Bridging phenomenology and neural mechanisms of inner speech: ALE meta-analysis on egocentricity and spontaneity in a dual-mechanistic framework

Jaydan Pratts [a], Gorana Pobric [a], Bo Yao [a, b, *]

[a] *Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, University of Manchester, UK*
[b] *Department of Psychology, Fylde College, Lancaster University, UK*

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.

1. Introduction

Inner speech is an internal, speech-like experience without the presence of an external sound (Alderson-Day and 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 and Barber, 2011; Yao and Scheepers, 2011; Yao et al., 2011), behavioural self-regulation (Diaz et al., 2014), as well as task switching and goal tracking (Emerson and 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 a unifying theoretical framework. This framework is examined and verified through an Activation Likelihood Estimation (ALE) meta-analysis of the existing literature.

1.1. Two mechanistic models of inner speech

1.1.1. 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 and Poeppel (2010) show that kinaesthetic estimation of articulatory imagery is followed by increased auditory neural activity ~ 170 ms 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 posits that inner speech arises from articulatory signals, empirical evidence shows that overt articulation of irrelevant speech (articulatory suppression) does not fully suppress inner speech (Wheeldon and Levelt, 1995). This implies that inner speech may not be driven entirely by articulation. Moreover, brain areas associated with speech production are not always activated during inner speech, especially when it occurs spontaneously (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 and Scheepers, 2011; Yao et al., 2011, 2021).

This mixed evidence also highlights two unresolved questions for the corollary discharge model. First, the speed of inner speech is often 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. Corollary discharges are derived from efference copies that match and monitor our speech output. This matching process is both time-sensitive and content-specific (Jack et al., 2019), leading to perceptual attenuation of our own speech (Ford and Mathalon, 2005). If corollary discharges unfold faster than overt speech, this mismatch in timing could lead to errors in attributing the speech as our own. Consequently, faster inner speech is less likely to fully rely on corollary discharges. Instead, it could take a more abstract format by losing its audible-speech qualities, such that it does not need to adhere to the timescale afforded by fully-fledged corollary discharges. Alternatively, it could be condensed into incomplete fragments, creating the illusion of being faster than fully articulated overt speech. The reduced reliance on corollary discharges calls for alternative, complementary mechanisms to account for the perceptual qualities of inner speech at this faster timescale.

Second, inner speech often incorporates voices of others which contain vocal features distinct from one's own (McCarthy-Jones and Fernyhough, 2011; Alderson-Day et al., 2018). These 'foreign' vocal features cannot be provided solely by corollary discharges as they are physically constrained by one's own articulator. While our articulator offers some flexibility, e.g., in raising or lowering pitch to mimic how other people talk, it has limitations in generating voices that we cannot produce accurately out loud, such as those of Darth Vader or a dragon from The Hobbit. To attribute inner speech to other people, one needs to recognise that the voices they experience cannot be explained by their own corollary discharges. This mismatch suggests the need for additional mechanisms to create perceptual features that are not predicted by one's corollary discharges.

1.1.2. 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 concept of novel perceptual experiences being generated using a finite number of stored memories is not limited to the embodied cognitive literature. Kosslyn et al. (2001) provide one early description of this mechanism in their examination of the neural foundations of mental imagery, where they describe a system in which mental images may be "created by combining and modifying stored perceptual information in novel ways". This mechanism whereby imagery is a simulation of perception occurring within sensory cortices continues to represent a dominant explanation of how visual imagery is generated (Dijkstra et al., 2019), and has been extended to the domain of inner speech where it is argued that memories of speech may be combined and reactivated such that an individual can generate novel inner speech experiences in the absence of motor activation (Carruthers, 2018). Initial support for such a mechanism supporting inner speech can be found in Tian et al. (2016) who evidence a memory-based route for the mental imagery of speech.

The perceptual simulation model of inner speech neatly complements the corollary discharge model. First, perceptual simulations of speech do not depend on speech production, which could explain the limited effectiveness of articulatory suppression on inner speech (Wheeldon and Levelt, 1995) and 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). Second, 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 and Dell, 2010). Third, 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.

Nevertheless, it is worth noting that the perceptual simulation model has its own limitations. Theoretically, it remains underspecified how memories of speech fragments are combined in ways that adhere to grammatical rules of our language (e.g., Carruthers, 2018). Empirically, 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. Moreover, 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 and 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.

1.2. 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 and Ferny-hough, 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).

1.3. 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, as well as the number of speakers within that speech. At one extreme inner speech can take the form of a monologue, consisting of a single voice speaking in a manner reminiscent of a soliloquy. At the opposite extreme, inner speech represents a dialogue involving the voice of the speaker and the imagined voices of others. *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 and Riekki, 2017; Yao et al., 2011). One attempt to explain these findings through the lens of the ConDialInt model is to view the involvement of the articulators, and speech production regions, as a dynamic process which may be modulated based on task demands. This is akin to Oppenheim and Dell's (2010) concept of flexible abstractness, whereby the involvement of speech articulators in inner speech may be increased or decreased depending on the need for phonological detail brought about by situational demands. In this vein it may be argued that the lack of observable involvement of speech production regions in some forms of inner speech indicates a low level of activation within the corollary discharge system, rather than the involvement of a separate mechanism of generation. While this flexible implementation of the ConDialInt model offers a plausible explanation of some of the divergent findings within inner speech neuroimaging literature, it offers a limited account for the lack of suppression of inner speech through concurrent articulation (Wheeldon and Levelt, 1995), and for the speed difference between inner and overt speech (Mackey, 1981; Netsell et al., 2016). Crucially, it does not explain the unique vocal features of inner speech that cannot be produced by one's own articulators. Perceptual simulation, a mechanism used in various types of mental imagery (e.g. visual imagery; Pearson et al., 2015), can help explain these gaps, particularly for inner speech in voices other than our own.

The ConDialInt model's dimension of *dialogality* raises further questions which require consideration. First, by categorising inner speech by its dialogality in a manner that ranges from a single voice to multiple distinct voices, 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 Grandchamp et al. (2019) concept of 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.

