Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–3152. https://doi.org/10.1162/jocn_a_00022

Yao, Bo, Belin, P., & Scheepers, C. (2012). Brain “talks over” boring quotes: top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *Neuroimage*, *60*(3), 1832–1842. https://doi.org/10.1016/j.neuroimage.2012.01.111

Yao, Bo, Taylor, J. R., Banks, B., & Kotz, S. A. (2021). Reading direct speech quotes increases theta phase-locking: Evidence for cortical tracking of inner speech? *Neuroimage*, *239*, 118313. https://doi.org/10.1016/j.neuroimage.2021.118313

Yaple, Z. A., & Yu, R. (2020). Upward and downward comparisons across monetary and status domains. *Human Brain Mapping*, *41*(16), 4662–4675. https://doi.org/10.1002/hbm.25148

Zatorre, R., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/s1364-6613(00)01816-7

Zatorre, Robert J, & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron*, *47*(1), 9–12. https://doi.org/10.1016/j.neuron.2005.06.013

Zatorre, R J, Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind’s ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*(1), 29–46. https://doi.org/10.1162/jocn.1996.8.1.29

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery - Congenital aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zhang, L., Li, B., Wang, H., Li, L., Liao, Q., Liu, Y., Bao, X., Liu, W., Yin, H., Lu, H., & Tan, Q. (2017). Decreased middle temporal gyrus connectivity in the language network in schizophrenia patients with auditory verbal hallucinations. *Neuroscience Letters*, *653*, 177–182. https://doi.org/10.1016/j.neulet.2017.05.042

Appendices

Appendix 1 - MRI coordinates inputted into GingerALE in Chapter 2

// Reference=Talairach
//Linden,2010: Imagery > Baseline ,Increases
// Subjects=7

43	-30	4
58	-41	14
34	-61	-18
46	14	5
19	7	6
-54	-34	9
-53	-39	23
-5	13	49
-46	20	9
-39	5	48
-17	4	8
-2	22	26
-1	-55	0

// Aleman,2005: Imagery > Baseline ,Increases
// Subjects=6

-43	-44	4
-2	0	55
-44	-10	47
-48	-27	47
-22	-67	34
-58	2	33
-29	47	27
-32	0	11
7	6	36
51	-23	45
29	-59	40
38	9	11
-41	-61	-3
33	-67	-5
-18	-84	3
12	-80	3

// Kleider-Offutt,2019: Auditory Imagery > Baseline ,Increases
// Subjects=28

-56.65	-24.19	4.63
-63.99	-34.63	-3.69
54.39	-28.49	6.1

-8.5	-86.39	-14.86
-8.65	-91.16	-4.51

// Lurito,2000: Rhyming > Baseline ,Increases
// Subjects=5

-46	32	25
-49	11	32
-55	13	7
-63	-31	26
-57	-46	1
-44	-59	-18

//Papathanassiou, 2000: Verb Generation > Rest ,Increases
// Subjects=8

-40.14	18.53	28.78
-54.76	7.13	11.23
-28.75	18.7	7.36
-40.35	-3.19	39.33
-21.56	0.67	18.39
-47.19	28.46	4.37
-48.95	29.17	-2.8
-24.22	43.54	22.41
45.49	-60.53	-31.32
28.16	-59.91	20.7
32.72	-62.8	-46.16
-51.15	-35.57	5.45
-3.44	1.34	49.39
6.7	-76.23	-42.47
6.57	-36.98	-20.74
56.44	5.91	0.39
52.81	22.87	0.13
-12.02	-33.85	-13.55
-41.62	-53.67	-21.33

// Nil,2000: Silent Reading > Baseline ,Increases
// Subjects=10

-36	-80	-20
-8	-92	-20
38	14	4
-10	46	24
-44	10	8

// Gulyas,2001: Letter Listing > Baseline ,Increases
// Subjects=10

-50	-8	42

// Rudner,2004: Rhyme Judgement > Baseline ,Increases
// Subjects=12

-4	6	44
4	17	32
-59	-1	-10
-63	-16	1
-55	11	-7
63	-20	-2
48	19	-11
40	-56	43
48	-41	43
48	-29	42
-32	-52	-24
-32	-63	-20
-24	-59	-17
28	-63	-20
12	-75	-16
-12	-16	1

// Okada,2018: Articulate TT > Baseline ,Increases
// Subjects=21

-47	-71	-16
-59	9	-2
49	-33	58
49	-37	4
-45	-3	56
-1	23	26
41	39	24
-39	45	22
-33	-51	36
41	-63	54
19	51	-14
-35	27	-2
37	45	30

// Okada,2017: Imagine TT > Baseline ,Increases
// Subjects=24

-45	-69	-24
-47	-3	50
1	1	54
-45	-49	52
19	3	8
55	-3	42
33	45	24
43	-43	42
-11	-17	4
-27	-17	-6

// Raij,2017: Verbal Thought > Baseline ,Increases

```
// Subjects=51
4.34   -63.54 1.03
18.22  46.09  33.27
23.93  -29.58 -.82
18.71  -73.67 -40.22
-10.15 -15.86 62.06
-9.75  6.32   37.14
-1.67  1.65   55.75
12.57  -72.49 5.73

// CurcicBlake, 2012: Stress > Baseline ,Increases
// Subjects=31
-31.65 -85.43 -14.26
32.11  -80.95 -4.65
29.71  -56.23 -26.67
-4.48  -4.94  55.98
32.38  22.46  5.15
-46.11 -40.27 43.82
-42.98 0.67   28.84
34.44  -51.35 38.73
31.62  -2.33  56.84

// Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
// Subjects=16
-47.39 14.2   15.63
-40.16 18.35  30.56
-26.71 31.09  -4.04
-42.15 -.96   35.9
13.64  -94.31 -11.63
28.79  -60.79 -28.02
-14.08 -86.54 -13.17
-5.37  6.24   57.03
-52.79 -45    -13.49
28.56  21.6   14.01
-23.91 -77.97 34.32

// MacSweeney, 2009: Rhyme Judgement > Picture Judgement ,Increases
// Subjects=7
-40    4      33
-47    22     3
-18    -78    36
0      15     50
36     -70    -13
-58    -37    -13

// Wilson, 2011: Generate word using probe > Rest ,Increases
// Subjects=26
-7     0      58
```

```
-51     2       29
-26     -16     55
-24     -72     26
-42     -37     -22
34      17      1
-32     13      3
-18     -6      14
-1      -44     -41
-38     -46     -39

// Booth,2003: Rhyme Judgement > Control ,Increases
// Subjects=13
-48.48  15.69   29.27
37.64   -28.17  13.06
54.52   -10.18  1.53
-51     -18.02  -.99

// Aparacio,2007: Rhyming > Baseline ,Increases
// Subjects=12
-45.71  1.96    25.31
-52.91  -44     -4.39

// Theys,2020: Reading > Symbols, Increases
// Subjects=11
-6.36   -4.26   57.72
-4.33   4.4     46.96
3.83    -.59    59.24
-48.81  -11.38  46.52
-48.7   -6.96   39.73
-53.76  6.8     4.91
-52.29  -40.77  20.25
-57.55  -27.65  1.59
-57.66  -44.94  5.35
23.2    -57.39  -24.19
6.42    -67.31  -18.21
26.96   -51.47  -27.17
-40.64  -42.32  -22.04
-35.03  -38.28  -25.17
46.8    20.44   34.96
41.26   26.06   35.37

//Yao,2011: Direct Quotations > Indirect ,Increases
// Subjects=16
-56.6   -32.98  -1.61
-59.47  -41.87  2.91
-56.81  -45.46  10.72
38.18   10.54   -17.5
43.67   12.79   -11.79
```

```
49.05	-13.18	-6.05
51.98	7.15	-12.18
57.39	-7.63	-5.38

//Alderson-Day,2020: Direct Quotations > Indirect ,Increases
// Subjects=21
37.26	-66.59	31.03
35.33	-72.69	35.82
50.38	-56.3	21.42
-47.38	25.2	18.48
-49.34	20.79	25.23
-41.66	28.26	6.25
-23.98	-78.5	39.68
-27.29	-86.52	6.43
-25.49	-77.74	12.7

// Grandchamp,2019: Mind Wandering > Baseline ,Increases
// Subjects=24
-9.86	10.02	47.4
-10.07	-2.47	59.73
-42.53	17.69	-1.07
-48.25	8.29	8.75
26.34	-82.58	12.21
-17.68	8.39	6.58
-46.04	-9.62	47.63
-14.72	-14.95	-14.5
48.43	-6.79	44.09
-28.79	44	17.87
32.31	13.72	7.93
21.18	2.6	6.68
15.63	19.14	10.86
43.57	-61.98	-26.08
32.48	-53.53	-25.47
57.1	3.9	20.93

// Grandchamp,2019: Other Voice > Baseline ,Increases
// Subjects=24
17.61	-60.78	49.26
-6.92	22.23	37.8
-9.98	6.44	55.17
-18.4	8.5	63.33
-28.83	43.74	20.55
-23.54	36.29	38.85
37.2	-40.53	43.41
46.07	-61.29	-4.35
31.9	-79.81	12.57
-18.53	-69.24	50.55
-48.83	-56.33	35.05
```

