A cognitive neuroscience view of inner language: to predict and to hear, see, feel

Hélène Loevenbruck, Romain Grandchamp, Lucile Rapin, Ladislas Nalborczyk, Marion Dohen, Pascal Perrier, Monica Baciu, Marcela Perrone-Bertolotti

Preliminary version produced by the authors.
In *Inner Speech: New Voices.* Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. ISBN: 9780198796640

A cognitive neuroscience view of inner language: to predict and to hear, see, feel

Hélène Lœvenbruck[1], Romain Grandchamp[1], Lucile Rapin[2], Ladislas Nalborczyk[1,3], Marion Dohen[4], Pascal Perrier[4], Monica Baciu[1] & Marcela Perrone-Bertolotti[1]

1. Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105 & Université Grenoble Alpes, Grenoble, France
2. Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Canada
3. Department of Experimental Clinical and Health Psychology, Ghent University, Belgium
4. GIPSA-Lab, Département Parole et Cognition, CNRS UMR 5216 & Université Grenoble Alpes, Grenoble, France

Abstract

The nature of inner language has long been under the scrutiny of humanities, through the practice of introspection. The use of experimental methods in cognitive neurosciences provides complementary insights. This chapter focuses on wilful expanded inner language, bearing in mind that other forms coexist. It first considers the abstract vs. concrete (or embodied) dimensions of inner language. In a second section, it argues that inner language should be considered as an action-perception phenomenon. In a third section, it proposes a revision of the « predictive control » account, fitting with our sensory-motor view. Inner language is considered as deriving from multisensory goals, generating multimodal acts (inner phonation, articulation, sign) with multisensory percepts (in the mind's ear, tact and eye). In the final section, it presents a landscape of the cerebral substrates of wilful inner verbalization, including multisensory and motor cortices as well as cognitive control networks.

Keywords: abstraction, simulation, embodiment, multisensory-motor, predictive control, agency

Introduction

Mental verbalization has long been under the scrutiny of writers, philosophers, literary scholars, psychoanalysts, psychologists and linguists, through the practice of thorough introspection, careful observation and reflection. Many terms have been used to describe it, including: inner language, inner speech, inner voice, covert speech, internal speech, silent speech, self-talk, internal monologue, internal dialogue, imagined speech, endophasia, private speech, verbal thought, subvocalisation, auditory imagery. The terms "inner speech" or "voice" are too restrictive, as mental verbalization is not always oral: consider deaf people who use sign language. We will therefore use the term inner language which captures the multimodal (auditory, somatosensory and visual) qualities of mental verbalisation.

The use of experimental methods and technology in neuroscience, psychology, psycholinguistics and psychiatry provides new insights into the nature of inner language. Inner language manifests in various ways. We often deliberately engage in inner language (e.g. when we count, make a list, schedule our objectives). This can be called "wilful/volitional inner language". But sometimes, our internal monologue is less deliberate, and "more passive". This latter form has been referred to as "verbal mind wandering" (Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Lœvenbruck, 2014), and often occurs during "resting states" (mind wandering can also be non-verbal, as in visual imagery, hence the adjective "verbal"). Verbal mind wandering consists of flowing, spontaneous, stimulus-independent verbal thoughts. Whereas wilful inner language is an

attention-demanding task, verbal mind wandering has been associated with the default mode network (Raichle, 2010), although it may also additionally activate executive regions (Christoff, Gordon, Smallwood, Smith & Schooler, 2009). Neural connectivity studies have shown that the attention and default mode networks fluctuate in an anticorrelated pattern (Ossandon et al., 2011). Therefore, in addition to core language regions that they presumably share, these two modes of inner language may recruit distinct regions, related to attention vs. default mode networks. Moreover, different levels of inner language have been identified, including condensed and expanded instances (Fernyhough, 2004).

In this chapter, we focus on the nature of wilful inner verbal production, in its expanded version, bearing in mind that other forms of inner language coexist. We first consider the abstract vs. concrete dimension of inner language. In a second section, we examine its sensory vs. motor dimension and argue that inner language should be considered as an action-perception phenomenon. We describe inner language as an act, spurring the mind's eye, ear, and tact. In a third section, we propose a revision of the "predictive control" account of inner speech, to fit with our sensory-motor view of inner language. In this integrated account, inner language is considered as deriving from multisensory goals, generating multimodal acts (inner phonation, articulation, sign) with multisensory percepts (in the mind's ear, tact and eye). In the final section, we present a landscape of the cerebral networks involved in wilful inner language production, including sensory and motor cortices as well as cognitive control networks.

1. The abstract-concrete dimension of inner language

In many studies of language and cognition, an Abstraction view is taken in which inner language involves symbolic and abstract representations, divorced from bodily experience. Alternative approaches, such as the Motor Simulation view, posit that inner language is concrete and embodied, involving physical processes that unfold over time[1]. These two views reflect different positions about internal processes, the first related to classical theories of mental architecture (Fodor & Pylyshyn, 1988; Newell & Simon, 1972) and the second, to the embodied cognition framework (Barsalou, 1999; Gallese & Lakoff, 2005; Pulvermüller & Fadiga, 2010).

1.1 Arguments for the abstractness and amodality of inner language

Introspective and psycholinguistic studies of inner language have led many scholars to view it as an abstraction, unconcerned with articulatory or auditory simulations. In the Abstraction view, inner speech is articulatorily impoverished and abstract (Oppenheim & Dell, 2010; Dell & Oppenheim, 2015). MacKay (1992, p.122) confidently stated that inner speech is amodal, i.e. nonarticulatory and nonauditory. According to him, articulatory movements 'are irrelevant to inner speech. Even the lowest level units for inner speech are highly abstract'.

A first argument in favour of the Abstraction view is condensation. Inner language is considered to be autonomous from perceptuo-motor processes and their operational details, condensing it, relative to overt speech, at different levels: articulation, phonology, lexicon and syntax. Its condensation would be manifest in the time course of its production, shorter than that of overt speech.

Introspective accounts of condensation are abundant. Although Egger (1881) provided many arguments for the embodied nature of inner speech, he was the first to clearly state why inner language may indeed be shorter. First, he listed physiological constraints. We cannot articulate overtly as quickly as covertly, the speed of our tongue movements being physiologically

[1] Abstract vs. concrete in the present paper relate to the format of the representation: symbolic and amodal vs. physical and modal. They do not refer to the semantic content of inner language, which may be abstract or concrete whatever its format.

limited. Also, when we speak aloud, we need to take breath between speech fragments, as speech only occurs during expiration. Inner speech, not being subjected to these physiological constraints, can be accelerated. Secondly, Egger mentioned social constraints. In order to be understood, we need to articulate more clearly and slowly than in covert speech. Egger simply meant that the absence of physiological and social constraints shortens inner production. But drawing from similar durational observations, several psychologists have claimed that inner speech is even phonologically reduced, many phonemes being dropped and only the word-initial sounds being clearly produced (e.g. Vygotsky, 1934/1986). In this view, covert words lack the full phonological and articulatory specification they have overtly, making them more abstract and amodal. Furthermore, according to Egger, some of our mentally used expressions bear meanings that are explicit only to ourselves. To be understood by an addressee, we would need to supplement them with contextual information. Therefore, condensation occurs not only at the phonological and articulatory levels, but also at the message level. Vygotsky (1934/1986) has further developed this notion of condensation. His theory is based on introspection, and on examination of children's private speech, in which children talk to themselves aloud, and which he claimed to be a precursor of adult inner speech (but see Perrone-Bertolotti et al., 2014, for developmental data challenging this view). He asserted that important words or affixes may be dropped in inner language, the syntax of inner speech being "predicated". Bergounioux (2001) likewise claims that inner speech entails 'a generalised use of asyndeton, anaphora and an over-representation of predication' (p.120, our translation). Examples of such linguistic operations can be found in literary works associated with the "*monologue intérieur*" movement, initiated by Dujardin (1887, 1931; Smadja, in press). Hence, introspective observations have led to the speculation that inner language is impoverished, at the syntactic, lexical, phonological and articulation levels. Such condensation implies that modality-specific processes (e.g. articulatory planning) may be suppressed in inner language, making it abstract and amodal. Empirical evidence for the condensed quality of inner language has been searched for.

At the syntactic and lexical levels, evidence for condensation can be found in a study of the rate of spontaneous covert speech (Korba, 1990). Participants were asked to mentally solve short verbal problems. They reported the inner speech used to solve each problem, which gave an estimation of the number of elliptical words used. Then they delivered a full statement of their strategies, which provided an extended word count. The equivalent speaking rate of the extended statement exceeded 4000 words per minute, an unattainable rate in overt mode. These findings suggest that such inner verbalization is condensed at the syntactic and lexical levels. At the phonological level, the condensation hypothesis receives support from empirical studies showing that production is faster in covert mode, even when syntactic and lexical contents are kept equal, i.e. when participants are asked to recite the same sets of words in both modes (Anderson, 1982; MacKay, 1981; Marshall & Cartwright, 1978; Marshall & Cartwright, 1980). These studies could suggest that some of the phonological or articulatory processes involved in overt speech are absent in covert mode. An alternative interpretation, described in 1.2, is that inner speech involves the same operations as overt speech but that, as suggested by Egger (1881), the execution of articulator movements takes longer than their simulation.

A second argument for the Abstraction view is that inner speech would be deprived of some articulatory specification. Speech errors during inner recitation of tongue-twisters do display the lexical bias observed in overt production, but they do not show the phonemic similarity bias, which is based on articulatory representations (Oppenheim & Dell, 2008). This second bias is a tendency to exchange phonemes with common articulatory features (e.g. REEF slips more often to LEAF, with /r/ and /l/ sharing voicing and approximant features, than REEF to BEEF, with /r/ and /b/ only sharing voicing). Oppenheim and Dell (2008, 2010) argue that reciprocal activations between articulatory and phonological levels can explain this effect. They have only observed it in overt mode or with inner speech accompanied with mouthing, which has led them to claim that

although inner speech is specified at the lexical level (because of the lexical bias), it is impoverished at lower (articulatory) levels. According to them, unarticulated inner speech is grounded on abstract linguistic representation and can emerge before any articulatory information is retrieved.

A third argument for the Abstraction view is that typical articulatory abilities are not required in inner speech. Patients with anarthria, who have motor cortex lesions disrupting articulatory abilities, may still have intact inner speech (Baddeley & Wilson, 1985; Vallar & Cappa, 1987). This could suggest that inner speech does not depend on articulation-specific processes. However, as discussed in 1.2, another explanation is that such lesions only affect speech execution, leaving earlier stages of speech planning (including articulatory specification) unaltered.