Additional questions arise regarding the condensation dimension of the ConDialInt model. Grandchamp et al. (2019) define condensed inner speech as that which "only involves the highest linguistic level (semantics), and has lost most of the acoustic, phonological, and even syntactic qualities of overt speech". Conceptually, the lack of these qualities implies that tasks requiring deliberate use of these qualities, like short-term memory rehearsal or phonological judgements, are unlikely to involve condensed inner speech. Rather, condensed inner speech seems to be inherently spontaneous. This raises questions whether the dimension of condensation is orthogonal to that of deliberateness-spontaneity. Empirically, due to its spontaneous nature, condensed inner speech is difficult to manipulate or measure. Although current neuroimaging techniques can in principle capture brain activity linked to semantic processing, such as in the left anterior temporal lobe (Visser et al., 2010), it remains challenging to attribute this neural activity to condensed inner speech specifically, as this activity can also be generated by other non-verbal semantic processing. Given the lack of empirical data on condensation, it is not possible to conduct a meta-analysis on this dimension, or to articulate how condensed inner speech may interact with known neurocognitive mechanisms related to corollary discharges and/or perceptual simulations. Therefore, the exclusion of condensation from our framework is not a critique of its existence or its inclusion in Grandchamp et al.'s model, but a reflection of practical concerns in the current meta-analysis. This does not preclude further adjustments to our framework as condensation becomes better understood and more empirically substantiated. As it currently stands, however, this dimension remains challenging to investigate or analyse.

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*. Rather than classifying inner speech by whether it is a monologue, dialogue, or the number of speakers represented within inner speech, the dimension of egocentricity measures the extent to which inner speech is a recreation of one's own voice (high egocentricity), or the voice of another individual (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 replaced by *spontaneity*. This renaming serves primarily to avoid the implication that spontaneous inner speech may not have an intent. For example, an individual might evaluate his plans for the day in a spontaneous manner, which demonstrates intent whilst also being spontaneous. It's crucial to note that 'deliberateness' and 'spontaneity' function on a continuum in our framework. The continuum ranges from highly deliberate inner speech, elicited by explicit task demands (e.g., rhyme judgements), to highly spontaneous inner speech, which emerges in the absence of cues or clear task demands (e.g., mind wandering). Between these extremes, other forms of inner speech exist but are less studied. These include, for example, rehearsing a conversation before it happens, engaging in internal monologue during problem-solving tasks like the Tower of Hanoi, or spontaneously generating musical lyrics. Our framework is conceptualised to accommodate these diverse types of inner speech along the continuum of deliberateness and spontaneity.

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 × spontaneity space (Fig. 1). That is, the more egocentric and deliberate inner speech is (i.e. more northwestward in Fig. 1), the more strongly it relies on articulation-derived corollary discharge. For example, tasks involving explicit phonological judgements are likely to elicit egocentric and deliberate inner speech and to activate speech production areas (Aleman et al., 2005; Lurito et al., 2000). In contrast, the less egocentric or deliberate inner speech is (i.e. more southeastward in Fig. 1), the more likely it resorts to perceptual simulation. For example, imagery of another person's voice (low egocentricity) and inner speech that
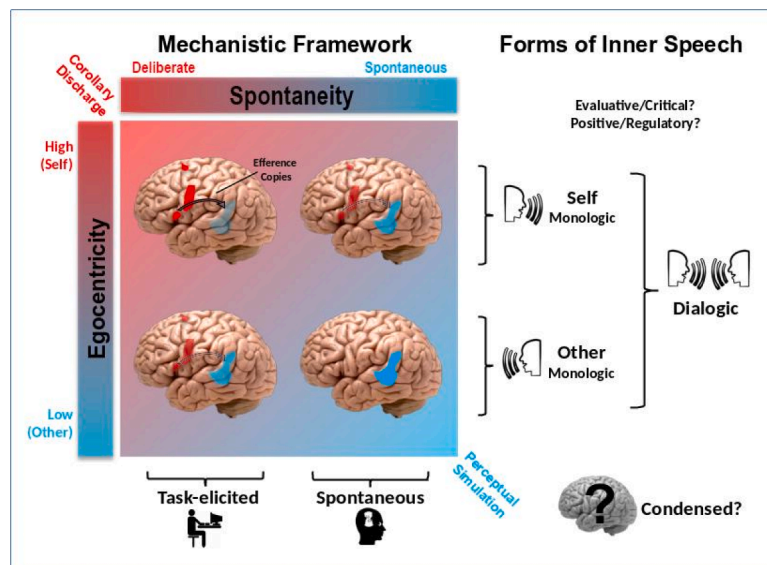

Fig. 1. Schematic illustration of the egocentricity × spontaneity framework, with its underpinning neural mechanisms and its hypothetical mapping with phenomenological qualities.

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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

emerges spontaneously 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 and 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).

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 the types of dialogic inner speech which include multiple speakers / voices. While the dimensions of *dialogality* 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 spontaneity dimensions. However, the dimension of *condensation* is not explicitly considered in the current framework as it likely correlates with egocentricity and spontaneity and cannot be easily manipulated or objectively observed.

1.4. 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 spontaneity. 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).

While the current state of neuroimaging literature is incomplete and focuses on specific forms such as those elicited by tasks (e.g., rhyme judgements), mind wandering, and silent reading, our meta-analysis relies on these well-studied forms to test our framework's predictions. The caveat lies in cautious generalisation to other types of inner speech. Less studied forms, such as internal dialogues incorporating voices of others, or those that occur in problem-solving tasks, are underexplored in the neuroimaging literature. The present framework aims to address this gap by conceptualising continuous dimensions of egocentricity and spontaneity in inner speech and serves as a generalised framework for further empirical validation and exploration of these under-studied forms.