51.71	13.36	10.92
32.34	11.19	4.98
43.53	17.24	0.34
26.55	42.4	32.17
12.18	11.39	61.42
-39.76	14.88	-1.29
-31.54	14.05	6.88
-34.57	-70.04	1.56
-31.3	-52.68	-31.87
20.46	5.5	63.7
43.49	30.7	7.02
23.95	40.93	18.47
29.47	43.44	21.5

Appendix 2 - Output of GingerALE in Chapter 2: All Studies - Unbalanced

```
Mask:
Reference Space            = Talairach
Dimensions                 = 80x96x70
Number of within-brain voxels = 198111
Mask Size                  = More Conservative (Smaller)

Foci:
Coordinate System          = Talairach
File of foci coordinates     = Pooled.txt
Number of foci             = 227
Number of experiments        = 22
Total number of subjects     = 383

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = Pooled_ALE.nii
FWHM minimum value            = 8.746152498730366
FWHM median value            = 9.6586470675157
FWHM maximum value            = 11.37334466796002
Minimum ALE score           = 2.6960556E-24
Maximum ALE score            = 0.03352563

P Values: Eickhoff (HBM, 2009)
File of P values            = Pooled_P.nii
Minimum P value              = 3.0931677E-11

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value           = 0.05
Thresholding Permutations     = 5000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value         = 0.001
Volume > Threshold            = 9576 mm^3
Chosen min. cluster size      = 672 mm^3
Thresholded ALE image         = Pooled_C05_5k_ALE.nii

Cluster Analysis:
 #1: 3608 mm^3 from (-12,-8,40) to (4,14,62) centered at (-4.6,2.3,53.8) with 1 peaks
    with a max value of 0.0335 ALE, 3.0931677E-11 P, 6.54 Z at (-4,0,56)
   Labels: (Gray Matter only)
    Hemisphere: 93.3% Left Cerebrum, 6.7% Right Cerebrum
    Lobe: 88.9% Frontal Lobe, 11.1% Limbic Lobe
    Gyrus: 64.4% Medial Frontal Gyrus, 24.4% Superior Frontal Gyrus, 11.1% Cingulate
Gyrus
    Cell Type: 81.3% Brodmann area 6, 10.2% Brodmann area 32, 8.4% Brodmann area 24

 #2: 2168 mm^3 from (-52,-14,24) to (-40,6,52) centered at (-45.6,-3.8,39.1) with 2 peaks
```

```
    with a max value of 0.0213 ALE, 7.267906E-7 P, 4.82 Z at (-46,-10,46)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 95.3% Frontal Lobe, 4.7% Parietal Lobe
    Gyrus: 76.6% Precentral Gyrus, 17.2% Inferior Frontal Gyrus, 4.7% Postcentral Gyrus,
1.6% Middle Frontal Gyrus
    Cell Type: 53.9% Brodmann area 6, 27.3% Brodmann area 4, 14.1% Brodmann area 9,
4.7% Brodmann area 3

 #3: 1280 mm^3 from (30,8,0) to (46,22,10) centered at (35.3,14.9,4.6) with 1 peaks
    with a max value of 0.0214 ALE, 6.3723417E-7 P, 4.84 Z at (34,14,6)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebrum
    Lobe: 95.5% Sub-lobar, 4.5% Frontal Lobe
    Gyrus: 81.8% Insula, 13.6% Claustrum, 4.5% Inferior Frontal Gyrus
    Cell Type: 65.2% Brodmann area 13, 3% Brodmann area 47

 #4: 936 mm^3 from (24,-62,-30) to (34,-50,-20) centered at (29.3,-56.9,-25.7) with 1 peaks
    with a max value of 0.0217 ALE, 5.208812E-7 P, 4.88 Z at (30,-56,-26)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebellum
    Lobe: 74.4% Anterior Lobe, 25.6% Posterior Lobe
    Gyrus: 57.3% Culmen, 11.1% Declive, 8.5% Tuber, 6% Pyramis
    null

 #5: 840 mm^3 from (-56,6,4) to (-44,14,14) centered at (-50,9.5,8.4) with 1 peaks
    with a max value of 0.0177 ALE, 1.0609114E-5 P, 4.25 Z at (-50,8,8)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 100% Frontal Lobe
    Gyrus: 91.3% Precentral Gyrus, 8.7% Inferior Frontal Gyrus
    Cell Type: 100% Brodmann area 44

 #6: 744 mm^3 from (-48,14,-4) to (-38,30,8) centered at (-43.2,20.7,2.2) with 2 peaks
    with a max value of 0.0163 ALE, 2.9530602E-5 P, 4.02 Z at (-42,16,0)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 57.1% Frontal Lobe, 42.9% Sub-lobar
    Gyrus: 57.1% Inferior Frontal Gyrus, 42.9% Insula
    Cell Type: 42.9% Brodmann area 13, 26.2% Brodmann area 47, 9.5% Brodmann area
45


Experiment Table:
[ 1 1 0 0 1 0 0 1 0 1 1 1 1 0 1 0 0 3 0 0 2 1 ]
[ 0 1 0 0 0 0 1 0 0 1 0 1 1 1 1 0 1 2 0 0 1 0 ]
[ 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 2 ]
[ 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 2 0 0 1 0 ]
```

```
[ 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 ]
[ 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 ]

Contributors to cluster #1
1 foci from //Linden,2010: Imagery > Baseline ,Increases
1 foci from Aleman,2005: Imagery > Baseline ,Increases
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from Raij,2017: Verbal Thought > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases
2 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
1 foci from Grandchamp,2019: Other Voice > Baseline ,Increases

Contributors to cluster #2
1 foci from Aleman,2005: Imagery > Baseline ,Increases
1 foci from Gulyas,2001: Letter Listing > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from MacSweeney, 2009: Rhyme Judgement > Picture Judgement ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Aparacio,2007: Rhyming > Baseline ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #3
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
2 foci from Grandchamp,2019: Other Voice > Baseline ,Increases

Contributors to cluster #4
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #5
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
```