1.2 Arguments for the concreteness and multimodality of inner language

In contrast with the Abstraction view, it has been suggested that inner speech is concrete in nature, i.e. expressed in a modal format and fully specified, down to physical, motor processes. The earliest claims of the concreteness of inner speech probably date back to Erdmann (1851) and Geiger (1868), who, as cited by Stricker (1885), introspectively observed that inner speech is accompanied by feelings of tension in the speech musculature. Stricker[2] explicitly associated inner speech with motor representations. He speculated that word representations consist in the awareness of impulsions driven from cerebral speech centres to speech muscles. In that vein, Watson (1919) described inner speech as a weakened form of overt speech. He considered inner language as a 'highly integrated bodily activity' (p.325). Although Oppenheim & Dell (2010) have held that he went as far as claiming that movements of the articulators are part of inner speech, he merely suggested that inner speech may, in some individuals, be accompanied with articulatory movement. Whether he actually alleged that movements necessarily occur in inner speech, or whether, by the term "integrated activity" he simply meant simulated action, is debatable. The extreme view that inner speech requires actual movement has been refuted by Smith, Brown, Toman, & Googman (1947) who showed that temporary paralysis induced by curare did not prevent verbal thought, memory storage and presumably inner speech. Thus, this extreme version cannot be upheld. A more nuanced view, referred to as the *Motor Simulation* hypothesis, is that inner speech is a mental simulation of articulation, without actual movement. In this view, inner speech production is described as similar to overt speech production, except that motor execution is blocked (Grèzes & Decety, 2001; Postma & Noordanus, 1996). Under the Motor Simulation hypothesis, a continuum exists between overt and covert speech, in line with the continuum between imagined and actual actions proposed by Decety and Jeannerod (1996). This has led some authors to claim that inner speech should share features with speech motor actions (Feinberg, 1978; Jones & Fernyhough, 2007) and that it may be associated with concrete physiological correlates. The Motor Simulation hypothesis is supported by several findings, which we turn to now.

a. Physiological correlates

Physiological measurements suggest that inner speech is physically planned, in the same way that overt speech is. First, as concerns respiratory rate, Conrad and Schönle (1979) have showed that the respiratory cycle varies along a continuum. During rest, breathing is symmetrical, with inspiration and expiration phases displaying equal durations. In overt speech, the cycle is strongly asymmetrical with a short inspiration and a long expiration during which speech is

[2] Stricker himself designed a clever introspective exercise to experience this orofacial activity: when one's mouth is positioned into the rounded shape required to pronounce 'o', if one tries to imagine uttering the phoneme 'm', a slight contraction is felt in the lip muscles, as if one was actually pressing lips for 'm'.

emitted. Conrad and Schönle have shown that inner speech displays a slightly prolonged expiratory phase. They concluded that motor processes are at play during inner speech (see also Chapell, 1994).

Speaking rate findings are more debated. As mentioned in 1.1, silent recitation has been found to be faster than overt recitation by many researchers. Some studies have found similar rates for covert and overt recitation, however (Landauer, 1962; Weber & Bach, 1969; Weber & Castleman, 1970). This suggests that the difference might be tenuous. Netsell and colleagues have examined spontaneous sentence production in both covert and overt modes (Netsell, Kleinsasser, & Daniel, 2016). Participants generated full sentences by saying the first thing that came to their mind. The rate of inner productions was found to be slightly faster than that of overt speech. The fact that the difference was small suggests that speaking aloud only differs from inner speech by the longer time needed to overtly articulate, once the motor plan is designed, compared with simulated articulation (see Section 3).

Concerning muscular activity, Stricker's introspective observation that inner speech is accompanied with muscular sensation finds support from a few electromyographic (EMG) studies of inner speech. Using electrodes inserted in the tongue or lips of five participants, Jacobson (1931) was able to detect EMG activity during several tasks requiring inner speech, including silent recitation. Sokolov (1972) carried out EMG measurements of lip and tongue muscles during tasks requiring different degrees of inner verbalisation. He recorded intense muscle activation during complex tasks requiring substantial inner speech production (problem solving). Conversely, a decrease in muscle activity was observed for automatized tasks, with lesser need for inner verbalisation. Surface EMG recordings carried out by McGuigan & Dollins (1989) indicated that the lips were significantly active when silently reading the letter "P" (an instance of bilabial articulation), but not when reading "T" (alveolar articulation) or a nonlinguistic control stimulus. On the opposite, the tongue was significantly active when reading "T", but not when reading "P" or the control. The authors concluded that the speech musculature used for the overt production of specific phonemes is also selectively active when covertly reading the same phonemes. Livesay, Liebke, Samaras, & Stanley (1996) measured labial EMG activity in twenty participants during rest and mental tasks. They found a significant increase in EMG activity during silent recitation compared to rest, but no increase during the non-linguistic visualisation task. A study during dreamed speech, using inserted electrodes, suggests that the silent (non-phonated) speech that occurs in dream is associated with EMG activity in *orbicularis oris* and *mentalis* muscles (Shimizu & Inoue, 1986). Surface EMG activity has also been detected in *orbicularis oris inferior* during auditory verbal hallucination (which has been described as inner speech attributed to an external source, see Section 3) in patients with schizophrenia (Rapin, Dohen, Polosan, Perrier, & Lœvenbruck, 2013). A study by Nalborczyk et al. (2017) on induced mental rumination, which can be viewed as a form of excessive negative inner speech, also shows an increase in labial EMG activity during rumination compared with relaxation. As concerns inner sign language, Max (1937) investigated activity in the *flexores digitorum*, a muscle in the forearm that flexes the fingers, in eighteen deaf participants during silent reading and mental verbal repetition. He observed that, compared to a baseline, these tasks were accompanied by an increase in EMG activity in the *flexores digitorum* in 84% of the cases. EMG activity in a control muscle did not vary as much. Overall, these results suggest that instances of inner speech or inner sign may be accompanied by activity in the orofacial or manual musculature.

b. Cerebral correlates

Several studies show that covert and overt speech production both recruit essential language areas in the left hemisphere, i.e. regions traditionally associated with speech production, such as motor and premotor cortex in the frontal lobe including Broca's area (or the left inferior frontal gyrus, LIFG), regions typically associated with speech perception, i.e. bilateral auditory areas and Wernicke's area in superior temporal gyrus (STG), and an associative region, the left inferior parietal lobule, including the left supramarginal gyrus (LSMG) (for a review, see Perrone-Bertolotti et al., 2014, 2016). However, there are differences. Consistent with the Motor Simulation hypothesis and the notion of a continuum between covert and overt speech, overt speech is associated with stronger activity in motor and premotor cortices than inner speech (e.g., Palmer et al., 2001). This can be related to the suppression of articulatory movements during inner verbal production. Moreover, overt speech recruits sensory areas more strongly than covert speech (Shuster & Lemieux, 2005). Overt speech is therefore not just inner speech with added motor processes, but it involves greater sensory activation, associated with the processing of one's speech. Reciprocally, inner speech involves cerebral areas that are not recruited during overt speech (Basho, Palmer, Rubio, Wulfeck, & Müller, 2007), such as those underlying the inhibition of overt response (cingulate gyrus, left middle frontal gyrus). Overall, these findings support the claim that inner speech is a motor simulation of speech, including motor planning, but excluding motor execution. The processes involved in overt speech therefore include those required for inner speech (except for inhibition). Lesion studies corroborate this conclusion: when overt speech is impaired, inner speech is either intact or altered, depending on the processes impacted. Several studies of brain-lesioned patients with aphasia have shown that the overt speech loss can be associated with an impairment in inner speech (e.g., Levine, Calvanio & Popovics, 1982; Martin & Caramazza, 1982). s.

Geva, Bennett, Warburton, & Patterson (2011a) have reported a dissociation that challenges this view, however. In three patients with chronic post-stroke aphasia (out of twenty-seven patients tested)[3], poorer homophone and rhyme judgement performance was observed in covert compared with overt mode. Drawing on accounts of speech production that include a speech comprehension system, such as Levelt, Roelofs & Meyer's (1999) model, Geva and colleagues suggested that inner speech relies on a connection between the production and comprehension systems, the latter being used to monitor internal representations. A damage in this connection could selectively impact inner speech while preserving overt speech. A limitation of this study, however, is that the task was to detect rhymes in written words. The deficit could have been induced by silent reading difficulties. To overcome this limitation, Langland-Hassan, Faries, Richardson, & Dietz (2015) have tested aphasia patients with a rhyming task using pictures rather than written words. The performance of patients on covert rhyming was poorer than that of controls, but many patients were unimpaired at overtly naming objects. The authors therefore suggested the deficit could be due to a specific inability to generate words in inner mode. Since the deficit was not due to an impairment in rhyme judgment (patients could judge whether words spoken to them rhymed) and since patients were also impaired in a generative naming task, the

[3] The other patients were similarly impaired in both inner and overt speech, or had an impairment with overt speech only, resulting from motor deficits or from articulatory encoding difficulties.

authors attributed the deficit in covert rhyming to a difficulty in generating multiple names for the same object to find a word rhyming with the companion picture. The authors left open the possibility that generating speech may be more cognitively and linguistically demanding in covert mode, and that inner speech may be a distinct ability, with specific neural substrates.

We suggest an alternative interpretation of this dissociation. First, the disconnection between production and comprehension systems invoked by Geva and colleagues would also impair overt speech, since the monitoring loop is recruited for overt speech, allowing for the repair of speech errors. Therefore, such a disconnection cannot explain these findings. Secondly, the specificity of inner speech defended by Langland-Hassan and colleagues is hard to reconcile with the fact that in Geva et al.'s study, correlations between inner and overt speech were significant. The lack of a comparable task in overt mode in Langland-Hassan et al.'s study makes it difficult to conclude along that line. According to our view, rhyme judgement relies on auditory representations of the stimuli (e.g. Paulesu, Frith, & Frackowiak, 1993). Overt speech generates a strong acoustic output, through the ear as well as through bone conduction, which is fed back to the auditory cortex and can be used to monitor speech. In the covert mode, the auditory information is the mentally simulated signal which is not as salient. White noise has been reported to interfere with rhyme judgments (Wilding & White, 1985), which confirms that inner auditory sensations are weak. The fact that even the control participants in Langland-Hassan et al.'s study did not reach perfect scores in the silent rhyming task supports this interpretation. In patients with aphasia, the weakness of auditory sensations may be accentuated for two reasons: first, because of an impairment in the final stages of articulatory simulation, and second, because of associated auditory deficits. Interestingly, one of the three patients in Geva et al's study had auditory comprehension deficits. Therefore, we speculate that the dissociation is due to an amplified lower saliency of the auditory sensations evoked during inner speech.

c. Articulatory specification

Another argument for the concreteness of inner speech comes from behavioural evidence of articulatory effects. Advocates of the Abstraction view have suggested that inner speech is impoverished at the articulatory level. This claim is still debated however, since a phonemic similarity bias has in fact been found by Corley, Brocklehurst & Moat (2011) during tonguetwister production, even in a covert mode. Moreover, Scott, Yeung, Gick & Werker (2013) have examined the influence of concurrent inner speech production on speech perception. They showed that the content of inner speech orients the perception of ambiguous syllables. They found that this influence even operates at the articulatory level: the inner production of /ɑ'fɑ/ vs. /ɑ'pɑ/ specifically biased perception towards /ɑ'vɑ/ vs. /ɑ'bɑ/, respectively. A recent fMRI study suggests that inner speech during reading codes detail as fine as voicing (Kell et al., 2017). In this study, the number of voiceless and voiced consonants in the silently read sentences was systematically varied. Increased voicing modulated voice-selective regions in auditory cortex. Overall, these data suggest that inner speech may indeed be specified at the articulatory level.