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 when higher in egocentricity and/or more deliberate, and rely more on the perceptual simulation mechanism as egocentricity decreases and/or spontaneity increases. In an ALE analysis, we predicted that inner speech which was deliberate and high in egocentricity 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

and 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 and Guenther, 2011). Inner speech which is low in egocentricity and/or spontaneous 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).

2. Materials and methods

2.1. 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 (see Fig. 2 for a visualisation). The final studies are presented in Table 1.

2.2. Eligibility criteria

2.2.1. 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 spontaneity *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 spontaneity, studies in which participants were required, explicitly or implicitly, to generate inner speech were classified as deliberate inner speech 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 spontaneous inner speech studies. 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 spontaneous inner speech but adopted a different approach. Participants were asked to report their internal state in the moments preceding the sounding of
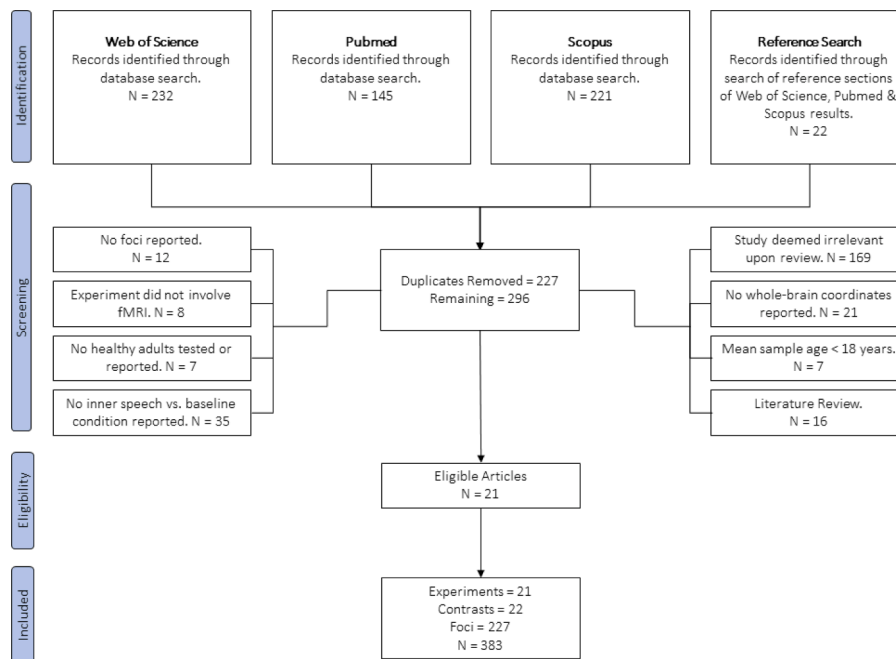
Fig. 2. Flow diagram of literature review process. Notation boxes in the *screening* section represent the various reasons for study exclusion and the number of studies excluded.

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 spontaneity yielded 14 studies in the deliberate inner speech group and 4 studies in the spontaneous inner speech group. Given that no studies were found which examined inner speech which was both spontaneous and low in 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 A1). 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.

2.3. 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 spontaneity 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 $p < 0.01$ is appropriate for smaller sample sizes, and has been adopted by previous ALE studies (Di et al., 2017; Falcone and Jerram, 2018; Ruiz Vargas et al., 2016).

Given the low number of studies in the low egocentricity ($N = 4$) and spontaneous ($N = 4$) conditions, both in absolute terms and relative to their high egocentricity/deliberate 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.

2.4. 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

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

Study	N	Foci	Task	Baseline	Contrast	Handedness	Spontaneity	Egocentricity
Lurito et al. (2000)	5	19	Rhyme judgement	Fixation symbol	Rhyme judgement > Fixation symbol	Right	Deliberate	High
Booth et al. (2003)	12	4	Rhyme judgement	Symbol matching	Rhyme judgement > Symbol matching	Right	Deliberate	High
Rudner et al. (2005)	12	16	Rhyme judgement	Rest	Rhyme judgement > Rest	Right	Deliberate	High
Aparacio et al. (2007)	12	2	Rhyme judgement	String matching	Rhyme judgement > String matching	Right	Deliberate	High
MacSweeney et al. (2009)	7	6	Rhyme judgement	Same picture judgement	Rhyme judgement > Same picture judgement	Right	Deliberate	High
Ćurčić-Blake et al. (2013)	31	9	Metrical stress judgement	Fixation symbol	Metrical stress judgement > Fixation symbol	Right	Deliberate	High
Hernandez et al. (2013)	16	11	Rhyme judgement	Font matching	Rhyme judgement > Font matching	Right	Deliberate	High
Aleman et al. (2005)	6	14	Imagine other person reading word and determine metrical stress	Fixation symbol	Imagine voice > Fixation symbol	Right	Deliberate	Low
Linden et al. (2011)	7	13	Imagine voice of familiar person	Rest	Imagine voice > Rest	Right	Deliberate	Low
Grandchamp et al. (2019)	24	24	Imagine other voice	N.S	Voice imagery > N.S*	Right	Deliberate	Low
Kleider-Offutt et al. (2019)	28	9	Read sentence in a learned voice	Rest	Read sentence > Rest	Not specified	Deliberate	Low
De Nil et al. (2000)	10	5	Mentally read presented 3-syllable words	Fixation symbol	Read words > Fixation symbol	Right	Deliberate	High
Gulyás (2001)	10	1	Internally list alphabet & first verse of national anthem	Rest	Internal listing > Rest	7 Right, 2 Left, 1 Ambidextrous	Deliberate	High
Theys et al. (2020)	11	16	Read visually presented pseudoword	View character string	Silent reading > View character string	Right	Deliberate	High
Papathanassiou et al. (2000)	8	19	Generate verbs related to given noun	Rest	Generate verb > Rest	Right	Deliberate	High
Wilson et al. (2011)	26	10	Generate words beginning with given letter	Rest	Word generation > Rest	Right	Deliberate	High
Alderson-Day et al. (2020)	21	10	Direct quotation speech	Fixation symbol	Direct quotation > Fixation symbol	Right	Spontaneous	High
Yao et al. (2011)	16	16	Direct quotation speech	Fixation symbol	Direct quotation > Fixation symbol	15 Right, 1 Left	Spontaneous	High
Raij et al. (2017)	51	8	Verbal thought (Direct Experience Sampling)	Fixation symbol	Verbal thought > Fixation symbol	Not specified	Spontaneous	High
Grandchamp et al. (2019)	24	16	Use joystick to report verbal mind wandering	N.S	Verbal Mind Wandering > N.S*	Right	Spontaneous	High
Okada et al. (2018a)	21	7	Silently read the visually presented tongue twister	Fixation symbol	Tongue twister > Fixation symbol	Right	Deliberate	High
Okada et al. (2018b)	21	13	Silently articulate the visually presented tongue twister	Rest	Tongue twister > Rest	Right	Deliberate	High