Contributors to cluster #6
1 foci from MacSweeney, 2009: Rhyme Judgement > Picture Judgement ,Increases
1 foci from //Alderson-Day,2020: Direct Quotations > Indirect ,Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
1 foci from Grandchamp,2019: Other Voice > Baseline ,Increases

Appendix 3 - Output of GingerALE in Chapter 2: All Studies - Balanced

```
Mask:
Reference Space             = Talairach
Dimensions                  = 80x96x70
Number of within-brain voxels = 198111
Mask Size                   = More Conservative (Smaller)

Foci:
Coordinate System           = Talairach
File of foci coordinates      = Pooled_Balanced copy.txt
Number of foci              = 163
Number of experiments         = 14
Total number of subjects      = 311

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = Pooled_Balanced copy_ALE.nii
FWHM minimum value            = 8.746152498730366
FWHM median value            = 9.203418613972946
FWHM maximum value            = 10.362276299774612
Minimum ALE score            = 1.976042E-25
Maximum ALE score             = 0.029073441

P Values: Eickhoff (HBM, 2009)
File of P values            = Pooled_Balanced copy_P.nii
Minimum P value              = 5.157227E-10

Thresholding:
Threshold Method            = Cluster-level Inference
Thresholding Value          = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value        = 0.001
Volume > Threshold           = 4776 mm^3
Chosen min. cluster size      = 632 mm^3
Thresholded ALE image        = Pooled_Balanced copy_C05_1k_ALE.nii

Cluster Analysis:
 #1: 2824 mm^3 from (-12,-8,46) to (4,10,62) centered at (-4.8,1,55.3) with 1 peaks
    with a max value of 0.0291 ALE, 5.157227E-10 P, 6.1 Z at (-4,0,56)
  Labels: (Gray Matter only)
    Hemisphere: 92.3% Left Cerebrum, 7.7% Right Cerebrum
    Lobe: 95.8% Frontal Lobe, 4.2% Limbic Lobe
    Gyrus: 76.8% Medial Frontal Gyrus, 19% Superior Frontal Gyrus, 4.2% Cingulate Gyrus
    Cell Type: 94.6% Brodmann area 6, 4.2% Brodmann area 24, 1.2% Brodmann area 32

 #2: 1128 mm^3 from (30,10,0) to (44,22,10) centered at (34.2,15.4,4.4) with 1 peaks
    with a max value of 0.0204 ALE, 6.936911E-7 P, 4.83 Z at (34,14,4)
```

```
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebrum
    Lobe: 98% Sub-lobar, 2% Frontal Lobe
    Gyrus: 80.4% Insula, 17.6% Claustrum, 2% Inferior Frontal Gyrus
    Cell Type: 64.7% Brodmann area 13, 2% Brodmann area 47

 #3: 824 mm^3 from (24,-62,-30) to (34,-50,-22) centered at (29.4,-55.9,-26.6) with 1 peaks
    with a max value of 0.0208 ALE, 5.02383E-7 P, 4.89 Z at (30,-56,-26)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebellum
    Lobe: 81.6% Anterior Lobe, 18.4% Posterior Lobe
    Gyrus: 57.3% Culmen, 9.7% Tuber, 8.7% Pyramis
    null


Experiment Table:
[ 1 0 1 0 0 1 1 1 1 1 3 0 0 2 ]
[ 2 0 0 1 0 0 0 1 0 1 0 0 0 1 ]
[ 0 0 0 0 0 0 0 1 1 0 2 0 0 1 ]

Contributors to cluster #1
1 foci from Grandchamp,2019: Other Voice > Baseline ,Increases
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from Raij,2017: Verbal Thought > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases
2 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #2
2 foci from Grandchamp,2019: Other Voice > Baseline ,Increases
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #3
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
```

Appendix 4 - Output of GingerALE in Chapter 2: High Intentionality - Unbalanced

```
Mask:
Reference Space              = Talairach
Dimensions                  = 80x96x70
Number of within-brain voxels = 198111
Mask Size                   = More Conservative (Smaller)

Foci:
Coordinate System            = Talairach
File of foci coordinates      = High Intent.txt
Number of foci               = 128
Number of experiments         = 14
Total number of subjects      = 206

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels           = High Intent_ALE.nii
FWHM minimum value            = 8.95555138448401
FWHM median value            = 9.755397007122182
FWHM maximum value            = 11.37334466796002
Minimum ALE score            = 8.3008935E-29
Maximum ALE score            = 0.021453783

P Values: Eickhoff (HBM, 2009)
File of P values             = High Intent_P.nii
Minimum P value              = 2.7246754E-8

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value           = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value         = 0.01
Volume > Threshold           = 14616 mm^3
Chosen min. cluster size      = 1728 mm^3
Thresholded ALE image         = High Intent_C05_1k_ALE.nii

Cluster Analysis:
 #1: 6320 mm^3 from (-54,-14,22) to (-36,22,56) centered at (-45,2.8,34.4) with 4 peaks
    with a max value of 0.0182 ALE, 5.066024E-7 P, 4.89 Z at (-44,2,30)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 95.9% Frontal Lobe, 4.1% Parietal Lobe
    Gyrus: 56.5% Precentral Gyrus, 19.8% Middle Frontal Gyrus, 19.5% Inferior Frontal
Gyrus, 4.1% Postcentral Gyrus
    Cell Type: 47.6% Brodmann area 6, 38.5% Brodmann area 9, 9.8% Brodmann area 4,
4.1% Brodmann area 3
```

```
 #2: 4152 mm^3 from (-12,-10,40) to (6,14,62) centered at (-3.6,1.2,53.5) with 2 peaks
    with a max value of 0.0215 ALE, 2.7246754E-8 P, 5.44 Z at (-6,-2,56)
  Labels: (Gray Matter only)
    Hemisphere: 84.4% Left Cerebrum, 15.6% Right Cerebrum
    Lobe: 87.8% Frontal Lobe, 12.2% Limbic Lobe
    Gyrus: 70.2% Medial Frontal Gyrus, 17.6% Superior Frontal Gyrus, 12.2% Cingulate
Gyrus
    Cell Type: 82.4% Brodmann area 6, 9.9% Brodmann area 24, 7.6% Brodmann area 32

 #3: 2280 mm^3 from (-58,4,-8) to (-42,16,18) centered at (-51.8,9.7,7.2) with 3 peaks
    with a max value of 0.0126 ALE, 6.367327E-5 P, 3.83 Z at (-54,8,6)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 71.3% Frontal Lobe, 26.7% Temporal Lobe, 2% Sub-lobar
    Gyrus: 46.5% Precentral Gyrus, 26.7% Superior Temporal Gyrus, 24.8% Inferior Frontal
Gyrus, 2% Insula
    Cell Type: 67.3% Brodmann area 44, 24.8% Brodmann area 22, 2% Brodmann area 6,
2% Brodmann area 13, 2% Brodmann area 45

 #4: 1864 mm^3 from (20,-66,-32) to (34,-50,-18) centered at (27.9,-58.1,-25.5) with 1
peaks
    with a max value of 0.0171 ALE, 1.3543103E-6 P, 4.69 Z at (28,-58,-26)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebellum
    Lobe: 57.9% Anterior Lobe, 42.1% Posterior Lobe
    Gyrus: 42.1% Culmen, 18.5% Declive, 9% Pyramis, 7.3% Tuber, 3.9% Uvula, 3.4%
Cerebellar Tonsil
    null


Experiment Table:
[ 1 2 0 1 0 0 1 1 2 1 1 1 1 2 ]
[ 0 1 0 0 1 0 1 1 1 0 1 0 0 3 ]
[ 1 1 1 0 1 0 0 0 1 0 0 0 0 1 ]
[ 0 0 0 0 1 0 0 1 1 0 0 0 0 2 ]

Contributors to cluster #1
1 foci from Lurito,2000: Rhyming > Baseline ,Increases
2 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Gulyas,2001: Letter Listing > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
2 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from MacSweeney, 2009: Rhyme Judgement > Picture Judgement ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Booth,2003: Rhyme Judgement > Control ,Increases
1 foci from Aparacio,2007: Rhyming > Baseline ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
```