Moreover, studies on articulatory difficulty also reveal articulatory effects during inner speech. Smith, Hillenbrand, Wasowicz, & Preston (1986) had participants repeat bisyllabic stimuli in both overt and covert modes. The stimuli covered a range of "production difficulty". An important durational range was found across stimuli, in both modes. Words which took longer to

be (covertly and overtly) produced involved alternations in similar phonemes in the same syllable position. They concluded that production difficulty (reflected by duration) is not solely due to execution but also to planning. We add that the finding that 'wristwatch' takes longer than 'wristband' in both modes suggests that articulatory specification does occur in inner speech. The labio-velar glide /w/ is articulated with lip rounding and protrusion, and so is the retroflex /r/ (Johnson, 1997). Both require precise control of the lip configuration. This is different from /b/ which involves a ballistic lip closing gesture. The phonemes in the /r/-/w/ alternation are therefore more similar articulatorily than those in /r/-/b/. Motor control studies show that alternating between the movements of two effectors is faster than repetition of a single effector movement, because in the first case, the motion of one effector can be anticipated during the movement of the other one (Rochet-Capellan & Schwartz, 2007). This explains why 'wristwatch' is longer to pronounce overtly than 'wristband'. The fact that it is also longer covertly suggests that articulatory coordination does take place in inner speech.

d. Gestural representation in covert sign language

Another line of reasoning for the modal nature of inner language comes from the examination of inner language in deaf signers. Behavioural studies have shown that the equivalent of inner speech in deaf signers involves internal representations of signs instead of auditory representations. In a verbal short term memory task, Bellugi, Klima, & Siple (1975) showed that errors made by hearing subjects were mainly sound based, and conform to previous experiments (e.g. 'vote' misrecalled as 'boat'). This suggests that hearing subjects were coding and remembering words in terms of their phonological properties. In deaf signing subjects, substitution errors reflected the visual configurational properties of the signs (e.g. 'noon' replaced by 'tree', both featuring the same arm position in American Sign Language). Other studies of the properties of verbal working memory in deaf signers reflect a transfer from the auditory to the visual modality, with a sign length effect instead of the auditory word-length effect in spoken language, or a manual suppression effect replacing articulatory suppression (Wilson & Emmorey, 1998). Such studies suggest that sign language is stored in terms of visual percepts as well as manuo-articulatory representations, just like speech is presumably stored in both auditory and oro-articulatory formats[4]. Therefore, inner language in deaf signers presumably involves an internal representation of signs. As reviewed in MacSweeney, Capek, Campbell & Woll (2008), lesion and neuroimaging studies corroborate these data: like inner speech, inner signing involves a predominantly left-lateralized perisylvian network. Differences exist between the networks supporting signed and spoken languages, reflecting specificities in the early stages of sensory processing (auditory vs. visual) or in higher-level language characteristics (e.g. referential use of space in sign language). Yet, inner language recruits a common core of regions, independent of the modality in which it is expressed.

As mentioned above, and as detailed in Section 3, auditory verbal hallucination (AVH) can be considered as a form of inner speech, which is attributed to an external source. Admittedly, because they often occur in delusional situations, AVH cannot be taken to be fully representative of inner speech. Yet they can be viewed as a specific case of inner speech, worth considering. The descriptions of AVH in deaf patients further illustrates the modality-specific qualities of inner

[4] More precisely, signs are expressed through movements of the arms, hands and also face; speech is expressed through movements of the larynx, tongue, mouth, face and is often accompanied with hand gestures; so both modalities are presumably stored in a bracchio-manuo-oro-facial articulatory format (see 2.2).

language. Atkinson, Gleeson, Cromwell, & O'Rourke (2007) showed that the hallucinatory phenomenon in deaf schizophrenia patients depended on their auditory experience. Patients born profoundly deaf reported that the "voices" they experienced were nonauditory. They reported seeing a moving image communicating with them through sign, lip motion or fingerspelling. Deaf patients with experience of hearing speech, due to residual hearing or predeafness experience, reported auditory features or uncertainty about mode of perception.

To summarize, behavioural measurements seem to indicate that phonatory-articulatory-gestural planning is at play during inner language and that inner language may be accompanied with activity in the speech and sign musculature. In terms of brain activity, overt and covert language seem to share common core neural correlates, with overt language recruiting motor and sensory areas more than inner language, and inner language recruiting inhibition circuits more than overt language. Therefore, contrary to the Abstraction view, some instances of inner language seem fully physically planned, including concrete articulatory (laryngeal, orofacial and manual) specifications that are coordinated, just like in overt language, but that are inhibited and not executed.

1.3 Coexistence of abstract-amodal and concrete-multimodal forms

The seemingly opposite views of Abstraction and Motor Simulation are not mutually exclusive, however. As explained in Fernyhough (2004), Alderson-Day and Fernyhough (2015) or Geva et al. (2011b), at least two levels of inner speech can be distinguished. The first one, condensed inner speech, is argued to correspond to Vygotsky's (1934) description: "inner speech is to a large extent thinking in pure meanings" (p.249). In Vygotsky's view, inner speech has lost most of the acoustic and structural qualities of external speech. As Vygotsky wrote, "[Th]e development of verbal thought takes the [following] course: from the motive that engenders a thought to the shaping of the thought, first in inner speech, then in meanings of words, and finally in words" (p.253). This level of inner speech can indeed be considered as abstract in format. Expanded inner speech, on the other hand, retains many of the phonological properties of external dialogue, and can be viewed as concrete in format. Fernyhough (2004) has suggested that inner speech varies with cognitive and emotional conditions between these two (or more) forms. We consider the expanded form as an outcome of the condensed form. The condensed form, we conjecture, is the conceptual message cast in a preliminary linguistic form, that involves lemmas[5], linearly ordered, but that does not yet have the full phonological (articulatory, gestural, acoustic) specification that expanded inner language has. A similar position is taken and defended in detail in Vicente & Martínez-Manrique (2016). Inner language can be defined as truncated overt verbalisation, but the level at which the production process is interrupted (abstract linguistic representation vs. articulatory/gestural representation) depends on which variant of inner language is at play. In the rest of this chapter, we will focus on expanded inner language.

2. The sensory-motor dimension of inner language

If we accept the concrete nature of inner language, at least in its expanded version, then we are still faced with another question related to its nature: is inner language motor or sensory? Are

[5] The term lemma in Levelt and colleagues' terminology refers to the word's syntax, see Levelt et al. (1999). It is different from the lexeme which denotes the word's phonological features and from the lexical concept which refers to the word's semantics.

inner speaking (or signing) and inner hearing (or viewing) different phenomena or are they two sides of the same coin?

2.1 Arguments for a motor or enactive nature

As explained in 1.2, the Motor Simulation view, also referred to as the 'Action' view (Jones & Fernyhough, 2007) or the 'Activity' view (Martinez-Manrique & Vicente, 2015) holds that inner language is an act, with a prior intention to express a certain thought, which is transformed into orofacial and/or manual motor commands. This view is grounded both on introspective experiments and empirical data (physiological recordings, behavioural measures as well as neuroimaging and brain lesion data) described above.

Inner language therefore seems to involve motor acts that are inhibited. If inhibition prevents motor acts from actually being executed, then the neurophysiological activity measured in peripheral muscles must be explained. We suggest that motor commands might be emitted, together with inhibitory signals blocking articulatory movement. This speculation is in line with Jeannerod & Decety's (1995) description of action imagery. According to them, during mental simulation of an action, "it is likely that the excitatory motor output generated for executing the action is counterbalanced by another parallel inhibitory output. The competition between two opposite outputs would account for the partial block of the motoneurons, as shown by residual EMG recordings and increased reflex excitability" (p.728).

Inner language therefore seems to involve the production of imaginary motor acts; be they articulatory, facial or manual. In a predictive control account, these imaginary motor acts can be viewed as the predicted actions that result from a copy of inhibited motor commands (see Section 3). They can be posited to correspond to the activations observed in premotor cortex and inferior frontal regions. They seem to have physiological sequels in orofacial muscles and in respiratory patterns. It could therefore be concluded from empirical data that inner language is fundamentally of a motor or enactive nature.

Yet, as explained in the preceding section, these imaginary motor acts give rise to sensory percepts, feelings in our muscles (Stricker, 1885) but also sounds in our heads. Taine (1870) himself was a precursor when he recognized the motor and sensory qualities of verbal thought: 'In normal state, we silently think with words that are mentally heard, read or uttered, and what is inside of us is the image of such sounds, letters, or of such muscular and tactile sensations in the throat, tongue and lips' (p.25-26, our translation). The sensory qualities of inner language are examined in the following section.

2.2 Arguments for a sensory nature

Early introspective works have claimed that inner speech is endowed with auditory qualities. Egger (1881) and Ballet (1886) claimed that rhythm, pitch, intensity and even timbre can be found in inner speech. The concept of an inner ear (or mind's ear) finds support in recent data.

The 'Verbal Transformation Effect' (VTE) refers to the perceptual phenomenon in which listeners report hearing a new percept when an ambiguous stimulus is repeated rapidly (Warren, 1961). Rapid repetitions of the word 'life', for example, produce a soundstream that is fully

compatible with segmentations into 'life' or 'fly'. Reisberg, Smith, Baxter, & Sonenshine (1989) examined the imagery analogue of the VTE. Participants were instructed to imagine the word "stress" being repeated by a friend's voice. The VTE was observed (subjects detected the compatible word "dress") showing that subjects are able to imagine an ambiguous soundstream, to parse it and find alternative construal of it. Smith, Wilson, & Reisberg (1995) further studied the VTE in a covert mode, using Baddeley's distinction between two components of the phonological loop involved in verbal short-term memory. According to Baddeley, the phonological loop relies on the "inner ear" – the phonological store –, and on the "inner voice"[6] – the articulatory rehearsal process. Smith and colleagues examined the VTE in a covert mode, asking whether the imagery judgement and reconstrual is based on the inner voice, the inner ear or both. Participants were instructed to imagine a friend repeating the word "stress" and to report any transformation. Repetition imagery was executed in three conditions: no-interference, articulatory suppression and irrelevant speech perception. A more important VTE was found in the no-interference condition than in the suppression and irrelevant-speech conditions. The disruptive impact of articulatory suppression was interpreted as a role for the inner voice in the VTE: to discern the transformations, subjects need to subvocally rehearse the material. The impact of irrelevant speech was taken to suggest that the VTE also depends on the inner ear. It was concluded that subjects seem to reinterpret ambiguous verbal images by using both components of the phonological loop, the inner voice and ear.

The neural correlates of the VTE have been examined by Sato et al. (2004). Participants were asked to silently repeat pseudo-words such as /psə/. In the baseline condition, participants were asked to covertly repeat a pseudo-word over and over. In the verbal transformation condition, they additionally had to actively search for a transformation (from /psə/ to /səp/ for instance). When compared with the baseline condition, active search for verbal transformation correlated with stronger activation in the left inferior frontal gyrus, left supramarginal gyrus, bilateral cerebellum as well as left superior temporal gyrus: when inner speech involves consciously attending to mental production, speech production as well as perception regions are more strongly activated. These results therefore corroborate the hypothesis of a close partnership between inner production and perception in the VTE.

Findings of error detection during covert tonguetwister repetition also seem to indicate that inner verbal production has sensory qualities that can be attended to. As mentioned in Section 1, several studies (reviewed in Dell & Oppenheim, 2015) show that participants are able to report the errors that they mentally hear. This can be interpreted as a role for the mind's ear in inner speech monitoring. A recent fMRI study of slip detection provides contradictory findings, however. Gauvin, Baene, Brass, & Hartsuiker (2016) investigated whether internal verbal monitoring takes place through the speech perception system. In a production condition, they had participants produce tongue-twisters overtly and judge whether their production was correct or incorrect, while white noise was presented via headphones to mask auditory feedback. Adding noise was meant to induce internal verbal monitoring, as participants could not hear their auditory feedback, while ensuring that the experimenter could judge repetition correctness. In a perception condition, participants simply heard the tonguetwister and made a correctness judgment. The superior temporal areas were found to be activated by error detection during the perception condition but not during production. The authors concluded that internal monitoring

[6] 'Inner voice' is taken here as the imaginary motor act (articulation and phonation). In the rest of the chapter, it refers to the result of that act, i.e. the auditory stimulus heard in the mind's ear.

occurs independently of speech perception systems. The fact that no activation was found in speech perception areas during the production condition could be due to the use of noise masking, however. Adding noise saturates the auditory system, which could mask subtle differences between contrasts. The examined contrasts were between erroneous and correct trials. Erroneous trials were instances in which an error was detected, which could indeed augment the activation of the auditory system, relative to a correct trial, but quite subtly. Since in both types of trials, the noise level was high, this subtle difference might have been undetectable. Therefore, we do not think that these results are conclusive.