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 spontaneity 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 %), deliberate (Nnull = 4; 28.6 %), spontaneous (Nnull = 1; 25 %). The balanced versions were re-analysed using the following additional null studies: low egocentricity (Nnull = 1; 25 %), high egocentricity (Nnull = 3; 30 %), deliberate (Nnull = 2; 25 %), spontaneous (Nnull = 1; 25 %). The clusters which survive the significance thresholds after the addition of ~30 % null studies are considered robust against potential file drawer effects.

2.5. 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 and 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 and Yu, 2020).

2.6. Ethics statement

All meta-analyses described in this research paper used pre-existing data collected across numerous peer-reviewed neuroimaging studies. These studies were all given ethical approval by their respective boards of ethics.

3. Results

3.1. ALE clusters for inner speech - all studies

The upper panel of Fig. 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
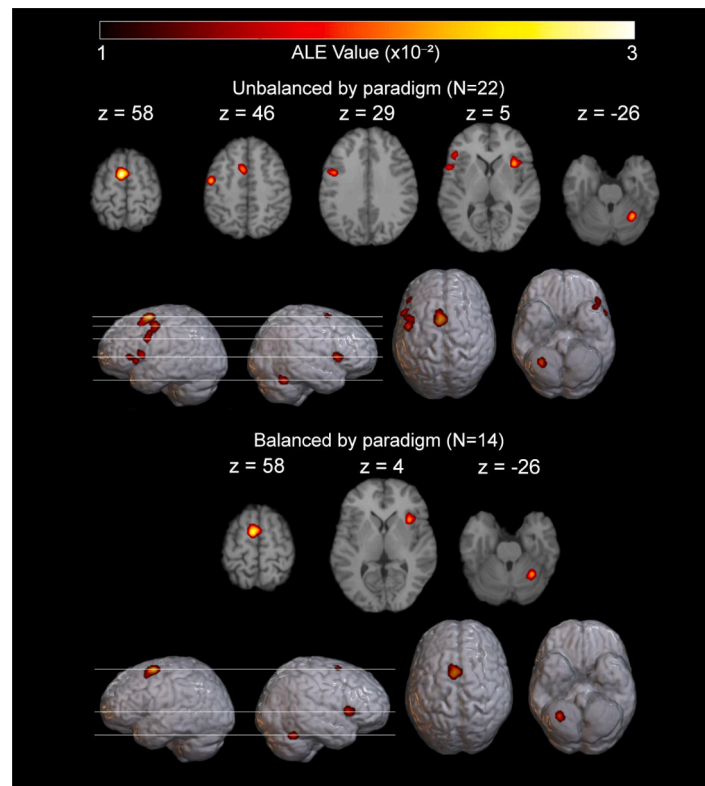
Fig. 3. Areas showing significant ALE statistic across all studies shown at FWE $p < 0.05$ at the cluster-level.

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.

gyrus (BA6 & BA44), left inferior frontal gyrus (BA45), right insula (BA13) and right culmen.

The lower panel of Fig. 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.

3.2. Inner speech as a function of egocentricity

The upper panel of Fig. 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 Fig. 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 /
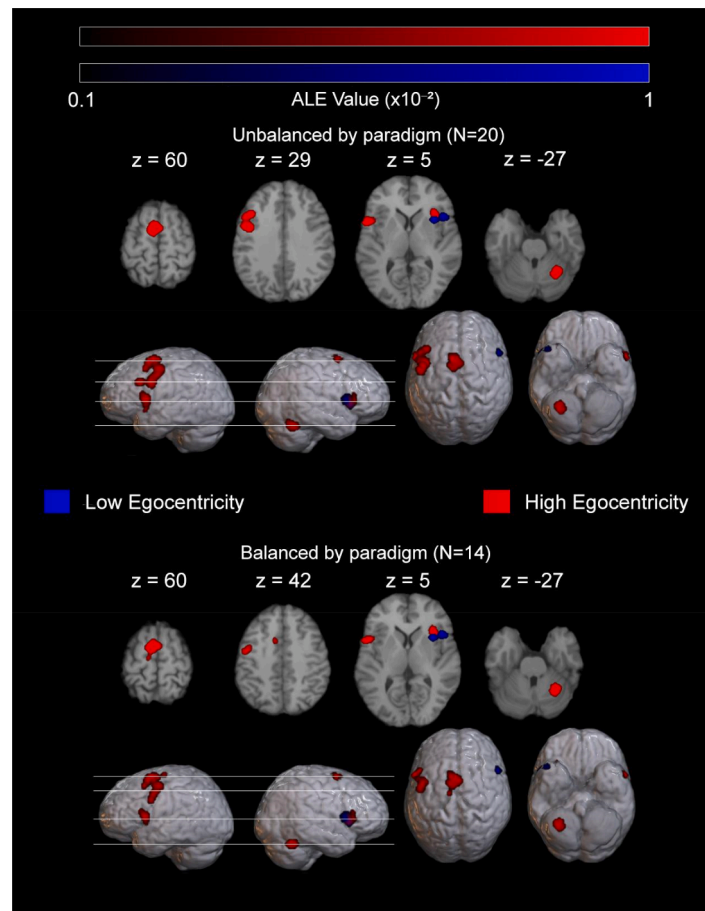
Fig. 4. Areas showing significant ALE scores in High Egocentricity (red) and Low Egocentricity studies (blue) at FWE $p < 0.05$ at the cluster-level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 ($\times 10^{-2}$)	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.