Contributors to cluster #2
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases

Contributors to cluster #3
1 foci from Lurito,2000: Rhyming > Baseline ,Increases
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Theys,2020: Reading > Symbols, Increases

Contributors to cluster #4
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases

Appendix 5 - Output of GingerALE in Chapter 2: High Intentionality - Balanced

```
Mask:
Reference Space            = Talairach
Dimensions                 = 80x96x70
Number of within-brain voxels = 198111
Mask Size                  = More Conservative (Smaller)

Foci:
Coordinate System           = Talairach
File of foci coordinates     = High_Intent.txt
Number of foci              = 93
Number of experiments        = 8
Total number of subjects     = 147

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = High_Intent_ALE.nii
FWHM minimum value           = 8.95555138448401
FWHM median value           = 9.437333897527274
FWHM maximum value           = 10.362276299774612
Minimum ALE score           = 1.0186681E-32
Maximum ALE score            = 0.021438507

P Values: Eickhoff (HBM, 2009)
File of P values            = High_Intent_P.nii
Minimum P value              = 1.0944036E-8

Thresholding:
Threshold Method            = Cluster-level Inference
Thresholding Value          = 0.05
Thresholding Permutations    = 1000
Cluster-Forming Method       = Uncorrected P value
Cluster-Forming Value        = 0.01
Volume > Threshold          = 6776 mm^3
Chosen min. cluster size     = 1920 mm^3
Thresholded ALE image        = High_Intent_C05_1k_ALE.nii

Cluster Analysis:
 #1: 3712 mm^3 from (-12,-10,44) to (6,10,62) centered at (-3.6,0.2,54.9) with 1 peaks
    with a max value of 0.0214 ALE, 1.0944036E-8 P, 5.6 Z at (-6,-2,56)
   Labels: (Gray Matter only)
    Hemisphere: 81.8% Left Cerebrum, 18.2% Right Cerebrum
    Lobe: 93.1% Frontal Lobe, 6.9% Limbic Lobe
    Gyrus: 78.8% Medial Frontal Gyrus, 14.3% Superior Frontal Gyrus, 6.9% Cingulate
Gyrus
    Cell Type: 90% Brodmann area 6, 6.9% Brodmann area 24, 3% Brodmann area 32

 #2: 3064 mm^3 from (-54,-14,26) to (-36,6,56) centered at (-44.9,-2.3,37.6) with 2 peaks
```

with a max value of 0.0135 ALE, 1.6694084E-5 P, 4.15 Z at (-42,-2,36)
Labels: (Gray Matter only)
Hemisphere: 100% Left Cerebrum
Lobe: 98.4% Frontal Lobe, 1.6% Parietal Lobe
Gyrus: 80.1% Precentral Gyrus, 12% Inferior Frontal Gyrus, 6.3% Middle Frontal Gyrus, 1.6% Postcentral Gyrus
Cell Type: 75.4% Brodmann area 6, 13.1% Brodmann area 4, 9.9% Brodmann area 9, 1.6% Brodmann area 3

Experiment Table:
[1 0 0 1 1 1 1 3]
[1 0 0 1 1 1 1 2]

Contributors to cluster #1
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases

Contributors to cluster #2
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases

Appendix 6 - Output of GingerALE in Chapter 2: High Egocentricity - Unbalanced

```
Mask:
Reference Space             = Talairach
Dimensions                  = 80x96x70
Number of within-brain voxels = 198111
Mask Size                   = More Conservative (Smaller)

Foci:
Coordinate System            = Talairach
File of foci coordinates      = High_Ego.txt
Number of foci               = 152
Number of experiments         = 16
Total number of subjects      = 281

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels           = High_Ego_ALE.nii
FWHM minimum value            = 8.746152498730366
FWHM median value            = 9.755397007122182
FWHM maximum value            = 11.37334466796002
Minimum ALE score            = 9.189723E-28
Maximum ALE score             = 0.02848524

P Values: Eickhoff (HBM, 2009)
File of P values             = High_Ego_P.nii
Minimum P value               = 2.3753938E-10

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value           = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value         = 0.01
Volume > Threshold           = 17912 mm^3
Chosen min. cluster size      = 1544 mm^3
Thresholded ALE image         = High_Ego_C05_1k_ALE.nii

Cluster Analysis:
 #1: 6128 mm^3 from (-54,-14,22) to (-36,22,56) centered at (-45.2,1.3,35.9) with 3 peaks
    with a max value of 0.0182 ALE, 1.6318077E-6 P, 4.65 Z at (-44,2,30)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 94.6% Frontal Lobe, 5.4% Parietal Lobe
    Gyrus: 59.9% Precentral Gyrus, 17.4% Inferior Frontal Gyrus, 17.4% Middle Frontal
Gyrus, 5.4% Postcentral Gyrus
    Cell Type: 46.4% Brodmann area 6, 34.1% Brodmann area 9, 14.1% Brodmann area 4,
5.4% Brodmann area 3
```

```
 #2: 5376 mm^3 from (-14,-10,34) to (6,12,64) centered at (-4.6,1.8,53) with 1 peaks
    with a max value of 0.0285 ALE, 2.3753938E-10 P, 6.23 Z at (-4,0,56)
  Labels: (Gray Matter only)
    Hemisphere: 88% Left Cerebrum, 12% Right Cerebrum
    Lobe: 78.9% Frontal Lobe, 21.1% Limbic Lobe
    Gyrus: 60.4% Medial Frontal Gyrus, 21.1% Cingulate Gyrus, 18.5% Superior Frontal
Gyrus
    Cell Type: 72.1% Brodmann area 6, 14.7% Brodmann area 24, 13.2% Brodmann area
32

 #3: 2536 mm^3 from (-58,4,-8) to (-42,16,18) centered at (-50.7,9.4,7.7) with 2 peaks
    with a max value of 0.0176 ALE, 2.7436797E-6 P, 4.55 Z at (-50,8,8)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 73.1% Frontal Lobe, 22.1% Temporal Lobe, 4.8% Sub-lobar
    Gyrus: 47.1% Precentral Gyrus, 26% Inferior Frontal Gyrus, 22.1% Superior Temporal
Gyrus, 4.8% Insula
    Cell Type: 70.2% Brodmann area 44, 20.2% Brodmann area 22, 4.8% Brodmann area
13, 1.9% Brodmann area 45

 #4: 2064 mm^3 from (22,-66,-32) to (36,-48,-18) centered at (28.9,-56.9,-25.7) with 1
peaks
    with a max value of 0.0213 ALE, 1.3274675E-7 P, 5.15 Z at (30,-56,-26)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebellum
    Lobe: 64.7% Anterior Lobe, 35.3% Posterior Lobe
    Gyrus: 47.7% Culmen, 14% Declive, 7.4% Pyramis, 7.4% Tuber, 4.3% Cerebellar Tonsil,
2.3% Uvula
    null

 #5: 1808 mm^3 from (26,10,-2) to (40,24,16) centered at (33.3,17.4,5.4) with 1 peaks
    with a max value of 0.016 ALE, 9.283292E-6 P, 4.28 Z at (34,16,4)
  Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebrum
    Lobe: 93.6% Sub-lobar, 6.4% Frontal Lobe
    Gyrus: 79.5% Insula, 11.5% Claustrum, 7.7% Inferior Frontal Gyrus, 1.3% Extra-Nuclear
    Cell Type: 70.5% Brodmann area 13, 5.1% Brodmann area 47, 3.8% Brodmann area 45


Experiment Table:
[ 1 2 0 1 0 0 1 0 1 2 1 1 1 1 2 1 ]
[ 0 1 0 0 1 0 1 2 1 1 0 1 0 0 3 2 ]
[ 1 1 1 0 1 0 0 0 0 1 0 0 0 0 1 1 ]
[ 0 0 0 0 1 0 0 0 1 1 0 0 0 0 2 1 ]
[ 0 0 1 0 0 0 0 0 1 1 0 1 0 0 0 1 ]

Contributors to cluster #1
1 foci from Lurito,2000: Rhyming > Baseline ,Increases
```