Neuroimaging studies of covert speech production themselves reveal auditory cortex, and specifically superior temporal gyrus, activation (Perrone-Bertolotti et al., 2014 for a review). Although this activation is lesser than the one observed in overt speech, it entails that an auditory experience accompanies inner speech. An interesting study suggests that inner speech can be disrupted during abnormal activity in the temporal lobe. Vercueil & Perrone-Bertolotti (2013) described the case of a woman who reported experiencing inner speech jargon (incomprehension of her own inner language) during her epileptic seizures which involved sharp theta waves in the left temporal regions.

Evidence for auditory sensations during reading has been provided by experimental psychology. Several studies suggest that silent reading is modulated by the knowledge of the author's speaking speed (Alexander & Nygaard, 2008), the talker's voice familiarity (Kurby, Magliano, & Rapp, 2009) or the reader's regional accent (Filik & Barber, 2011). The involvement of the mind's ear during silent reading has been recently confirmed by fMRI experiments (Yao, Belin, & Scheepers, 2011, 2012). Several areas in the auditory cortex, called temporal voice area (TVA), are selectively involved during human voice perception (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Yao and colleagues contrasted silent reading of direct (e.g., Mary said: "I'm hungry") and indirect speech (e.g., Mary said that she was hungry) sentences. The direct speech condition induced greater activation of the right TVA than the indirect speech condition, which suggest that voice-related perceptual representations are more engaged when silently reading direct speech statements. Further support for the assumption that silent reading involves the mind's ear comes from an fMRI study by Lœvenbruck, Baciu, Segebarth & Abry (2005). In the baseline condition, participants silently read a sentence in French, with a neutral prosody (*Madeleine m'amena*, "Madeleine brought me around"). In the prosodic focus condition, they silently read the same sentence, adding contrastive focus on the subject. In an overt mode, this would correspond to higher pitch and longer duration on the focused subject, followed by pitch compression on the post-focal constituents (*MADELEINE$_F$ m'amena*). When compared with the baseline, the silent prosodic focus condition yielded greater activity in the left inferior frontal gyrus, insula, supramarginal gyrus as well as in Wernicke's area. These results suggest that when we silently read, we can use specific prosodic contours, with distinctive auditory qualities. These auditory variations correspond to objectively measurable cerebral correlates. Further evidence for TVA activation during silent reading comes from intracranial EEG recording of TVA in four epileptic patients (Perrone-Bertolotti, Kujala, Vidal et al., 2012). Patients were instructed to perform a silent reading task in which attention was manipulated: they were asked to only attend to the words written in grey, ignoring the white words. Consecutive grey words formed a story, about which they were questioned after the experiment. The results not only showed that silent reading activate the TVA, but also, that the neural response to written words was increased during attended compared to unattended words. This suggest that TVA activity increase is under top-down attentional control. It must be noted however, that reading is not systematically associated

with inner speech, even when attention is high. A few aphasia case reports suggest that some reading abilities may be maintained even when inner speech is impaired (Levine et al., 1982; Saffran & Marin, 1977). This can be explained by the fact that silent reading of frequent words may take a direct route from orthography to meaning, without necessarily recurring to inner speech (Coltheart, 2005). To sum up, behavioural and neuroimaging data suggest that auditory sensations are often present during silent reading.

The concept of a mind's ear is appropriate, but it is insufficient. As we have argued, the imaginary sensory consequences of imaginary motor acts may be multimodal: they may lead to sounds in our heads and, as hinted by Taine (1870) or Stricker (1885), to imaginary proprioceptive and tactile sensations. Paulhan (1886) claimed that inner speech involves visual, auditory and motor images. By visual images he meant the form, shape and colour of the letters that compose written words. He stated that these were rare. He qualified auditory images as dominant in inner speech. He defined motor images as the sensations in the speech organs that sometimes accompany inner speech. Contrary to Stricker who considered inner speech as purely motor, he claimed that motor images cannot be isolated from auditory images, whereas the reverse is possible.

A few terminology precautions are necessary here. It is not always clear what the nineteenth century authors meant by "motor" and "articulatory" representations. Even nowadays, "articulatory" is often opposed to "auditory" or "acoustic", with some confusion. Sometimes, the process is targeted: what is meant by "articulatory" is motion (action), in contrast to audition (perception). Sometimes, modality is at play: "articulatory" refers to somatosensory sensations, in contrast to auditory percepts. Bearing in mind this confusion, we use the term "motor" to refer to action and "somatosensory" to describe bodily sensations. Although Stricker clearly claimed that inner speech consisted of imagined actions, Paulhan's intuitive notion of "motor images" are related to somatosensory percepts, i.e. to the evocation of sensations, rather than to the simulation of actual speech movements.

Nevertheless, inner speech seems indeed to involve somatosensory sensations, which include proprioception and tactile sensations. Proprioception provides information about articulator location and movement and is sent by receptors in the muscles, joints and skin. Tactile information corresponds to the touch sense from mechanoreceptors that report contact (e.g. between tongue and palate). According to Lackner & Tuller (1979), speech errors can be detected by means of proprioceptive and tactile information and it has been claimed that proprioceptive and tactile feedback play a role in speech motor control (Levelt, 1989; Postma, 2000; Gick, 2015). We speculate that imagined proprioceptive and tactile feedback are part of inner speech: in addition to the mind's ear, the mind's 'tact' should also be considered. Moreover, the fact that, as explained above, motor commands may reach muscles during inner speech, could explain the actual (not imagined) sensations in the speech muscles introspectively reported by Stricker and Paulhan. We will further address the co-existence of motor, auditory and proprioceptive representations in Section 3.

Finally, the 'mind's eye' certainly plays a role in inner language. As mentioned earlier, inner language representations in deaf signers include visual information. Gestures are not only used in the deaf population. They accompany speech in normal hearers and play a fundamental role in thought and speech (De Ruiter, 2007). Moreover, speech is audiovisual: lip reading enhances speech comprehension when the acoustic signal is degraded by noise (Sumby & Pollack, 1954). Lip reading occurs even with nondegraded acoustic signals, as the McGurk effect shows (McGurk

& MacDonald, 1976). Auditory and visual speech information include common stages of processing (Nahorna, Berthommier, & Schwartz, 2015). These findings suggest that visual information (facial and manual) could be involved in inner speech, even in hearing subjects. A preliminary work by Arnaud, Schwartz, Lœvenbruck, & Savariaux (2008) provides tentative suggestions that speakers can have visual representations of their own lip movements. Furthermore, as suggested by Paulhan, visual written representations may occur during inner speech. More research is needed to confirm that inner language involves visual (labial, facial, manual, written) representations, even in the hearing population.

Inner verbalizing therefore involves the reception of imaginary sensory signals, presumably including auditory, proprioceptive, tactile and visual elements, handled by the mind's ear, tact and eye.

To wrap up, the nature of inner language is both motor and sensory. One can conceive that imaginary acts give rise to multisensory percepts. But these acts themselves could stem from prior sensory goals, as Paulhan hinted in 1886. The precedence of some sensory representations over motor ones in wilful inner verbalization will now be discussed, in a motor control framework.

3. Integrating the sensory-motor nature of inner language into the 'Predictive Control' account

The 'sensory-motor' nature of inner verbalization can be accounted for in a motor control perspective in which intended sensory goals can give rise to motor acts which themselves generate sensory percepts. The 'predictive control' account of inner speech, also called 'comparator model', pertains to this perspective. This account is based on the hypothesis that action control uses internal models, i.e. systems that simulate the behaviour of a natural process (Kawato, Furukawa, & Suzuki, 1987; Jordan & Rumelhart, 1992). Two kinds of internal models, forward and inverse models, are supposed to be coupled and regulated through several comparators. A forward model is an internal representation of the system (body, limb, organ) that captures the forward or causal relationship between the inputs to the system (motor commands) and the outputs (Wolpert & Kawato, 1998). An inverse model performs the inverse computation, i.e. provides motor commands from desired sensory states. During the execution of a goal-directed motor task, an inverse model computes motor commands from the specification of desired changes in the sensory state of the motor apparatus. A copy of the motor commands, called "efference copy", is fed to a forward model that, given the current state of the apparatus, generates a prediction of the upcoming sensory consequences of the action. Thanks to its negligible delay, this sensory prediction, also called "*internal* feedback", ensures a stable feedback control of actions (Miall, Weir, Wolpert, & Stein, 1993; Miall & Wolpert, 1996; Wolpert & Kawato, 1998). The propagation of the *actual* feedback to the central nervous system is indeed delayed, due to axon transmission and synaptic delays (during speech production, the delay between auditory feedback perturbation and motor command adaptation is about 200 ms, *i.e.* the duration of one syllable, Houde, Nagarajan, Sekihara, & Merzenich, 2002). Because of these delays, a control based on actual feedback would either require very slow execution or be unstable. Forward models, by providing an internal feedback that occurs earlier than the actual experience, can trigger early error correction and allow for stable action control.

The efference copy mechanism is not only crucial to smooth motor control. It is also considered to play a role in the awareness of action. It has been hypothesized that disruptions in the predictive mechanism could lead to delusions of control and, in the case of speech, to auditory hallucination (e.g., Frith, 1992; Frith, Blakemore, & Wolpert, 2000). A model is presented in Figure 1, that explains this hypothesis in the context of overt speech (and that includes an adaptation to

inner speech, detailed below). To make things clear, we take as example the goal of 'uttering vowel /i/', although it is debatable whether such a mechanism is necessary in isolated vowel production. In Frith and colleagues' view, the goal is associated with a desired[7] multisensory state, which can be expressed in terms of articulatory properties (anterior elevated tongue position, lip spreading, phonation) as well as acoustic properties (first two spectral formants spread apart). An inverse model transforms the desired sensory state into motor commands, which are sent to the articulatory-phonatory motor system. This leads to the production of labial, lingual, and laryngeal movements, and to an acoustic signal. In turn, these movements and sound generate long-delay somatosensory and auditory feedbacks, the actual sensory experience. An efference copy of the motor commands is also sent to a forward model, which generates predicted somatosensory and auditory feedbacks. A delay is applied to the internal feedback signals (which become the "corollary discharge") so that they are synchronized with the actual feedbacks. The efference copy mechanism is depicted in dashed line in Figure 1. The predictive model includes three state comparisons which have each a specific role in overt speech production.

Insert Figure 1 about here

The first comparison (referred to as C1 in Figure 1) takes place between the actual sensory feedback and the desired sensory state. If a discrepancy results from C1, the inverse model receives an error signal and the motor commands are adjusted. C1 is irrelevant in ongoing actions, as the time necessary for the actual feedback to reach the central nervous system is of about one syllable. This would lead to utterly slow speech production. C1 is instead supposed to play a role in speech learning, by tuning the inverse model to produce motor commands that are best adapted to new goals. Moreover, it has been suggested that C1 contributes to the sense of body ownership, the pre-reflective experience that it is our own body that is currently moving, voluntarily or not (Gallagher, 2000; Franck & Thibaut, 2003; Tsakiris, Schütz-Bosbach, & Gallagher, 2007).

The second comparison (C2) is the one involved in the stable control of actions, using internal instead of actual feedbacks. It compares desired and predicted states. Via C2, errors can be detected in the motor commands, and be corrected, before actual feedback reaches the central nervous system[8].