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).

3.3. Inner speech as a function of spontaneity

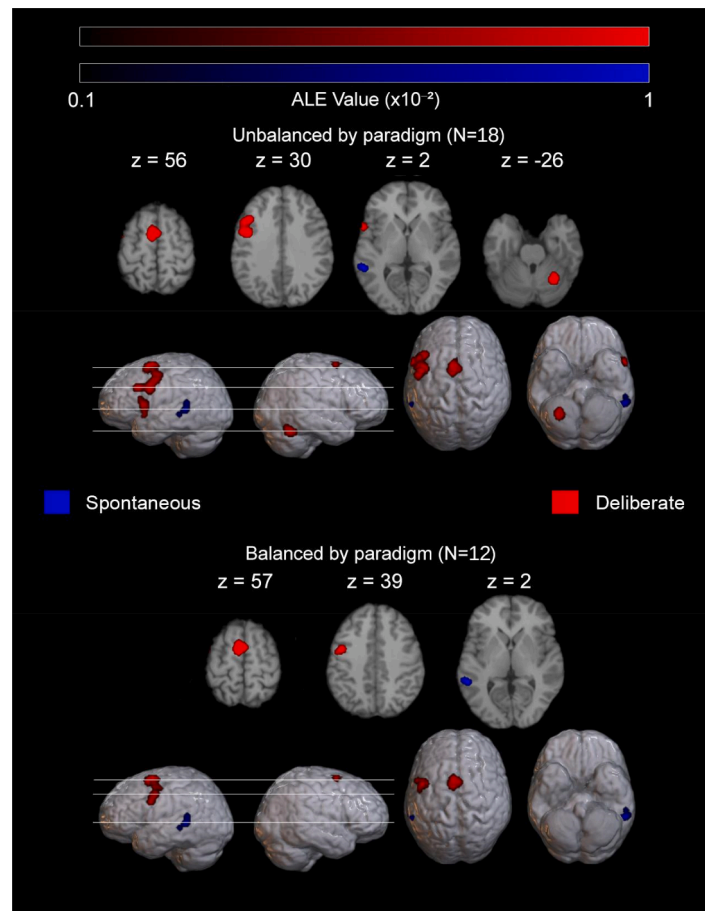The upper panel of Fig. 5 shows brain areas which displayed

Fig. 5. Areas showing significant ALE scores in Spontaneous studies (blue) and Deliberate (red) and at FWE $p < 0.05$ at the cluster-level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significant convergence for spontaneous and deliberate studies, respectively, in the *unbalanced* dataset. The Talairach coordinates associated with spontaneity analyses are presented in the 'unbalanced' section of Table 4. The *Spontaneous* studies converged on a single cluster centred on the left middle temporal gyrus and extending across the left superior temporal gyrus. The *Deliberate* 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 lower panel of Fig. 5 shows brain areas which displayed significant convergence for spontaneous and deliberate studies, respectively, in the *balanced* dataset. The associated Talairach coordinates are presented in the 'balanced' section of Table 4. The *Spontaneous* studies converged a single cluster centred on the left middle temporal gyrus (BA22) and extending across the left superior temporal gyrus (BA21). The *Deliberate* studies converged on two clusters. The largest cluster was

Table 4

Clusters showing significant ALE statistic across Spontaneous and Deliberate 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)										
Spontaneous										
1	Middle Temporal Gyrus	L	BA22	-56	-43	3	1.03	1440	0	25 %
Deliberate										
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 %*
Balanced by Paradigm (N = 12)										
Spontaneous										
1	Middle Temporal Gyrus	L	BA22	-56	-43	3	1.03	1440	0	25 %
Deliberate										
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 %*

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.

centred on the left medial frontal gyrus / supplementary motor area (BA6), with an additional cluster centred on the left precentral gyrus (BA6).

4. 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 spontaneity (deliberate 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 hypothesised that inner speech would primarily engage the corollary discharge mechanism when higher in egocentricity and/or more deliberate, and rely more on the perceptual simulation mechanism as egocentricity decreases and spontaneity increases. 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 continuous dimensions of egocentricity and spontaneity.

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 spontaneity, 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. *Spontaneous* inner speech converged on the left middle temporal gyrus (L-MTG; BA22) with a substantial portion of the cluster (> 25 %; Appendix A3) covering the left superior temporal gyrus (L-STG; BA22). *Deliberate* 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. 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.

4.1. 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 corollarydischarge model of inner speech. However, given the proximity of the L-MFG cluster to regions associated with hand/finger movement (Amiez and 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 this region is under-investigated

in the inner speech literature. 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 and Desmond, 2012; Modinos et al., 2009; Morin and 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 and Hamper, 2012; Morin and 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 compassing both the L-PCG and L-IFG (Appendix A2). 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 spontaneity. When inner speech is more egocentric and deliberate, it may be generated predominantly by corollary discharge, which can attenuate neural activity within speech perception areas (Ford et al., 2021; Hurlburt et al., 2016; Leube et al., 2010; Shergill et al., 2013). At opposite ends of these dimensions, perceptual simulation is more strongly engaged and activation is more likely to converge in speech perception areas.

4.2. 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 spontaneity at low egocentricity, and that perceptual simulation may only gain predominance when inner speech is both of low egocentricity and spontaneous. This type of inner speech is under-investigated and as such is not represented in our dataset, with all studies examining low egocentricity but deliberate 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.