2 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Gulyas,2001: Letter Listing > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
2 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from MacSweeney, 2009: Rhyme Judgement > Picture Judgement ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Booth,2003: Rhyme Judgement > Control ,Increases
1 foci from Aparacio,2007: Rhyming > Baseline ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #2
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
2 foci from Raij,2017: Verbal Thought > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases
2 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #3
1 foci from Lurito,2000: Rhyming > Baseline ,Increases
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #4
1 foci from Rudner,2004: Rhyme Judgement > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #5
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Appendix 7 - Output of GingerALE in Chapter 2: High Egocentricity - Balanced

```
Mask:
Reference Space           = Talairach
Dimensions                = 80x96x70
Number of within-brain voxels = 198111
Mask Size                 = More Conservative (Smaller)

Foci:
Coordinate System         = Talairach
File of foci coordinates     = High_Ego Balanced.txt
Number of foci            = 117
Number of experiments        = 10
Total number of subjects     = 222

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = High_Ego Balanced_ALE.nii
FWHM minimum value            = 8.746152498730366
FWHM median value            = 9.203418613972946
FWHM maximum value            = 10.362276299774612
Minimum ALE score            = 1.0186681E-32
Maximum ALE score             = 0.028442714

P Values: Eickhoff (HBM, 2009)
File of P values            = High_Ego Balanced_P.nii
Minimum P value              = 1.0499338E-10

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value           = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value         = 0.01
Volume > Threshold            = 14280 mm^3
Chosen min. cluster size      = 1688 mm^3
Thresholded ALE image         = High_Ego Balanced_C05_1k_ALE.nii

Cluster Analysis:
 #1: 5208 mm^3 from (-14,-18,34) to (6,12,64) centered at (-4.8,0.5,54.4) with 3 peaks
    with a max value of 0.0284 ALE, 1.0499338E-10 P, 6.35 Z at (-4,0,56)
  Labels: (Gray Matter only)
    Hemisphere: 86.9% Left Cerebrum, 13.1% Right Cerebrum
    Lobe: 84.7% Frontal Lobe, 15.3% Limbic Lobe
    Gyrus: 63.6% Medial Frontal Gyrus, 19.9% Superior Frontal Gyrus, 15.3% Cingulate
Gyrus, 1.2% Precentral Gyrus
    Cell Type: 78.6% Brodmann area 6, 10.7% Brodmann area 24, 10.7% Brodmann area
32
```

```
 #2: 3192 mm^3 from (-54,-14,26) to (-38,4,56) centered at (-45.2,-4.1,40.1) with 2 peaks
     with a max value of 0.0156 ALE, 8.280594E-6 P, 4.31 Z at (-46,-8,48)
   Labels: (Gray Matter only)
     Hemisphere: 100% Left Cerebrum
     Lobe: 92.9% Frontal Lobe, 7.1% Parietal Lobe
     Gyrus: 79.7% Precentral Gyrus, 9.1% Inferior Frontal Gyrus, 7.1% Postcentral Gyrus,
4.1% Middle Frontal Gyrus
     Cell Type: 65.5% Brodmann area 6, 20.3% Brodmann area 4, 7.1% Brodmann area 9,
7.1% Brodmann area 3

 #3: 2072 mm^3 from (-58,4,-2) to (-42,16,18) centered at (-49.8,8.9,8.7) with 1 peaks
     with a max value of 0.0161 ALE, 5.3986123E-6 P, 4.4 Z at (-50,8,8)
   Labels: (Gray Matter only)
     Hemisphere: 100% Left Cerebrum
     Lobe: 74.3% Frontal Lobe, 17.6% Temporal Lobe, 8.1% Sub-lobar
     Gyrus: 52.7% Precentral Gyrus, 21.6% Inferior Frontal Gyrus, 17.6% Superior Temporal
Gyrus, 8.1% Insula
     Cell Type: 70.3% Brodmann area 44, 16.2% Brodmann area 22, 8.1% Brodmann area
13, 2.7% Brodmann area 45, 1.4% Brodmann area 6

 #4: 1984 mm^3 from (26,10,-2) to (40,26,16) centered at (33.3,17.5,5.6) with 1 peaks
     with a max value of 0.016 ALE, 5.4430407E-6 P, 4.4 Z at (34,16,4)
   Labels: (Gray Matter only)
     Hemisphere: 100% Right Cerebrum
     Lobe: 94.2% Sub-lobar, 5.8% Frontal Lobe
     Gyrus: 79.1% Insula, 12.8% Claustrum, 7% Inferior Frontal Gyrus, 1.2% Extra-Nuclear
     Cell Type: 70.9% Brodmann area 13, 4.7% Brodmann area 47, 3.5% Brodmann area 45

 #5: 1824 mm^3 from (22,-64,-32) to (36,-48,-22) centered at (29.1,-55.9,-26.5) with 1
peaks
     with a max value of 0.0208 ALE, 1.09221794E-7 P, 5.18 Z at (30,-56,-26)
   Labels: (Gray Matter only)
     Hemisphere: 100% Right Cerebellum
     Lobe: 72.4% Anterior Lobe, 27.6% Posterior Lobe
     Gyrus: 51.3% Culmen, 8.8% Tuber, 8.3% Pyramis, 6.1% Cerebellar Tonsil, 2.6%
Declive, 1.8% Uvula
     null


Experiment Table:
[ 1 0 0 1 1 1 1 3 2 3 ]
[ 1 0 0 1 1 1 1 2 1 0 ]
[ 1 1 0 0 0 1 0 1 1 0 ]
[ 0 1 0 0 1 1 1 0 1 0 ]
[ 0 0 0 0 1 1 0 2 1 0 ]

Contributors to cluster #1
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
```

1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
3 foci from Theys,2020: Reading > Symbols, Increases
2 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases
3 foci from Raij,2017: Verbal Thought > Baseline ,Increases

Contributors to cluster #2
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Okada,2017: Imagine TT > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #3
1 foci from //Papathanassiou, 2000: Verb Generation > Rest ,Increases
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #4
1 foci from Nil,2000: Silent Reading > Baseline ,Increases
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
1 foci from Wilson, 2011: Generate word using probe > Rest ,Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Contributors to cluster #5
1 foci from CurcicBlake, 2012: Stress > Baseline ,Increases
1 foci from Hernandez,2013: Rhyme Judgement > Font Matching ,Increases
2 foci from Theys,2020: Reading > Symbols, Increases
1 foci from Grandchamp,2019: Mind Wandering > Baseline ,Increases