A third comparison (C3) is involved, between the actual sensory state and the delayed prediction (corollary discharge). If the afferent sensory feedback and the corollary discharge do not match, the forward model is adjusted. This forward model updating, together with the inverse model tuning via C1, are claimed to improve performance when learning new actions, by generalizing the tuning for future productions (but see Tremblay, Houle, & Ostry, 2008; Rochet-Capellan, Richer, & Ostry, 2012, who have only observed limited generalization to future actions). It has been suggested that C3 could also play a role in self-monitoring (Wolpert, Ghahramani, & Jordan, 1995; Wolpert, 1997). If the actual sensory feedback matches the predicted sensory signal, then the sensory cortex could be informed that the perceived stimuli are self-generated, which would provide a sense of agency. Frith (1992) posited that a defective predictive system could

[7] We use the term « desired » rather than « intended », to allow for unintended action to be monitored via this mechanism (see below, on unbidden thoughts).

[8] Discrepancies between desired and predicted feedbacks could also be due to an inaccurate forward model. C2 could therefore also contribute to adjust the forward model, not just the inverse model. But this would require an additional mechanism by which the discrepancy would be sent either to the inverse or the forward model. In the absence of evidence for such a mechanism, we stick to the classical view, with C2 only affecting the inverse model.

explain why a self-initiated action may be experienced as externally controlled in delusions of control: if the predicted and actual sensory feedbacks do not match, then some external influence must have taken place.

In line with agency, another advantage of the predictive model is that it explains the observed modulation of sensory cortex activity during self-initiated actions. If subjects can predict the sensations they are going to feel, then these are not informative and can be attenuated, relative to externally caused sensations which need to be attended. When actual and predicted feedbacks match, the sensory consequence of the motor act is thus attenuated, compared with the same stimulation produced by an external agent (Blakemore, 2003; Blakemore, Frith, & Wolpert, 1999; Frith, 2002). This mechanism has been invoked to explain why we cannot tickle ourselves (Blakemore, Wolpert, & Frith, 2000).

According to some authors, the mere presence of a predicted signal (even before C3 takes place) could itself contribute to the awareness of initiating a movement, to feeling in control (Blakemore, Wolpert, & Frith, 2002; Frith, 2002). Temporal measurements by Libet, Gleason, Wright, & Pearl (1983) or Haggard, Newman, & Magno (1999) indeed indicate that subjects are aware of initiating a movement about 80 ms before the actual movement occurs.

In sum, predictive control seems to play an important role in self-monitoring. It is claimed that it provides senses of ownership and agency, that are essential components of self-awareness, via C1 and C3 comparisons, involving desired, predicted and actual states.

The predictive control framework has been fruitful in the speech domain (Guenther, Ghosh, & Tourville, 2006; Houde & Nagarajan, 2011; Postma, 2000). It has even been applied to covert speech production (Feinberg, 1978; Frith, 1992). Several researchers, including Frith (1992), Jones & Fernyhough (2007), Seal, Aleman, & McGuire (2004), have claimed that disruptions in the predictive control mechanism explain auditory verbal hallucination (AVH). According to their view, if the prediction is faulty, the actual sensory consequences of inner speech are not attenuated and agency is not felt. Either because of attributional biases (Seal et al., 2004) or simply because self-authorship is not felt (Jones & Fernyhough, 2007), inner speech would then be experienced as other-generated.

The involvement of a corollary discharge in inner speech control is supported by several studies. Dampening or delaying of auditory cortex responsivity has been observed during inner speech (with EEG: Ford & Mathalon, 2004; with MEG: Numminen & Curio, 1999) and interpreted as a modulatory influence of frontal speech production areas on temporal speech reception areas. This finding should be interpreted with caution, however. In Ford & Mathalon's EEG study, the inner production was preceded by an auditory stimulus, which could have dampened the subsequent auditory response (via auditory suppression). Nevertheless, in an fMRI study, Shergill et al. (2002) did find increased fronto-temporal connectivity during inner speech, associated with the increase in inner speaking rate. Tian, Zarate, & Poeppel (2016) also found temporal cortex activation during inner speech, which they related to the presence of a corollary discharge. Scott (2013) provided behavioural evidence for auditory attenuation in inner speech. The "Mann effect" refers to the influence of contextual speech sounds on the perception of subsequent speech sounds. Scott showed that this effect was specifically weakened when the contextual sound was played during matching speech imagery, suggesting that the impact of the auditory stimuli was only attenuated when inner speech matched.

Some researchers have questioned the functional relevance of a monitoring system in inner speech, however. MacKay (1992) specifically asked "*why speakers must independently 'listen to' the meaning and sound of what they are saying internally when they know all along the meaning and sound of what they are saying"* (p.140). Stephens & Graham (2000), Gallagher (2004) and Langland-Hassan (2008) also argue that the predictive mechanism is redundant in inner speech.

This critique has been addressed by Jones & Fernyhough (2007) who inscribe the necessity for self-monitoring in a Vygotskyan developmental perspective. According to them, children start off by overt "private speech", simulating dialogues with interlocutors. Verbal thought would only become covert after several years, through a gradual process of internalization. During this process, it is crucial for children to be able to label the received auditory stimuli as self- or other-generated. This means that the efference copy is not an *ad hoc* mechanism solely invoked to explain delusions of thought insertion, but it is ontogenetically necessary for inner speech to develop from private speech. We further claim that distinguishing self-generated from other-generated voices remains compelling in adult inner speech. As argued in Section 2, we can hear our inner voice, its timbre, and its intonational variations, we can even detect inner speech errors. We can have imaginary dialogues, involving various voices. We claim that it is through self-monitoring that we do not mistake these internal voices for external voices, and that we are aware that we have imagined them. A broader role for the predictive mechanism will be discussed below, related to awareness and distinguishing wilful inner speech from unbidden thoughts. But before this, we need to address a further critique, stemming from the direct application of the predictive control model to inner speech. Although Frith (1992, see also Feinberg, 1978) was one of the first to suggest that inner speech control could rely on such a model, he himself questioned the notion of actual sensory feedback during inner speech, which by essence is silent and motionless (Frith, 2012). In the case of inner speech, C3 is irrelevant, as it would compare a predicted sensory signal with an absent actual feedback. Rapin et al. (2013) have offered two alternative accounts (see also Rapin, Dohen, & Lœvenbruck, 2016).

The first account relies on the hypothesis that, as argued in 2.1, inhibitory signals may be sent to prevent motor command amplitude from reaching a sufficient threshold for speech movement to occur. But even though the speech apparatus may not move, the motor commands could slightly increase muscle tension. The actual sensory feedback during inner speech would thus consist of some residual proprioceptive feedback (rather than auditory). During AVHs, this residual signal could be the sensory feedback that does not match the faulty prediction and that leads to self-generated signals being interpreted as external.

In the second account, the relevant comparison for agency-monitoring during inner speech is not C3 because the actual feedback is silent and motionless. Agency during inner speech, which is faulty during AVH, cannot come either from the mere presence of a prediction (see above), because we claim that the predicted signal is precisely what becomes identified as an external voice (or manual/facial gestures in deaf subjects). Instead, we suggest that agency comes from C2, the comparison between desired and predicted states (a distinction between predicted state and predicted experience is made below). The AVH symptoms could be explained as follows: if the prediction is defective, then there is no match between predicted and desired states, agency is not felt and the inner voice or gesture (predicted experience) could feel alien. In addition, C2 would signal a discrepancy, which would abort the perceptual attenuation and which would reinforce the saliency of the inner voice (or gesture), accentuating its alien character. It must be reminded that C2 was originally introduced to explain stable feedback control of action, by early tuning of the motor commands when the predicted state does not match the goal. C2 should therefore still issue a sense of agency in case of goal unattainment, i.e. when the prediction and the goal are only slightly discrepant.

The first account, which entails that some proprioceptive sensations could subsist in muscles during inner speech, is supported by introspective experiments (Stricker, 1885). The second account has been concurrently formulated by Tian & Poeppel (2012) as well as Swiney & Sousa (2014) who have similarly proposed that C2 is the suitable comparison for agency and perceptual attenuation in inner speech. This was even proposed by Frith (2005) himself. It can also be found, incidentally, in Gallagher (2000). These two accounts are compatible and are integrated in Figure 1. The lines and boxes shaded in light grey are irrelevant in inner speech and

In *Inner Speech: New Voices.* Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. ISBN: 9780198796640

only apply to overt speech. The red arrow corresponds to inhibitory signals sent in parallel with the goal of inner speaking (this arrow is irrelevant in overt speech).

We have added to Figure 1 the concept of "inner language percepts", at the level of the predicted experience. Tian & Poeppel, Swiney & Sousa, as well as Scott (2013) and Scott et al. (2013), argue, like us, that the voice perceived during inner speech precisely consist of the predicted signal. In other domains, researchers have claimed that the forward model could be used during mental training, to predict the sensory consequences of an action without having to execute it. This could tune the inverse model for future actions (Jeannerod & Pacherie, 2004; Pacherie, 2008). The predicted signal would thus correspond to the subjective feeling in mental imagery (Grush, 2004). During inner speech, the predicted signal would thus equate to the voice mentally heard (and the somatosensory sensations felt), or the sign/lip gesture internally seen. As explained above, this simulated signal occurs earlier than the actual experience would, which explains why inner speech may be shorter than overt speech.

We also include in Figure 1 two perceptual attenuation mechanisms, at the level of C3 for external sensory signals and at C2 for internally generated signals. During overt speech control, we speculate that agency could result from both C2 and C3 and would therefore be stronger that during inner speech. C2 would attenuate the predicted sensory signal (the inner voice) and C3 would dampen the external feedback.

According to our view, the efference copy is therefore more than a mechanism that identifies self vs. other-generated voices. It is what makes our own verbal thoughts come to awareness (Frith, 2010). As mentioned above, it has been argued that the desired state, used by the inverse model to derive the efference copy, could itself be the inner voice consciously heard and felt as our own, with no need for a prediction (Gallagher, 2004; Langland-Hassan, 2008). In Langland-Hassan's 'filter model', the mere existence of an efference copy, without computing a prediction, and without using a comparator, could act as a filter during inner speech. This filter would itself endow our actions with a sense of agency. In this alternative model, however, the comparator mechanism is argued to be necessary for other somatosensory modalities than those involved in inner speech. This entails that different mechanisms would be required depending on the modality, which does not seem parsimonious. We add that the desired state itself cannot be experienced as a voice. In many motor control theories, the comparisons take place between end sensory states, not between ongoing experiences. We speculate that the desired state, being expressed in terms of goals in acoustic and articulatory spaces, is a coarse plan, not a full speech experience, with the unfolding of speech muscle movements and sounds over time. In Figure 1, we have included our speculated distinction between predicted experience and predicted end state. The predicted experience, developing over time, is the inner voice. The predicted state is the end sensory product, compared with the desired goal. The inner voice is not thoroughly felt until it is fully simulated over time, through the efference copy. And it is not felt as self-intended before its end product, the predicted state, is compared with the desired state. We further speculate that top-down executive signals presumably control for the generation of a prediction. Three types of verbal thought can then be explained. First, unbidden thoughts, i.e. verbal thoughts without a feeling of agency (Gallagher, 2004), can be viewed as desired states with no corresponding predicted states. They sound evanescent and muffled, because they are not fully specified over time. They do not feel alien, because no comparison is made at all, presumably because no top-down signal has launched the generation of a prediction. In Figure 1, we have added "unbidden thoughts", at the level of the desired state. A second type is wilful inner speech, in which top-down signals initiate the generation of a prediction. A sensory experience unfolds over time and an inner voice is distinctly heard. The desired and predicted states match (even only slightly): agency is felt. A third type is AVH, in which top-down signals initiate the generation of a prediction, but, due to a dysfunction, the desired and predicted states do not match at all and the prediction feels alien. The alien voice is vividly heard, as the absence of perceptual attenuation (due to the discrepant

comparison) makes the predicted experience more salient than an ordinary inner voice. The efference copy mechanism therefore contributes to creating the rich sensory qualities of inner speech, as well as the feeling of agency, of awareness of our thoughts.