4.3. Inner speech as a function of spontaneity

Within the dimension of spontaneity we hypothesised that deliberate inner speech would preferentially recruit the corollary discharge mechanism, with spontaneous 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 deliberate 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 and 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 an repository of speech motor programs for each phonemic, syllabic or multi-syllabic sound a speaker might want to produce (Tourville and 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 inner speech paradigms which deliberately generate inner speech. Given our previous proposal that the associated 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 deliberate inner speech studies which were class as phonological-judgement tasks in the *Contrast Selection & Grouping* stage of data analysis (see Section 2.3.). The results of the ALE analysis demonstrated convergence over the L-IFG across phonological judgement tasks (Appendix A2). While the subdivision of the *deliberate* 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 paradigms recruit different neurocognitive mechanisms to generate inner speech?

Analysis of spontaneous 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 spontaneous 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 and 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 are 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.

4.4. 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 spontaneous inner speech 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 spontaneous conditions, 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 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.

Despite 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 to fully elucidate the mechanisms by which it can be generated.

4.5. 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 spontaneity 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 centring the framework around two fundamental dimensions (egocentricity and spontaneity) 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 spontaneous 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 and Fernyhough, 2015; Jones and 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 deliberate, it remains unclear as to whether the observed neural correlates are specific to studies which are both low in egocentricity and deliberate, or whether they are a feature of low egocentricity inner speech more broadly. The investigation of inner speech which is low in egocentricity and spontaneous 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 and 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 and 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. Moreover, the exact processes which facilitate the modification and combination of stored memories in order to create novel speech experiences remains poorly understood. Future research could examine the extent to which the more comprehensive literature on memory modification and reactivation within the visual cortex applies to the auditory cortex and inner speech (Favila et al., 2020).

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 and 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 spontaneous 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- s 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)'s 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 spontaneity 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).

5. 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 spontaneity. In particular, speech production areas implicated in the motor-route of generation are consistently engaged in highly egocentric and deliberate inner speech, but not with inner speech which is low in egocentricity or spontaneous. 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 spontaneity. 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.

Data availability

Analysis stimuli and output are made available to the public at: https://osf.io/n7rwq/.

CRediT authorship contribution statement

Jaydan Pratts: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Gorana Pobric:** Writing – review & editing, Supervision. **Bo Yao:** Conceptualization, Validation, Resources, Data curation, Writing – review & editing, Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Appendix A1
Studies included in the 'Balanced' dataset.

Study	Contrast	Paradigm Category
Ćurčić-Blake et al. (2013)	Metrical stress judgement > Fixation symbol	Phonological judgement
Hernandez et al. (2013)	Rhyme judgement > Font matching	Phonological judgement
Grandchamp et al. (2019)	Voice imagery > N.S	Other voice imagery
Kleider-Offutt et al. (2019)	Read sentence > Rest	Other voice imagery
De Nil et al. (2000)	Read words > Fixation symbol	Single word Reading
Theys et al. (2020)	Silent reading > View character string	Single word Reading
Papathanassiou et al. (2000)	Generate verb > Rest	Word generation
Wilson et al. (2011)	Word generation > Rest	Word generation
Alderson-Day et al. (2020)	Direct quotation > Fixation symbol	Direct quotation reading
Yao et al. (2011)	Direct quotation > Fixation symbol	Direct quotation reading
Raij et al. (2017)	Verbal thought > Fixation symbol	Mind wandering
Grandchamp et al. (2019)	Verbal Mind Wandering > N.S*	Mind wandering
Okada et al. (2018a)	Tongue twister > Fixation symbol	Tongue Twister imagery
Okada et al. (2018b)	Tongue twister > Rest	Tongue Twister imagery

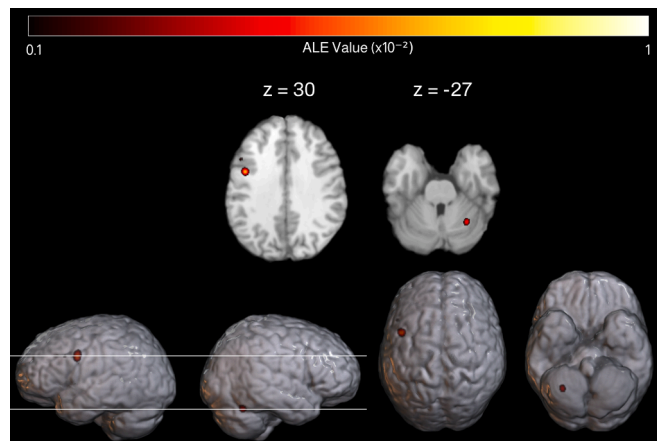A2. Phonological Judgement ALE Results (*N* = 7)

Fig. A2. ALE scores in an analysis of phonological judgement studies at FWE $p < 0.05$ at the cluster-level.

masque:

Reference Space = Talairach

Dimensions = 80 × 96 × 70

Number of within-brain voxels = 198,111 masque Size = More Conservative (Smaller)

Foci:

Coordinate System = Talairach

File of foci coordinates = Phonological Judgements.txt

Number of foci = 54

Number of experiments = 7

Total number of subjects = 96

ALE - Random Effects, Turkeltaub Non-Additive (HBM, 2012):

File of ALE voxels = Phonological Judgements_ALE.nii

FWHM minimum value = 8.95555138448401

FWHM median value = 9.755397007122182

FWHM maximum value = 11.37334466796002

Minimum ALE score = 1.411269E-37

Maximum ALE score = 0.016396433

P Values: Eickhoff (HBM, 2009)

File of P values = Phonological Judgements_P.nii

Minimum P value = 8.2646736E-8

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 = 2576 mm^3

Chosen min. cluster size = 560 mm^3

Thresholded ALE image = Phonological Judgements_C05_1k_ALE.nii

Cluster Analysis:

#1: 1848 mm^3 from (-52,-4,24) to (-38,20,38) centred at (-43.7,5,30.5) with 2 peaks with a max value of 0.0164 ALE, 8.2646736E-8 P, 5.23 Z at (-42,2,30)

Labels: (grey Matter only)