Appendix 8 - Output of GingerALE in Chapter 2: Low Intentionality - Unbalanced

```
Mask:
Reference Space             = Talairach
Dimensions                  = 80x96x70
Number of within-brain voxels = 198111
Mask Size                   = More Conservative (Smaller)

Foci:
Coordinate System           = Talairach
File of foci coordinates    = Low_Intent.txt
Number of foci              = 58
Number of experiments       = 4
Total number of subjects    = 93

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = Low_Intent_ALE.nii
FWHM minimum value          = 8.939044780271026
FWHM median value           = 9.203418613972946
FWHM maximum value          = 9.437333897527274
Minimum ALE score           = 1.5501979E-37
Maximum ALE score           = 0.013614843

P Values: Eickhoff (HBM, 2009)
File of P values            = Low_Intent_P.nii
Minimum P value             = 8.5963065E-6

Thresholding:
Threshold Method            = Cluster-level Inference
Thresholding Value          = 0.05
Thresholding Permutations   = 1000
Cluster-Forming Method      = Uncorrected P value
Cluster-Forming Value       = 0.01
Volume > Threshold          = 1440 mm^3
Chosen min. cluster size    = 1400 mm^3
Thresholded ALE image       = Low_Intent_C05_1k_ALE.nii

Cluster Analysis:
 #1: 1440 mm^3 from (-62,-48,-8) to (-50,-32,14) centered at (-56.1,-42.8,2.9) with 3 peaks
    with a max value of 0.0103 ALE, 9.723458E-5 P, 3.73 Z at (-54,-44,2)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 100% Temporal Lobe
    Gyrus: 74.1% Middle Temporal Gyrus, 25.9% Superior Temporal Gyrus
    Cell Type: 50% Brodmann area 22, 46.3% Brodmann area 21


Experiment Table:
```

[0 0 1 3]

Contributors to cluster #1
1 foci from Alderson-Day, 2020: Direct Quotation > Fixation ,Increases
3 foci from //Yao,2011: Direct Quotations > Baseline ,Increases

Appendix 9 - Output of GingerALE in Chapter 2: Low Intentionality - Balanced

```
Mask:
Reference Space             = Talairach
Dimensions                  = 80x96x70
Number of within-brain voxels = 198111
Mask Size                   = More Conservative (Smaller)

Foci:
Coordinate System           = Talairach
File of foci coordinates     = Low_Intent.txt
Number of foci               = 58
Number of experiments        = 4
Total number of subjects     = 93

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels           = Low_Intent_ALE.nii
FWHM minimum value           = 8.939044780271026
FWHM median value            = 9.203418613972946
FWHM maximum value           = 9.437333897527274
Minimum ALE score            = 1.5501979E-37
Maximum ALE score            = 0.013614843

P Values: Eickhoff (HBM, 2009)
File of P values             = Low_Intent_P.nii
Minimum P value              = 8.5963065E-6

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value           = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method       = Uncorrected P value
Cluster-Forming Value        = 0.01
Volume > Threshold           = 1440 mm^3
Chosen min. cluster size     = 1336 mm^3
Thresholded ALE image        = Low_Intent_C05_1k_ALE.nii

Cluster Analysis:
 #1: 1440 mm^3 from (-62,-48,-8) to (-50,-32,14) centered at (-56.1,-42.8,2.9) with 3 peaks
    with a max value of 0.0103 ALE, 9.723458E-5 P, 3.73 Z at (-54,-44,2)
  Labels: (Gray Matter only)
    Hemisphere: 100% Left Cerebrum
    Lobe: 100% Temporal Lobe
    Gyrus: 74.1% Middle Temporal Gyrus, 25.9% Superior Temporal Gyrus
    Cell Type: 50% Brodmann area 22, 46.3% Brodmann area 21


Experiment Table:
```

[0 0 1 3]

Contributors to cluster #1
1 foci from Alderson-Day, 2020: Direct Quotation > Fixation ,Increases
3 foci from //Yao,2011: Direct Quotations > Baseline ,Increases

Appendix 10 - Output of GingerALE in Chapter 2: Low Egocentricity - Unbalanced

```
Mask:
Reference Space            = Talairach
Dimensions                 = 80x96x70
Number of within-brain voxels = 198111
Mask Size                  = More Conservative (Smaller)

Foci:
Coordinate System           = Talairach
File of foci coordinates     = Low_Ego.txt
Number of foci               = 58
Number of experiments         = 4
Total number of subjects      = 65

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = Low_Ego_ALE.nii
FWHM minimum value           = 9.01191592595652
FWHM median value           = 10.61174813139395
FWHM maximum value           = 10.935527854620133
Minimum ALE score           = 7.696232E-36
Maximum ALE score            = 0.011280302

P Values: Eickhoff (HBM, 2009)
File of P values            = Low_Ego_P.nii
Minimum P value              = 1.6697322E-5

Thresholding:
Threshold Method             = Cluster-level Inference
Thresholding Value          = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method       = Uncorrected P value
Cluster-Forming Value        = 0.01
Volume > Threshold           = 2712 mm^3
Chosen min. cluster size      = 1912 mm^3
Thresholded ALE image        = Low_Ego_C05_1k_ALE.nii

Cluster Analysis:
 #1: 2712 mm^3 from (28,6,-4) to (56,22,14) centered at (42.7,13.4,5.9) with 3 peaks
    with a max value of 0.0113 ALE, 1.6697322E-5 P, 4.15 Z at (44,16,2)
   Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebrum
    Lobe: 50.3% Sub-lobar, 49.7% Frontal Lobe
    Gyrus: 43% Insula, 32.5% Inferior Frontal Gyrus, 17.2% Precentral Gyrus, 7.3%
Claustrum
    Cell Type: 37.7% Brodmann area 13, 32.5% Brodmann area 44, 10.6% Brodmann area
47, 4% Brodmann area 45
```

Experiment Table:
[1 1 0 3]

Contributors to cluster #1
1 foci from //Linden,2010: Imagery > Baseline ,Increases
1 foci from Aleman,2005: Imagery > Baseline ,Increases
3 foci from Grandchamp,2019: Other Voice > Baseline ,Increases

Appendix 11 - Output of GingerALE in Chapter 2: Low Egocentricity - Balanced

```
Mask:
Reference Space           = Talairach
Dimensions                = 80x96x70
Number of within-brain voxels = 198111
Mask Size                 = More Conservative (Smaller)

Foci:
Coordinate System          = Talairach
File of foci coordinates     = Low_Ego.txt
Number of foci             = 58
Number of experiments        = 4
Total number of subjects      = 65

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):
File of ALE voxels          = Low_Ego_ALE.nii
FWHM minimum value            = 9.01191592595652
FWHM median value           = 10.61174813139395
FWHM maximum value            = 10.935527854620133
Minimum ALE score           = 7.696232E-36
Maximum ALE score            = 0.011280302

P Values: Eickhoff (HBM, 2009)
File of P values            = Low_Ego_P.nii
Minimum P value              = 1.6697322E-5

Thresholding:
Threshold Method            = Cluster-level Inference
Thresholding Value          = 0.05
Thresholding Permutations     = 1000
Cluster-Forming Method        = Uncorrected P value
Cluster-Forming Value        = 0.01
Volume > Threshold           = 2712 mm^3
Chosen min. cluster size      = 1912 mm^3
Thresholded ALE image         = Low_Ego_C05_1k_ALE.nii

Cluster Analysis:
 #1: 2712 mm^3 from (28,6,-4) to (56,22,14) centered at (42.7,13.4,5.9) with 3 peaks
    with a max value of 0.0113 ALE, 1.6697322E-5 P, 4.15 Z at (44,16,2)
   Labels: (Gray Matter only)
    Hemisphere: 100% Right Cerebrum
    Lobe: 50.3% Sub-lobar, 49.7% Frontal Lobe
    Gyrus: 43% Insula, 32.5% Inferior Frontal Gyrus, 17.2% Precentral Gyrus, 7.3%
Claustrum
    Cell Type: 37.7% Brodmann area 13, 32.5% Brodmann area 44, 10.6% Brodmann area
47, 4% Brodmann area 45
```

Experiment Table:
[1 1 0 3]

Contributors to cluster #1
1 foci from //Linden,2010: Imagery > Baseline ,Increases
1 foci from Aleman,2005: Imagery > Baseline ,Increases
3 foci from Grandchamp,2019: Other Voice > Baseline ,Increases

Appendix 12 - Adapted Psi-Q used in Chapters 3 Pre-experiment.