In summary, in our adapted version of the predictive control account, wilful inner language is seen as a process in which verbal goals are converted to motor commands. These motor commands are inhibited but still transmitted to the orofacial and manual systems, giving rise to residual proprioceptive feedback. This residual feedback may provide a sense of ownership and is probably felt by some individuals. In parallel, a copy of the motor commands is sent to a forward model which computes multisensory inner language percepts associated with the simulated acts: sounds – the inner voice heard by our mind's ear –, proprioceptive sensations in the orofacial and manual musculature – the inner movements perceived by our mind's tact –, and visible manual/lip gestures – the inner signs potentially seen by our mind's eye.

We note here that this account does not explain how visual information about one's manual and facial movements may be derived from motor commands. The role of forward models is to map motor commands onto resulting sensory percepts, through a simulation of the motor and sensory systems. During overt speech, the sensory percepts resulting from the motor commands sent to the speech musculature to pronounce an /i/, for instance, correspond to the audition of the /i/ sound and the associated proprioceptive sensation. But are they linked with visual information about the associated facial configuration? In sign language, the sensory percepts associated with the motor commands sent to form the sign for 'tree', for instance, correspond to the proprioceptive sensation of a raised arm and hand as well as to the vision of a straight arm and extended hand, in an egocentric perspective. Lip-reading being so important in deafness, is visual information about one's arm linked with information about one's lips, in an allocentric perspective? If future research shows that visual information about one's face indeed plays a role in language production, even in hearing subjects (as suggested in Section 2), then an additional mechanism needs to be included to handle the presence of predicted allocentric facial visual information in addition to the predicted egocentric visual feedback about the arm and hand.

4. A cerebral landscape

We will now sketch a landscape of the cerebral regions involved in wilful inner language production. Our sketch is based on findings and theoretical assumptions in linguistics, psycholinguistics, and neurolinguistics described in the previous sections. It shares some of the hypotheses described in the functional anatomic models of overt speech production by Guenther & Vladusich (2012), Hickok (2012) or Tian & Poeppel (2013), but it differs in specific points. It is displayed in Figure 2. The diagram in Figure 1 is also complemented with anatomical locations corresponding with this sketch. In both figures, the efference copy mechanism is depicted in dashed lines.

A few words of caution are first needed. The functional anatomic sketch proposed here specifically describes volitional inner language production. Additional regions, or perhaps a different network altogether, may be at play for the more evanescent and less wilful form of inner speech that corresponds to verbal mind wandering (see Introduction). Moreover, Hurlburt (2011) makes a phenomenological distinction between "inner hearing" and "inner speaking", which does not coincide with our view of inner hearing as the sensory prediction elicited by the act of inner speaking. Hurlburt claims that another form of inner hearing exists, in which subjects feel as the recipient of the voice, not their creator. A different network might mediate this particular phenomenon of "inner hearing". A neuroimaging study seems to confirm this intuition (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016), although we think the inner hearing phenomenon elicited was in fact related to verbal mind wandering. Hurlburt's inner hearing could be close to

verbal mind wandering, or it could be akin to auditory verbal hallucination. Our sketch is restricted to Hurlburt's "inner speaking". Moreover, we only consider the very last concrete stages of inner language production, sometimes referred to as "expanded inner speech". As argued in 1.3, condensed inner speech may correspond to the initial stages of inner language. These include two of the stages described in Levelt (1989): the conceptualizer, the output of which is a linearized preverbal message, and grammatical encoding, which consists in selecting the appropriate lemmas from the lexicon and arranging them in a syntactic order. We start our sketch of inner language production precisely where condensed inner speech may most certainly stop, *i.e.* once lemmas have been retrieved and arranged. To simplify things, we restrict the landscape to the production of single words with no consideration of syntax or prosody. Inner speech can be produced with one's own or someone else's voice (e.g. Geiselman & Glenny, 1977). The cerebral networks underlying the monitoring of different voices remain to be described (but see Grandchamp et al., 2016), so our sketch is limited to own-voice inner speech. Finally, current data on inner sign are too scarce, but we speculate that the auditory processes invoked in our sketch may be replaced with visual processes in inner sign. Therefore, our sketch only applies to wilful spoken inner production of isolated words with one own's voice.

Insert Figure 2 about here

So, if we skip conceptual preparation and grammatical encoding, we can start off with a lemma being retrieved. The meta-analyses by Indefrey & Levelt (2004) and Indefrey (2011) suggest that lemma retrieval is handled by the mid-section of the left middle temporal gyrus (Brodmann Area (BA) 21). Tian & Poeppel (2013) locate this process in the left posterior middle temporal gyrus. Until more research decides between these two proposals, we broadly associate lemma retrieval with the left middle temporal gyrus. So, inner word production presumably starts with an activated left middle temporal gyrus mediating lemma retrieval.

According to Levelt et al. (1999), the next stage is phonological code retrieval, which generates the lexeme. It is not clear whether Levelt and colleagues think it implies sound as well as articulation. Although Levelt (1994) states that phonological encoding generates "an articulatory or phonetic shape for all words" (p.91), Indefrey & Levelt (2004) in fact reduce this stage to activations in Wernicke's area. Tian & Poeppel (2013)'s model also limits this stage to auditory specification. We see things differently. In our revised predictive control account, the inner language goal (or the lemma) is associated with a desired state, expressed in a multisensory format. Because more research is needed to confirm the role of visual information, we restrict the sketch to auditory and somatosensory information. The sketch remains fully compatible with the inclusion of visual information, via visual cortex activation, however (and is also compatible with inner sign production). According to us, the lemma is converted to a lexeme in a multisensory format, through two pathways, one for auditory and one for somatosensory representations. These two pathways are presumably parallel, but auditory specification may in fact be sequentially followed by somatosensory specification, or the reverse. A similar view is proposed in Hickok (2012)'s model. Hickok makes the additional claim that these two pathways correspond to two hierarchical levels. The higher level codes speech information at the syllable level and involves auditory goals, whereas the lower level deals with articulatory feature clusters, roughly corresponding to phonemes, and involves somatosensory goals. It is not clear to us whether the auditory and somatosensory pathways are reserved to one level each. Further research will help better specifying this stage. Meanwhile, we remain agnostic as to whether a parallel or a sequential scheme applies, and as to whether each pathway is linked to a specific speech level or

not. In Figure 2, these two pathways are simply labelled as 'a' and 'b', for auditory and somatosensory, respectively. Following suggestions by Indefrey & Levelt (2004), Guenther et al. (2006), Hickok (2012) as well as Tian & Poeppel (2013), we posit that the auditory specification activates the left posterior superior temporal gyrus (pSTG) and the superior temporal sulcus (STS), arrow 1a. Following Guenther et al. (2006) or Hickok (2012) we further suggest that the parallel somatosensory pathway activates the anterior supramarginal gyrus (aSMG) and the primary somatosensory cortex (S1), arrow 1b.

The next stage in Levelt and colleagues' model is syllabification, which is supposed to operate directly from the phonological code (the desired state) and to be mediated by the left inferior frontal gyrus (Indefrey, 2011). In line with the predictive control account, we suggest that a transformation is first needed, from the desired state expressed in a multisensory format, to commands, expressed in a motor format. This inverse model transformation involves two pathways. The auditory specification is fed to the temporo-parietal junction (TPJ, arrow 2a). The somatosensory specification is sent to the cerebellum (arrow 2b). Activities in the cerebellum has indeed been observed during the execution of a motor task and has been related to the generation of motor commands (Gomi et al., 1998; Grush, 2004; Imamizu & Kawato, 2009; Kawato et al., 1987; Wolpert, Miall, & Kawato, 1998).

This transformation will make it possible for motor programs to be specified, again following two pathways, as in Hickok (2012): the transformed auditory goals are sent from the TPJ to the left inferior frontal gyrus (LIFG) and to the premotor cortex (ventral BA6), arrow 3a; the transformed somatosensory goals are sent from the cerebellum to the lower primary motor cortex (M1), arrow 3b. We add to Hickok (2012) the speculation that the motor programs issued by LIFG are themselves sent to M1 (arrow 4) presumably leading to a unique motor plan, specified in M1, and integrating the two motor programs, from the auditory and the somatosensory pathways.

Articulation is then inhibited, via a signal presumably emitted when the desired state was specified (or even at an earlier stage, during lemma retrieval or conceptual preparation). This signal is probably issued in regions involved in inhibitory control, i.e. in rostral prefrontal cortex (BA 10) and anterior cingulate gyrus (BA 32) (Basho *et al.*, 2007). It may be sent to M1 only (if the assumption that a unique motor program is specified in M1 is correct) or to both LIFG and M1, as suggested in Figure 2. A residual somatosensory feedback may be felt, resulting from attenuated motor commands being sent to the motor system. This may activate the aSMG and S1.

Once motor commands are computed, an efference copy is used by the forward model to simulate a predicted state. We assume that this transformation involves the inverse pathways from the ones used in the transformation from sensory states to motor commands. A similar step is taken by Tian et al. (2016). But whereas they assume a sequential pathway, from motor representations in the LIFG to auditory consequence in pSTG and STS via somatosensory consequences in SMG, we stick to the two-pathway scheme. The efference copy mediated by LIFG is sent to the TPJ (arrow 4a) and is transformed into a predicted auditory signal that activates pSTG and STS (arrow 5a). The other copy, in M1, is sent to the cerebellum (arrow 4b) and is transformed into a predicted somatosensory signal that activates aSMG and S1 (arrow 5b). We conjecture that C2 (between predicted and desired states) takes place at two sites, in auditory and somatosensory cortices. This comparison is presumably under the supervision of cognitive control regions, but too little research has been carried out in this field to make any speculation.

Conclusion

Although we are still far from having a complete picture of the nature of inner language, we argue that our integrated approach, in which inner language is conceived of as multimodal acts with multisensory percepts, stemming from coarse multisensory goals, is backed up by data in linguistics, psycholinguistics, and neurolinguistics. Many issues still need to be resolved. First, the

dynamics of cerebral activation proposed here still needs empirical evidence. We have claimed for instance that inner language stems from a coarse desired multisensory state that originates from temporal and parietal regions, and is converted into motor commands in the frontal regions. A copy of these motor commands is itself converted back into a predicted multisensory signal, with activations in temporal and parietal regions. A temporo-parieto-fronto-temporo-parietal loop is therefore hypothesized and should be demonstrated. Further research is needed to assess the dynamic pattern of activation and connectivity of the cerebral regions involved in inner word production. Second, we have limited ourselves to word-level production. Further research should examine the additional processes involved in full sentence generation. Third, we have only focused on the later stages of inner language, once conceptual preparation and grammatical encoding have taken place. These early stages should be examined for a full picture of inner language. Fourth, we have mainly focused on wilful inner speech, yet verbal mind wandering is a very frequent inner language instance, which may be related to the experience of inner hearing without feeling in control (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016). Better understanding its mechanism would provide important insights into the origin of auditory verbal hallucination. Further examining the fluctuations between unvoluntary and wilful inner language could also help explaining verbal rumination, an excessive form of negative inner speech, during which supervisory mechanisms seem faulty. Finally, current theories do not provide satisfactory accounts of how cognitive control is unfolded during inner language. Although many of the subcomponents of inner language processes can be associated with specific regions or networks, several stages remain unknown. In particular, it is still unclear which regions process the results of the comparisons supposed to occur in the predictive control account and how cognitive control integrates these outcomes. We are currently carrying out research to explore these issues.