Hemisphere: 100 % Left Cerebrum

Lobe: 100 % Frontal Lobe

Gyrus: 45.5 % Precentral Gyrus, 30.3 % Inferior Frontal Gyrus, 24.2 % Middle Frontal Gyrus

Cell Type: 50.5 % Brodmann area 6, 49.5 % Brodmann area 9

#2: 728 mm^3 from (26,-64,-30) to (32,-54,-20) centred at (29,-59.3,-25.9) with 1 peaks with a max value of 0.0144 ALE, 9.343978E-7 P, 4.77 Z at (30,-58,-26)

Labels: (grey Matter only)

Hemisphere: 100 % Right Cerebellum

Lobe: 50.5 % Posterior Lobe, 49.5 % Anterior Lobe

Gyrus: 38.5 % Culmen, 17.6 % Declive, 13.2 % Pyramis, 13.2 % Tuber, 6.6 % Uvula null

Experiment Table:

[1 0 1 2 1 1 1]

[0 1 1 1 0 0 0]

Contributors to cluster #1

1 foci from Lurito,2000: Rhyming > 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 Booth,2003: Rhyme Judgement > Control, Increases

1 foci from Aparacio,2007: Rhyming > Baseline, Increases

Contributors to cluster #2

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
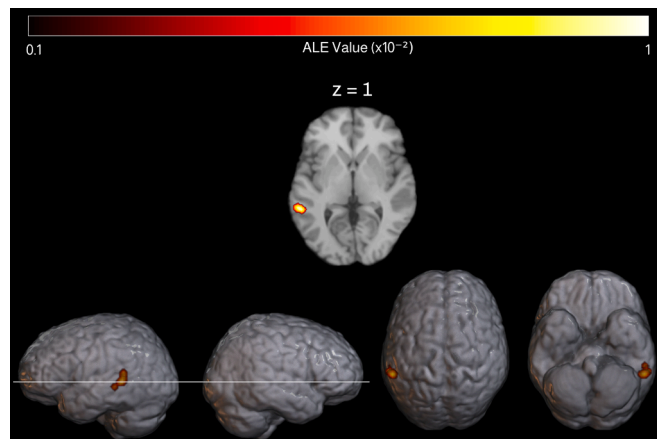
A3. Spontaneous Inner Speech ALE Results

Fig. A3. ALE scores in an analysis of spontaneous studies at FWE $p < 0.05$ at the cluster-level.

masque:

    Reference Space = Talairach

    Dimensions = $80 \times 96 \times 70$

    Number of within-brain voxels = 198,111 masque 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) centred 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: (grey 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

References

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., 2015. Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychol. Bull. 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. Conscious. Cogn. 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. J. Cogn. Neurosci. 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. Soc. Cogn. Affect. Neurosci. 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. Hum. Brain Mapp. 34 (3), 738–752. https://doi.org/10.1002/hbm.21471.

Amiez, C., Petrides, M., 2014. Neuroimaging evidence of the anatomo-functional organization of the human cingulate motor areas. Cereb. Cortex 24 (3), 563–578. https://doi.org/10.1093/cercor/bhs329.

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. J. Cogn. Neurosci. 17 (6), 928–938. https://doi.org/10.1162/0898929054021157.

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

Barsalou, L.W., Simmons, W.K., Barbey, A.K., Wilson, C.D., 2003. Grounding conceptual knowledge in modality-specific systems. Trends Cogn. Sci. (Regul. Ed.) 7 (2), 84–91. https://doi.org/10.1016/s1364-6613(02)00029-3.

Barsalou, L.W., 2008. Grounded cognition. Annu. Rev. Psychol. 59, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Barsalou, L.W., 1999. Perceptual symbol systems. Behav. Brain Sci. 22 (4), 577–660. https://doi.org/10.1017/S0140525x99002149.

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. Hum. Brain Mapp. 19 (3), 155–169. https://doi.org/10.1002/hbm.10111.

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. Cogn. Sci. 25 (5), 695–709. https://doi.org/10.1207/s15516709cog2505_4.

Carruthers, P., 2018. The causes and contents of inner speech. In: Langland-Hassan, P., Vicente, A. (Eds.), Inner Speech: New Voices. Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0002.

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. J. Cogn. Neurosci. 23 (10), 2892–2904. https://doi.org/10.1162/jocn.2010.21589.

Clowes, R., 2007. A self-regulation model of inner speech and its role in the organisation of human conscious experience. J. Cons. Stud. 14 (7), 59–71.

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

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. Psychol. Med. 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. Schizophr. Bull. 39 (5), 1087–1095. https://doi.org/10.1093/schbul/sbs107.

D'Esposito, M, 2007. From cognitive to neural models of working memory. In: Philos. Trans. R. Soc. Lond., B, Biol. Sci., 362, pp. 761–772. https://doi.org/10.1098/rstb.2007.2086.

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. J. Speech, Lang, Hear. Res. 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.

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

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

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.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. NeuroImage 59 (3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017.

Emerson, M.J., Miyake, A., 2003. The role of inner speech in task switching: a dual-task investigation. J. Mem. Lang. 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., 1996. The dialogic mind: A dialogic approach to the higher mental functions. New Ideas Psychol. 14 (1), 47–62. https://doi.org/10.1016/0732-118X(95)00024-B.

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

Favila, S.E., Lee, H., Kuhl, B.A., 2020. Transforming the concept of memory reactivation. Trends Neurosci. 43 (12), 939–950. https://doi.org/10.1016/j.tins.2020.09.006.

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.

Ford, J.M., Mathalon, D.H., 2005. Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? Int. J. Psychophysiol. 58 (2), 179–189. https://doi.org/10.1016/j.ijpsycho.2005.01.014.

Ford, J.M., Roach, B.J., Mathalon, D.H., 2021. Vocalizing and singing reveal complex patterns of corollary discharge function in schizophrenia. Int. J. Psychophysiol. 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. Psychol. Med. 17 (3), 631–648. https://doi.org/10.1017/s0033291700025873.

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. Front. Psychol. 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 Res. Bull. 54 (3), 319–328. https://doi.org/10.1016/S0361-9230(00)00441-X.