Item	Modality	Prompt
1	Sight	Imagine the appearance of a bonfire.
2	Sight	Imagine the appearance of the front door of your house.
1	Auditory	Imagine yourself saying the word “barn”.
2	Auditory	Imagine the word “farm” being said by a familiar person.
3	Auditory	Imagine the melody of a familiar song.
4	Auditory	Imagine the sound of hands clapping in applause.
1	Olfactory	Imagine the smell of burning wood.
2	Olfactory	Imagine the smell of fresh paint.
1	Gustatory	Imagine the taste of sea water.
2	Gustatory	Imagine the taste of toothpaste.
1	Tactile	Imagine touching a soft towel.
2	Tactile	Imagine touching the point of a pin.

Appendix 13 - VVIQ used in Chapters 3 Main experiment.

Item	Subject	Prompt
1	Relative/Friend	Think of the exact contours of face, head, shoulders, and body of a friend or relative you frequently see.
2	Relative/Friend	Visualize their characteristic poses of head and attitudes of the body.
3	Relative/Friend	Picture their precise carriage, length of step, etc., when walking.
4	Relative/Friend	Imagine the different colors worn in some of their familiar clothes.
1	Rising sun	Visualize a rising sun above the horizon into a hazy sky.
2	Rising sun	Picture the sky as it clears and surrounds the sun with blueness.
3	Rising sun	Imagine clouds as a storm blows up with flashes of lightning.
4	Rising sun	Visualize a rainbow appearing in the sky.
1	Shop	Think of the overall appearance of a shop you often visit from the opposite side of the road.
2	Shop	Picture the window display of that shop, including colors, shapes, and details of individual items for sale.
3	Shop	Visualize yourself near the entrance of the shop, focusing on the color, shape, and details of the door.
4	Shop	Imagine entering the shop, going to the counter, where an assistant serves you and money changes hands.
1	Country scene	Visualize the contours of a country scene involving trees, mountains, and a lake.
2	Country scene	Picture the color and shape of the lake in that scene.
3	Country scene	Imagine the color and shape of the trees in that landscape.
4	Country scene	Visualize a strong wind blowing on the trees and on the lake, causing reflections in the water.

Appendix 14 - Stimuli from the scenario rating task used across Chapters 3, 4 & 5.

Item	Condition	Prompt
1	Sound	Imagine the sound of a bag of coins rattling.
2	Sound	Imagine the sound of a bicycle ringing its bell.
3	Sound	Imagine the sound of a boat’s foghorn out at sea.
4	Sound	Imagine the sound of a car horn.
5	Sound	Imagine the sound of a coat being unzipped.
6	Sound	Imagine the sound of a crackling fire.
7	Sound	Imagine the sound of a crowd clapping.
8	Sound	Imagine the sound of a door slamming shut.
9	Sound	Imagine the sound of a fire alarm.
10	Sound	Imagine the sound of a firework bang.
11	Sound	Imagine the sound of a galloping horse.
12	Sound	Imagine the sound of a hammer striking a nail.
13	Sound	Imagine the sound of a kettle boiling.
14	Sound	Imagine the sound of a passing train.
15	Sound	Imagine the sound of a police siren.
16	Sound	Imagine the sound of a printer printing.
17	Sound	Imagine the sound of a raging river.
18	Sound	Imagine the sound of a running shower.
19	Sound	Imagine the sound of a tap-dancer’s shoes on stage.
20	Sound	Imagine the sound of an alarm bell ringing.
21	Sound	Imagine the sound of heavy rain.
22	Sound	Imagine the sound of popcorn popping.
23	Sound	Imagine the sound of thunder after a lightning strike.
24	Sound	Imagine the sound of waves crashing against rocks.
1	Own	Using your own inner voice, say “Commuting in dense cities is hard work”.
2	Own	Using your own inner voice, say “He is an agent for the local police station”.
3	Own	Using your own inner voice, say “I had to give my first speech yesterday”.
4	Own	Using your own inner voice, say “I have a massive problem with that”.
5	Own	Using your own inner voice, say “I have made a list of people to talk to”.
6	Own	Using your own inner voice, say “I have to return my faulty headphones”.
7	Own	Using your own inner voice, say “I made a list of things to bring with me”.
8	Own	Using your own inner voice, say “I never realised it was that big a deal”.
9	Own	Using your own inner voice, say “I will call him first thing in the morning”.
10	Own	Using your own inner voice, say “It should not be too hard to understand”.
11	Own	Using your own inner voice, say “Plenty of tourists stay here over summer”.

12	Own	Using your own inner voice, say “The art gallery was closed on Sunday”.
13	Own	Using your own inner voice, say “The beach was way too crowded”.
14	Own	Using your own inner voice, say “The blue van was by far the cheapest”.
15	Own	Using your own inner voice, say “The checkout queue was too long”.
16	Own	Using your own inner voice, say “The garden was low maintenance”.
17	Own	Using your own inner voice, say “The morning news says a big storm is coming”.
18	Own	Using your own inner voice, say “The score was two-one at half-time”.
19	Own	Using your own inner voice, say “The snow delayed my travel plans”.
20	Own	Using your own inner voice, say “The train leaves in around ten minutes”.
21	Own	Using your own inner voice, say “The tree fell over unexpectedly”.
22	Own	Using your own inner voice, say “There is a good chance it will snow later”.
23	Own	Using your own inner voice, say “They have built dozens of houses downtown”.
24	Own	Using your own inner voice, say “Winter days are generally cold”.
1	Other	Imagine another person saying, “Amsterdam is full of narrow buildings”.
2	Other	Imagine another person saying, “February is the shortest month of the year”.
3	Other	Imagine another person saying, “Going shopping can be quite expensive”.
4	Other	Imagine another person saying, “I have my hair cut every single month”.
5	Other	Imagine another person saying, “I think I want to move to a larger house”.
6	Other	Imagine another person saying, “Internet browsing can be a lot of fun”.
7	Other	Imagine another person saying, “It was a warm sunny day for a picnic”.
8	Other	Imagine another person saying, “Most guitars have six strings”.
9	Other	Imagine another person saying, “Most siblings try to help one another”.
10	Other	Imagine another person saying, “Osaka is Japan’s second biggest city”.
11	Other	Imagine another person saying, “Shopping centres are a great place to shop”.
12	Other	Imagine another person saying, “The campus is very quiet over summer”.
13	Other	Imagine another person saying, “The car was speeding down the road”.
14	Other	Imagine another person saying, “The church has stained glass windows”.
15	Other	Imagine another person saying, “The garden hedge had been freshly cut”.
16	Other	Imagine another person saying, “The hotel room had a great view”.
17	Other	Imagine another person saying, “The house was old but quite clean”.
18	Other	Imagine another person saying, “The neighbours were very generous”.
19	Other	Imagine another person saying, “The translator spoke eight languages”.
20	Other	Imagine another person saying, “The tree fell over unexpectedly”.
21	Other	Imagine another person saying, “The trousers were too large for me”.
22	Other	Imagine another person saying, “The university course lasted three years”.
23	Other	Imagine another person saying, “There was a huge parade yesterday”.
24	Other	Imagine another person saying, “Travelling has become non-existent this year”.