Acknowledgements

This research was funded by the ANR project INNERSPEECH [grant number ANR-13-BSH2-0003-01], http://lpnc.univ-grenoble-alpes.fr/InnerSpeech. We thank Anne Vilain, Maëva Garnier, Jean-Philippe Lachaux, Luciano Fadiga, Christopher Moulin, Cédric Pichat, Yanica Klein, Laurent Lamalle, Jean-Luc Schwartz, Irène Troprès for helpful advice. We are grateful to the editors Agustin Vicente and Peter Langland-Hassan for their insightful comments and suggestions on an earlier version of this chapter.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychological Bulletin*, 141(5), 931-965.

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 446-459.

Anderson, R. E. (1982). Speech imagery is not always faster than visual imagery. *Memory & Cognition,* 10, 371-380.

Arnaud, L., Schwartz, J.-L., Lœvenbruck, H., & Savariaux, C. (2008). Perception as a (Shaped) Mirror of Action: It Seems Easier to Lipread One's Own Speech Gestures than those of Somebody Else. *Workshop on Speech and Face to Face Communication*, Grenoble, 27-29 Oct. 2008.

Atkinson, J. R., Gleeson, K., Cromwell, J., & O'Rourke, S. (2007). Exploring the perceptual characteristics of voice-hallucinations in deaf people. *Cognitive neuropsychiatry,* 12, 339-361.

Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language,* 24, 490-502.

Ballet, G. (1886). Le langage intérieur et les diverses formes de l'aphasie. *Félix Alcan, éditeur, Paris.*

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, **22**, 577-609.

Basho, S., Palmer, E. D., Rubio, M. A., Wulfeck, B., & Müller, R.-A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. *Neuropsychologia*, *45*(8), 1697–1706.

Belin, P., Zatorre, R., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, Nature Publishing Group,* 403, 309-312.

Bellugi, U., Klima, E., & Siple, P. (1975). Remembering in signs. *Cognition*, 3 (2), 93-125.

Bergounioux, G. (2001). Endophasie et linguistique [Décomptes, quotes et squelette]. *Langue française,* 132, 106-124.

Blakemore, S. (2003). Deluding the motor system. *Consciousness and Cognition,* 12 (4), 647-655.

Blakemore, S., Frith, C., & Wolpert, D. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience,* 11(5), 551-559.

Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport, 11*(11lemma), R11-R16.

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends In Cognitive Sciences, 6*, 237-242.

Chapell, M. S. (1994). Inner Speech and Respiration: Toward a Possible Mechanism of Stress Reduction. *Perceptual and Motor Skills,* 79, 803-811.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, 106, 8719-8724.

Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. *The science of Reading: a Handbook, Oxford: Blackwell Publishing*, 6-23.

Conrad, B., & Schönle, P. (1979). Speech and Respiration. *Archiv Für Psychiatrie Und Nervenkrankheiten*, *226*, 251–268.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 162–175.

Decety, J., & Jeannerod, M. (1996). Mentally simulated movements in virtual reality: does Fitt's law hold in motor imagery? *Behavioral Brain Research, 72*, 127-134.

Dell, G., & Oppenheim, G. M. (2015). Insights for Speech Production Planning from Errors in Inner Speech. *The Handbook of Speech Production,* Redford, M. (Ed.), *John Wiley & Sons*, West Sussex, UK, 404-418.

De Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture,* 7(1), 21-38

Dujardin, E. (1887). Les lauriers sont coupés. Published in instalments in *Revue indépendante,* May-August 1887, Paris (subsequent one-volume edition Paris : éditions Messein, 1925).

Dujardin, E. (1931). *Le Monologue intérieur*, *son apparition, ses origines, sa place dans l'œuvre de James Joyce et dans le roman contemporain*. Paris : Messein.

Egger, V. (1881). *La Parole intérieure. Essai de psychologie descriptive.* G. Baillière, (Ed.), Paris.

Erdmann, J. E. (1851). *Psychologische Briefe,* Aufl. Leipzig: Reichardt.

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin, 4*, 636–640.

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, 22, 49-68.

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PloS One*, *6*(10), e25782.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-7.

Ford, J., & Mathalon, D. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research, 38*(1), 37-46.

Franck, N., & Thibaut, F. (2003). Les hallucinations. *Encyclopédie Médico-Chirurgicale*, 37120 A10.

Frith, C., Blakemore, S., & Wolpert, D. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews,* 31, 357-363.

Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. *Psychology Press.*

Frith, C. D. (2002). Attention to action and awareness of other minds. *Consciousness and Cognition,* 11(4), 481-487.

Frith, C. D. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition,* 14(4), 752-770.

Frith, C. D. (2010). What is consciousness for? *Pragmatics & Cognition, 18*(3), 497-551.

Frith, C. D. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition,* 21(1), 52-54.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences,* 4(1), 14-21.

Gallagher, S. (2004). Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology, 37*(1), 8-19.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology,* 22, 455-479.

Gauvin, H. S., Baene, W. D., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage, 126*, 96-105.

Geiger, L. (1868). *Ursprung und Entwickelung der menschlichen Sprache und Vernunft,* Verlag der J. G. Cotta'schen Buchhandlung, Stuttgart.

Geiselman, R., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, 5, 499-504.

Geva, S., Bennett, S., Warburton, E. A., & Patterson, K. (2011a). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology,* 25, 323-343.

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

Gick, B. (2015). Speech production and articulatory phonetics. *Proceedings of the 18th ICPhS, Glasgow, University of Glasgow.*

Gomi, H., Shidara, M., Takemura, A., Inoue, Y., Kawano, K., & Kawato, M. (1998). Temporal firing patterns of Purkinje cells in the cerebellar ventral paraflocculus duringocular following responses in monkeys I. Simple spikes. *Journal of Neurophysiology*, 80, 818–831.

Grandchamp, R., Rapin, L., Lœvenbruck, H., Perrone-Bertolotti, M., Pichat, C., Lachaux, J. P., & Baciu, M. (2016). Inner Speech with your own or someone else's voice. Cerebral correlates assessed with fMRI. *Society for the Neurobiology of Language (SNL) Conference 2016*, London 17-20 August 2016.

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1-19.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and brain sciences,* 27(3), 377-396.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280-301.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25, 408 - 422.

Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology, 90*(2), 291-303.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience,* 13(2), 135-145.

Houde, J., & Nagarajan, S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience,* 5, 82.

Houde, J., Nagarajan, S., Sekihara, K., & Merzenich, M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neurosci*ence, *14*(8), 1125-1138.

Hurlburt, R. T. (2011). *Investigating pristine inner experience: Moments of truth*. (Cambridge University Press).

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. http://doi.org/10.1371/journal.pone.0147932

Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychological Research PRPF*, 73(4), 527-544.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology,* 2, 255.

Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition,* 92, 101-144.

Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities. VII. Imagination, recollection, and abstract thinking involving the speech musculature. *American Journal of Physiology, 97*, 200-209.

Jeannerod, M., & Decety, J. (1995). Mental motor imagery: a window into the representational stages of action. *Current Opinion in Neurobiology,, 5*, 727-732.

Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & Language, 19*(2), 113-146.

Johnson, K. (1997). *Acoustic and auditory phonetics*. Blackwell, Oxford.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*, 391-399.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive science,* 16(3), 307-354.

Kawato, M., Furukawa, K., & Suzuki, R. (1987). Hierarchical neural-network model for control and learning of voluntary movement. *Biological cybernetics,* 57 (3), 169-185.

Kell, C. A., Darquea, M., Behrens, M., Cordani, L., Keller, C., & Fuchs, S. (2017). Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading. *Human Brain Mapping, 38*, 493-508.

Korba, R. J. (1990). The Rate of Inner Speech. *Perceptual and Motor Skills,* 71, 1043-1052.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition, Elsevier,* 112 (3), 457-461.

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper, & E. C. T. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merril Garret*. Hillsdale, NJ: Lawrence Erlbaum.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and motor skills*, 15, 646-646.

Langland-Hassan, P. (2008). Fractured phenomenologies: Thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language, 23*(4), 369-401.

Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, 6, 1-10.

Levelt, W. J. M. (1989). *Speaking: from intention to articulatio*n. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1994). The skill of speaking. In Bertelson P., Eelen P., d'Ydewalle G. (Eds.), *International Perspectives on Psychological Science Vol. 1: Leading themes*, Hillsdale, NJ: Lawrence Erlbaum Associates, 89–103.

Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences,* 22, 1-38.

Levine, D. N., Calvanio, R., & Popovics, A. (1982). Language in the absence of inner speech. *Neuropsychologia*, *20*(4), 391–409.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*(3), 623–642.

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perception and Motor Skills, 83*, 1355–1362.

Lœvenbruck, H., Baciu, M., Segebarth, C., & Abry, C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, *18*(3), 237–258.

MacKay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of motor behavior,* 13, 274-285.

MacKay, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory Imagery*, NJ/England: Erlbaum, 121–49.

MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences,* 12, 432 – 440.

Marshall, P. H., & Cartwright, S. A. (1978). Failure to replicate a reported implicit-explicit speech equivalence. *Perceptual and Motor Skills,* 46, 1197-1198.

Marshall, P. H., & Cartwright, S. A. (1980). A final (?) note on implicit/explicit speech equivalence. *Bulletin of the Psychonomic Society,* 15, 409-409.

Marshall, R., Rappaport, B., & Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language,* 24, 297-313.

Martin, R. C., & Caramazza, A. (1982). Short-term memory performance in the absence of phonological coding. *Brain and Cognition,* 1, 50 - 70.

Martinez-Manrique, F., & Vicente, A. (2015). The activity view of inner speech. *Frontiers in psycholo*gy, 6, 232, doi: 10.3389/fpsyg.2015.00232

Max, L. W. (1937). Experimental study of the motor theory of consciousness. IV. Action-current responses in the deaf during awakening, kinaesthetic imagery and abstract thinking. *Journal of Comparative Psychology,* 24, 301.

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *Pavlovian Journal of Biological Science, 24*, 19–26.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature,* 264, 746-748.

Miall, R., & Wolpert, D. (1996). Forward models for physiological motor control. *Neural networks,* 9(8), 1265-1279.

Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a Smith Predictor? *Journal of Motor Behaviour,* 25, 203-216.

Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2015). Audio-visual speech scene analysis: characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America,* 137 (1), 362-377

Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Polosan, M., Spinelli, E., … & Lœvenbruck, H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology, 127*, 53-63.

Netsell, R., Kleinsasser, S., & Daniel, T. (2016). The Rate of Expanded Inner Speech During Spontaneous Sentence Productions. *Perceptual and Motor Skills,* 123, 383–393.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. New-York: Prentice-hall.

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters, 272*(1), 29 – 32.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528-537.

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

Ossandon, T., Jerbi, K., Vidal, J. R., Bayle, D. J., Henaff, M.-A., Jung, J., Minotti, L., Bertrand, O., Kahane, P., & Lachaux, J.-P. (2011). Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance. *The Journal of Neuroscience*, *31*(41), 14521–14530.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179-217.

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

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature,* 362, 342-345.

Paulhan, F. (1886). Le langage intérieur et la pensée. *Revue Philosophique de la France et de l'Étranger, PUF,* 21, 26-58.

Perrone-Bertolotti, M., Grandchamp, R., Rapin, L., Baciu, M., Lachaux, J. P., & Lœvenbruck, H. (2016). Langage intérieur. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : Du cerveau au langage*. Louvain-la-Neuve : De Boeck Supérieur, Collection Neuropsychologie, 109–24.

Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., … & Lachaux, J.-P. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *J Neurosci*, *32*(49), 17554–17562.

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

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*, 97-132.

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

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience,* 11, 351-360.

Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, *14*(4), 180–190.