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.

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. Conscious. Cogn. 22 (4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003.

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. Am. J. Psychol. 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. Proc. Natl. Acad. Sci. U.S.A. 109 (25), 10077–10082. https://doi.org/10.1073/pnas.1200143109.

Johnson, J.F., Belyk, M., Schwartze, M., Pinheiro, A.P., Kotz, S.A., 2021. Expectancy changes the self-monitoring of voice identity. Eur. J. Neurosci. 53 (8), 2681–2695. https://doi.org/10.1111/ejn.15162.

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

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

Kosslyn, S.M., Ganis, G., Thompson, W.L., 2001. Neural foundations of imagery. Nat. Rev. Neurosci. 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.

Lancaster, J.L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum. Brain Mapp. 28 (11), 1194–1205. https://doi.org/10.1002/hbm.20345.

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 Psychol. (Amst) 181, 62–74. https://doi.org/10.1016/j.actpsy.2017.10.004.

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. Cogn. Neurosci. 1 (2), 111–117. https://doi.org/10.1080/17588921003646156.

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. Hum. Brain Mapp. 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. J. Mot. Behav. 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 J. Neurol. 132 (7), 1928–1940. https://doi.org/10.1093/brain/awp129. Pt.

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 Lang. 120 (1), 42–51. https://doi.org/10.1016/j.bandl.2011.08.005.

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. Conscious. Cogn. 20 (4), 1586–1593. https://doi.org/10.1016/j.concog.2011.08.005.

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 Psychol. (Amst) 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., Group, PRISMA, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339. https://doi.org/10.1136/bmj.b2535 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. Open Neuroimag. J. 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 Res. Bull. 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. Conscious. Cogn. 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. Front. Neurosci. 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. Neurosci. Biobehav. Rev. 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. Percept. Mot. 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 Lang. 135, 96–103. https://doi.org/10.1016/j.bandl.2014.06.003.

Okada, K., Matchin, W., Hickok, G., 2018b. Phonological feature repetition suppression in the left inferior frontal gyrus. J. Cogn. Neurosci. 30 (10), 1549–1557. https://doi.org/10.1162/jocn_a_01287.

Okada, K., Matchin, W., Hickok, G., 2018a. Neural evidence for predictive coding in auditory cortex during speech production. Psychon. Bull. Rev. 25, 423–430. https://doi.org/10.3758/s13423-017-1284-x.

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

Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2000. A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. NeuroImage 11 (4), 347–357. https://doi.org/10.1006/nimg.2000.0546.

Pearson, J., Naselaris, T., Holmes, E.A., Kosslyn, S.M., 2015. Mental imagery: functional mechanisms and clinical applications. Trends Cogn. Sci. 19 (10), 590–602. https://doi.org/10.1016/j.tics.2015.08.003.

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. Behav. Brain Res. 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. Psychol. Bull. 84 (5), 977–1001. https://doi.org/10.1037/0033-2909.84.5.977.

Raij, T.T., Riekki, T.J.J., 2017. Dorsomedial prefontal cortex supports spontaneous thinking per se. Hum. Brain Mapp. 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. Curr. Opin. Neurobiol. 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.

Rudner, M., Rönnberg, J., Hugdahl, K., 2005. Reversing spoken items—Mind twisting not tongue twisting. Brain Lang. 92 (1), 78–90. https://doi.org/10.1016/j.bandl.2004.05.010.

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

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. Res. Synth. Methods 11 (6), 866–883. https://doi.org/10.1002/jrsm.1448.

Scott, M., Yeung, H.H., Gick, B., Werker, J.F., 2013. Inner speech captures the perception of external speech. J. Acoust. Soc. Am. 133 (4), EL286–EL292. https://doi.org/10.1121/1.4794932.

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

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, 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. Psychol. Med. 31 (2), 241–253. https://doi.org/10.1017/s003329170100335x.

Sommer, I.E., Diederen, K.M., Blom, J.D., Willems, A., Kushan, L., Slotema, K., Boks, M.P., Daalman, K., Hoek, H.W., Neggers, S.F., Kahn, R.S., 2008. Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain 131 (Pt 12), 3169–3177. https://doi.org/10.1093/brain/awn251.

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

Theys, C., Kovacs, S., Peeters, R., Melzer, T.R., van Wieringen, A., et al., 2020. Brain activation during non-habitual speech production: Revisiting the effects of simulated disfluencies in fluent speakers. PLoS One 15 (1), e0228452. https://doi.org/10.1371/journal.pone.0228452.

Tian, X., Poeppel, D., 2010. Mental imagery of speech and movement implicates the dynamics of internal forward models. Front. Psychol. 1, 166. https://doi.org/10.3389/fpsyg.2010.00166.

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.

Tourville, J.A., Guenther, F.H., 2011. The DIVA model: a neural theory of speech acquisition and production. Lang. Cogn. Process 26 (7), 952–981. https://doi.org/10.1080/01690960903498424.

Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P., 2012. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. Hum. Brain Mapp. 33 (1), 1–13. https://doi.org/10.1002/hbm.21186.

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. J. Cogn. Neurosci. 22 (6), 1083–1094. https://doi.org/10.1162/jocn.2009.21309.

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

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

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

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

Wilson, S.J., Abbott, D.F., Lusher, D., Gentle, E.C., Jackson, G.D., 2011. Finding your voice: a singing lesson from functional imaging. Hum. Brain Mapp. 32 (12), 2115–2130. https://doi.org/10.1002/hbm.21173.

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, B., Belin, P., Scheepers, C., 2011. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. J. Cogn. Neurosci. 23 (10), 3146–3152. https://doi.org/10.1162/jocn_a_00022.

Yao, B., 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. Hum. Brain Mapp. 41 (16), 4662–4675. https://doi.org/10.1002/hbm.25148.

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. Neurosci. Lett. 653, 177–182. https://doi.org/10.1016/j.neulet.2017.05.042.