Appendix 15 - Stimuli from spontaneous imagery task used in Chapters 3, 4 & 5.

Item	Condition	Prompt
1	Own	Close your eyes and relax for 3 minutes. During this time, press the spacebar if you ‘hear’ your own inner voice. You may press it multiple times.
1	Other	Close your eyes and relax for 3 minutes. During this time, you are free to think of anything you would like as long as you do not think of the movie quote. If you do ‘hear’ the movie quote in your head, press the spacebar. You may press it multiple times.
1	Visual	Close your eyes and relax for 3 minutes. During this time, you are free to think of anything you would like as long as you do not think of a white bear. Every time a white bear comes to mind, press the spacebar. You may press it multiple times.

Appendix 16 - R Code used to perform statistical analysis on scenario rating task data.

```
# Title and Output type
title: "R Notebook"
output: html_notebook

# ----------------------------
# STEP 1: SETUP AND LIBRARY LOADING
# ----------------------------

# Clear the R environment of all variables
rm(list = ls())

# Load necessary libraries for data manipulation, modeling, and visualization
library(tidyverse)
library(lme4)
library(lmerTest)
library(fitdistrplus)
library(emmeans)
library(ggplot2)
library(ggpubr)
library(ordinal)
library(stats)
library(rstatix)

# Set the root directory to the directory of the current R script or R Notebook
knitr::opts_knit$set(root.dir = dirname(rstudioapi::getSourceEditorContext()$path))

# ----------------------------
# STEP 2: DATA LOADING AND INITIAL PROCESSING
# ----------------------------

# Load data from the specified path
D  <- read.csv("path/to/data", header=T, sep=",")

# Summarize the number of missing values for each column to ensure data quality
D %>% summarise(across(where(~any(is.na(.))), ~sum(is.na(.))))

# Convert the 'Response.Time' column to numeric for proper processing and analysis
D$Response.Time <- as.numeric(D$Response.Time)

# ----------------------------
# STEP 3: OUTLIER HANDLING
# ----------------------------

# Filter out outliers based on median and MAD method, grouped by 'scenario_type' and
'suppression_type'
```

```
D2 = D %>%
  group_by(scenario_type, suppression_type) %>%
  filter(between(Reaction.Time,
              median(Reaction.Time, na.rm=TRUE) - (2.5 * mad(Reaction.Time)),
              median(Reaction.Time, na.rm=TRUE) + (2.5 * mad(Reaction.Time)))) %>%
  ungroup()

# Calculate the proportion of data retained after outlier removal
1-nrow(D2)/nrow(D)

# Replace the original dataset with the filtered one
D = D2

# ----------------------------
# STEP 4: DATA TRANSFORMATION AND SUMMARIZATION
# ----------------------------

# Visualize the distribution of Reaction Time post outlier removal
hist(D$Reaction.Time, 100)

# Convert 'Response' and 'Reaction.Time' columns to numeric for subsequent analysis
D$response <- as.numeric(D$Response)
D$response_time <- as.numeric(D$Reaction.Time)

# Summarize data by 'Participant.Private.ID', 'scenario_type', and 'suppression_type'
by_cyl <- D %>%
  group_by(Participant.Private.ID, scenario_type, suppression_type) %>%
  summarise(mean_response = mean(response), mean_rt = mean(response_time))

# ----------------------------
# STEP 5: HYPOTHESIS TESTING
# ----------------------------

# Conduct an ANOVA to test the interactive effects of 'suppression_type' and
'scenario_type' on the mean response
model <- aov(mean_response ~ suppression_type * scenario_type, data = by_cyl)

# Display the summary of the ANOVA model to interpret the statistical significance and
effects
summary(model)
```

Appendix 17 - R Code used to perform statistical analysis on spontaneous imagery task data.

```
# Title and Output type
title: "R Notebook"
output: html_notebook

# ----------------------------
# STEP 1: SETUP AND LIBRARY LOADING
# ----------------------------

# Clear the R environment of all variables
rm(list = ls())

# Load necessary libraries
library(tidyverse)
library(lme4)
library(lmerTest)
library(fitdistrplus)
library(emmeans)
library(ggplot2)
library(ggpubr)
library(ordinal)
library(stats)

# Set the root directory to the directory of the current R script or R Notebook
knitr::opts_knit$set(root.dir = dirname(rstudioapi::getSourceEditorContext()$path))

# ----------------------------
# STEP 2: DATA LOADING AND INITIAL PROCESSING
# ----------------------------

# Load data from the specified path
D  <- read.csv("path/to/data", header=T, sep=",")

# Summarize the number of missing values for each column
D %>% summarise(across(where(~any(is.na(.))), ~sum(is.na(.))))

# Convert the response_time column to numeric for proper processing
D$response_time <- as.numeric(D$response_time)

# Plot a histogram of response times to visualize its distribution
hist(D$response_time, 100)

# ----------------------------
# STEP 3: OUTLIER HANDLING AND DATA GROUPING
# ----------------------------
```

```
# Filter outliers based on median and MAD, and group by 'type' and 'condition'
d2 = D %>%
  group_by(type, condition) %>%
  filter(between(response_time,
                 median(response_time, na.rm=TRUE) - (2.5 * mad(response_time)),
                 median(response_time, na.rm=TRUE) + (2.5 * mad(response_time)))) %>%
  ungroup()

# Calculate the proportion of data retained after outlier removal
1-nrow(d2)/nrow(D)

# Assign the filtered dataset back to D
D = d2

# Group D by 'type' and 'condition' for subsequent analyses
D <- D %>%
  group_by(type, condition)

# -----------------------------
# STEP 4: DATA SUMMARIZATION
# -----------------------------

# Summarize response, response_time, and cs by subject, type, and condition
by_cyl <- D %>%
  group_by(subject_nr, type, condition) %>%
  summarise(mean_response = mean(response), mean_rt = mean(response_time),
mean_cs = mean(cs))

# -----------------------------
# STEP 5: DATA TRANSFORMATION AND SPECIFICATION
# -----------------------------

# Convert subject_nr and type to factors for proper statistical analysis
D$subject_nr = as.factor(D$subject_nr)
D$Condition = as.factor(D$type)

# Specify order for the response variable and convert to factor
D$response.f = factor(D$response, levels = c("1","2","3","4","5"), order = TRUE)

# Specify levels for Suppression_Type
D$Suppression_Type <- factor(D$condition, levels = c("suppression_1", "suppression_2",
"suppression_3"))

# Display the structure and summary of the data to verify transformations and understand
data characteristics
str(D)
summary(D)
```

```
# -----------------------------
# STEP 6: HYPOTHESIS TESTING
# -----------------------------

# Conduct an ANOVA to test the effects of Suppression_Type and Condition on the response
model <- aov(response ~ Suppression_Type*Condition, data = by_cyl)

# Display the summary of the ANOVA model to understand statistical significance and effects
summary(model)
```