Rapin, L., Dohen, M., & Lœvenbruck, H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : Du cerveau au langage*. Louvain-la-Neuve : De Boeck Supérieur, Collection Neuropsychologie, 347-370.

Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Lœvenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research, 56*(6), S1882-S1893.

Reisberg, D., Smith, J., Baxter, D., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *The Quarterly Journal of Experimental Psychology,* 41, 619-641.

Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology, 107*(6), 1711-1717.

Rochet-Capellan, A., & Schwartz, J.-L. (2007). An articulatory basis for the labial-to-coronal effect:/pata/seems a more stable articulatory pattern than/tapa. *The Journal of the Acoustical Society of America,* 121, 3740-3754.

Saffran, E. M. & Marin, O. S. M. (1977). Reading without phonology: Evidence from aphasia. *Quarterly Journal of Experimental Psychology*, 29, 515-525.

Sato, M., Baciu, M., Lœvenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C., & Abry, C. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *Neuroimage,* 23 (3),1143–51.

Scott, M. (2013). Corollary Discharge Provides the Sensory Content of Inner Speech. *Psychological Science,* 24, 1824-1830.

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America,* 133, EL286-EL292.

Seal, M. L., Aleman, A., & McGuire, P. K. (2004). Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry, 9*(1-2), 43-72.

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping, 16*(4)*,* 219-227.

Shimizu, A., & Inoue, T. (1986). Dreamed speech and speech muscle activity. *Psychophysiology*. 23, 210-4.

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono-and multisyllabic words. *Brain and Language*, *93*(1), 20–31.

Smadja, S. (in press). *La Recherche en littérature : approches stylistiques du monologue intérieur*. Paris: Hermann.

Smith, B., Hillenbrand, J., Wasowicz, J., & Preston, J. (1986). Durational characteristics of vocal and subvocal speech-implications concerning phonological organization and articulatory difficulty. *Journal of Phonetics,* 14, 265-281.

Smith, J., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia, 33*, 1433-1454.

Smith, S., Brown, H., Toman, J., Googman, L. (1947). The Lack of Cerebral Effects of d-Tubocurarine. *Anesthesiology*, 8(1), 1-14.

Sokolov, A. (1972). *Inner speech and thought*. Plenum Press New York.

Stephens, G. L., & Graham, G. (2000). *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press.

Stricker, S. (1885). *Du langage et de la musique.* Traduit de l'allemand par Frédéric Schwiedland, Alcan, Paris.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America,* 26, 212-215.

Swiney, L., & Sousa, P. (2014). A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech. *Frontiers in Human Neuroscience, 8*, 72-86.

Taine, H. (1870). *De l'intelligence*, 2 vols. Paris: Hachette.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience, 6*, 152-162.

Tian, X., & Poeppel, D. (2013). The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive Neuroscience,* 25 (7), 1020-1036.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex, 77*, *1 – 12.*

Tremblay, S., Houle, G., & Ostry, D. J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, *28*(10), 2426–2434.

Tsakiris, M., Schütz-Bosbach, S., & Gallagher, S. (2007). On agency and body-ownership: Phenomenological and neurocognitive reflections. *Consciousness and Cognition, 16*(3), 645-660.

Vallar, G., & Cappa, S. F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology,* 4, 55-77.

Vercueil, L., & Perrone-Bertolotti, M. (2013). Ictal inner speech jargon. *Epilepsy & Behavior*, *27*(2), 307-309.

Vicente, A., & Martínez-Manrique, F. (2016). The Nature of Unsymbolized Thinking. *Philosophical Explorations*, 19, 173-187.

Vygotski, L. S. (1934/1986). *Thought and Languag*e. English Translation by Alex Kozulin. The MIT Press, Cambridge, MA: MIT Press.

Warren, R. M. (1961). Illusory changes of distinct speech upon repetition -- the verbal transformation effect. *British Journal of Psychology,* 52, 249-258.

Watson, J. B. (1919). *Psychology from the standpoint of a behaviorist*. (J. B. Lippincott, Ed.). Philadelphia.

Weber, R. J., & Bach, M. (1969). Visual and speech imagery. *British Journal of Psychology,* 60, 199-202.

Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Psychophysics,* 8, 165-168.

Wilding, J., & White, W. (1985). Impairment of rhyme judgments by silent and overt articulatory suppression. *The Quarterly Journal of Experimental Psychology Section A,* 37, 95-107.

Wilson, M., & Emmorey, K. (1998). A "word length effect" for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, 26, 584-590.

Wolpert, D. (1997). Computational approaches to motor control. *Trends in cognitive sciences,* 1(6), 209-216.

Wolpert, D., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks,* 11, 1317-1329.

Wolpert, D., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9), 338–347.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science,* 269, 1880-1882.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience,* 23, 3146-3152.

Yao, B., Belin, P., & Scheepers, C. (2012). Brain 'talks over' boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage,* 60, 1832-1842.

In *Inner Speech: New Voices.* Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. ISBN: 9780198796640

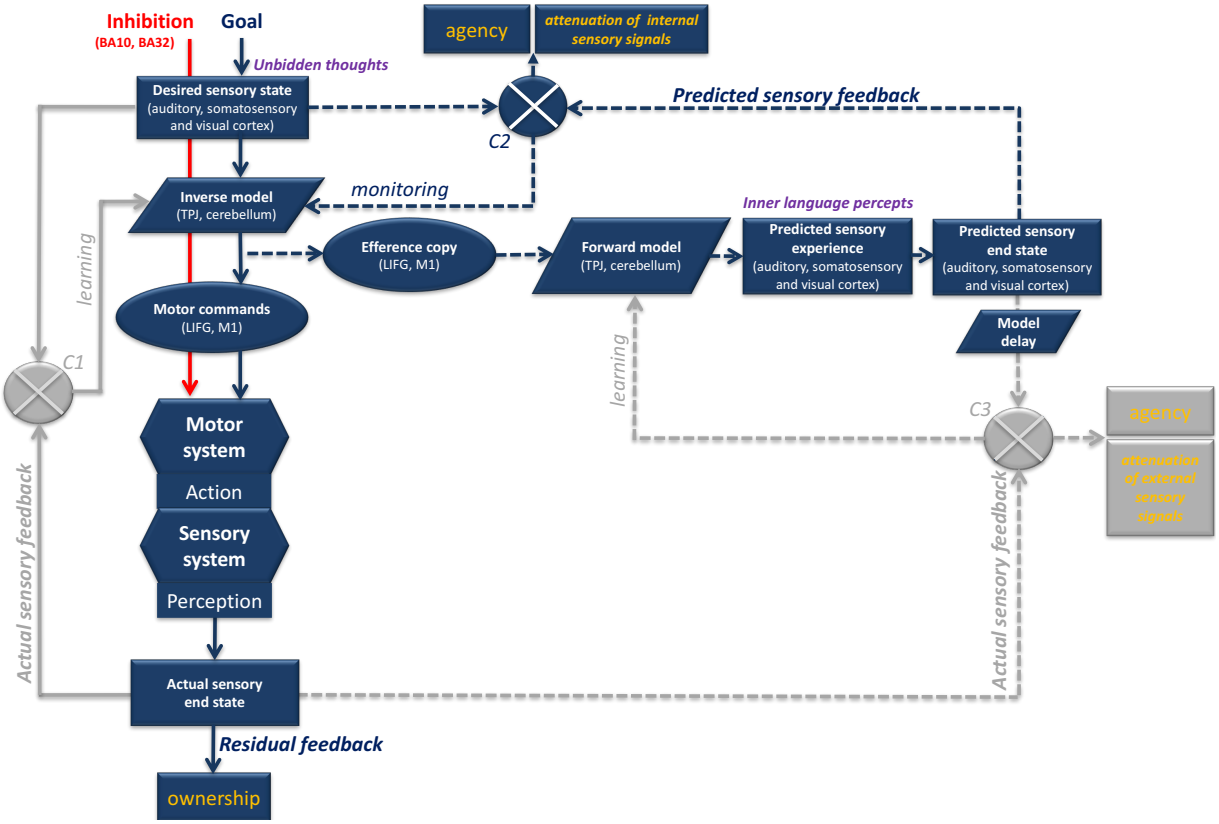
Figure captions

Figure 1. Our adaptation of the Predictive Control Account of inner speech. During overt speech, given a desired sensory state, an inverse model computes motor commands that are sent to the motor system, which produces speech movements and sounds. These are then processed by the sensory system, producing an actual sensory experience and resulting in an actual sensory end state. This actual sensory state provides a sense of ownership and can be compared with the desired state (C1) to improve the inverse model. In parallel, an efference copy mechanism takes place, depicted in dashed lines. A forward model predicts the sensory consequences of the motor commands. The predicted sensory feedback (or rather its end state) can be compared with the desired sensory state (C2) to adjust the motor commands, even before the action is executed. In addition, when the two states are close enough, a sense of agency is felt. The predicted sensory feedback, to which a delay is applied, is also compared with the actual sensory feedback (C3), to improve the forward model, and to further contribute to agency. During covert speech, the lines and boxes in light grey are irrelevant. In parallel with the motor commands, inhibitory signals (in red) are sent to the motor system, preventing actual articulator movement from occuring. A residual actual sensory feedback may still be experienced, giving rise to the sense of ownership. The predicted sensory signal computed by the efference copy mechanism yields inner language percepts: the inner voice heard and/or the inner articulation felt and/or the inner sign/gesture seen. Its end product is compared with the desired sensory state (C2) to adjust the motor commands while providing a sense of agency if the two states are sufficiently similar. TPJ, temporo-parietal junction; LIFG, left inferior frontal gyrus; M1, primary motor cortex.

Figure 2. A cerebral landscape of wilful covert word production with one own's voice. Lemma retrieval is handled by the left middle temporal gyrus. The lemma is converted to a lexeme, in a multisensory format, through two pathways, one for auditory representation (a) and one for somatosensory (b) representations. The auditory specification of the desired auditory state activates the left pSTG and STS, arrow 1a. The parallel somatosensory pathway activates the aSMG and S1, arrow 1b. An inverse model transformation then takes place, involving two pathways. The auditory specification is fed to the TPJ, arrow 2a. The somatosensory specification is sent to the cerebellum (arrow 2b). Motor programs are then specified: the transformed auditory goals are sent from the TPJ to the LIFG and to the left ventral premotor cortex, arrow 3a; the transformed somatosensory goals are sent from the cerebellum to the lower M1, arrow 3b. The motor programs issued by LIFG are themselves sent to M1 (arrow 4) integrating the two motor programs computed in the auditory and the somatosensory pathways. Articulation is inhibited, via a signal issued in rostral prefrontal cortex (BA 10) and anterior cingulate gyrus (BA 32) and sent to M1 only, or to both LIFG and M1. A residual somatosensory feedback may be felt (aSMG and S1), resulting from attenuated motor commands being sent to the motor system. The efference copy mediated by LIFG is sent to the TPJ (arrow 4a) and is inversed into a predicted auditory signal, activating pSTG and STS (arrow 5a). The other copy, in M1, is sent to the cerebellum (arrow 4b) and is inversed into a predicted somatosensory signal, activating aSMG and S1 (arrow 5b). C2 (between predicted and original desired states) takes place at two sites, in auditory and somatosensory cortices. pSTG, posterior superior temporal gyrus; STS, superior temporal sulcus; aSMG , anterior supramarginal gyrus; S1, primary somatosensory cortex; TPJ, temporo-parietal junction; LIFG, left inferior frontal gyrus; M1, primary motor cortex.

Preliminary version produced by the authors.
In *Inner Speech: New Voices.* Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. ISBN: 9780198796640

Figure 1



Preliminary version produced by the authors.
In *Inner Speech: New Voices.* Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. ISBN: 9780198796640

Figure 2