Neurolinguistics for Continuous Direct-Speech Brain-Computer Interfaces

Ciaran Cooney[1,*], Raffaella Folli[2], and Damien Coyle[1]

[1]Intelligent Systems Research Centre, Ulster University, Derry, Northern Ireland, United Kingdom

[2]Institute for Research in Social Sciences, Ulster University, Jordanstown, Northern Ireland, United Kingdom

*Correspondence: cooney-c@ulster.ac.uk

Abstract

A direct-speech brain-computer interface (DS-BCI) acquires neural signals corresponding to imagined speech, then processes and decodes these signals to produce a linguistic output in the form of phonemes, words or sentences. Recent research has shown the potential of neurolinguistics to enhance decoding approaches to imagined speech with the inclusion of semantics and phonology in experimental procedures. As neurolinguistics research findings are beginning to be incorporated within the scope of DS-BCI research, it is our view that a thorough understanding of imagined speech, and its relationship with overt speech, must be considered an integral feature of research in this field. With a focus on imagined speech, we provide a review of the most important neurolinguistics research informing the field of DS-BCI, and suggest how this research may be utilised to improve current experimental protocols and decoding techniques. Our review of the literature supports a cross-disciplinary approach to DS-BCI research, in which neurolinguistics concepts and methods are utilised to aid development of a naturalistic mode of communication.

1 Seeking a naturalistic form of communication through Direct-speech BCI

A direct-speech brain-computer interface (DS-BCI) is one that captures and decodes neural signals corresponding directly to speech production, enabling a naturalistic mode of communication (Iljina et al., 2017). Such a system has the potential to transform the lives of patients with severe motor dysfunction, including pathologies such as amyotrophic lateral sclerosis resulting in locked-in syndrome. Loss of verbal communication has a profound effect

30 on those inflicted, with loss of social interaction and the potential for isolation. In parallel with

31 this personal degeneration, a caregiver faces a more difficult challenge in ascertaining the needs

32 of the patient. These factors have played a crucial role in driving the development of DS-BCIs

33 (Brumberg et al., 2011; Oken et al., 2014).

34 It is our view that development of a functional DS-BCI must be predicated on imagined speech

35 (see section 3 for a detailed description) as the communicative modality. However, several

36 other types of speech have been utilised in experiments referenced throughout this text, making

37 it important to define their meanings. Table 1 is a categorisation of the different types of speech

38 typically used in DS-BCI experimentation. Three types of speech are presented, namely overt

39 (Blakely et al., 2008), intended (Guenther et al., 2009) and imagined (D'Zmura et al., 2009),

40 and these are subcategorised according to whether the speech is being produced or perceived

41 by a subject. Overt speech production results in an audible output that can be heard by the

42 person speaking, and by others within range of the sounds produced. Intended speech is the

43 name given to describe when a person tries to speak but does not have the capacity to produce

44 an audible output. Imagined speech is the internal pronunciation of words without any audible

45 output or associated movement. These are types of speech production and possible methods of

46 communication with DS-BCI. However, several studies have used decoding approaches

47 applied to the neural correlates of speech perception as evidence for the potential of decoding

48 speech processeses for communication (Di Liberto et al., 2015; Wang et al., 2018). We consider

49 it to be extremely important to distinguish speech perception studies from speech production

50 studies, and to be aware that the 'speech' in these studies refers to different phenomena. In

51 perception studies, the speech being considered is the stimulus provided by the experimenter.

52 The corresponding response of the subject, typically in the auditory cortex, is the neural activity

53 being decoded. This differs greatly from the study of speech production in which the subject is

54 actively producing phones, words or sentences, whether prompted or unprompted, with neural

55 correlates typically corresponding to brain regions associated with speech production.

56 Although speech perception studies are important for DS-BCI research, this review, is

57 primarily concerned with speech production, and in particular, imagined speech production.

58 **Table 1 Categorisation of types of speech typically used in DS-BCI experiments.**

	Production	Perception
Overt	Fully-articulated speech with audible output.	Active or passive hearing of audible speech (one's own

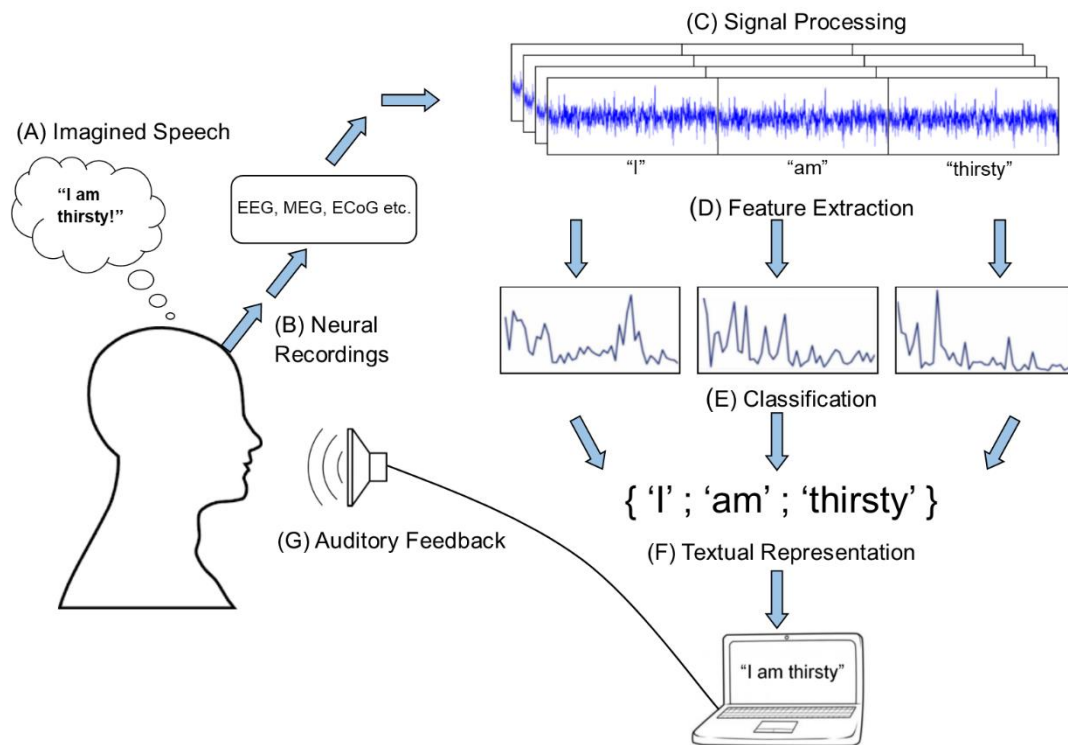|		speech or from another source).
Intended	Intention to produce overt speech but without the capacity to produce audible output.	Perception of one's own intended speech production.
Imagined	Internal pronunciation of words, independent of movement and without any audible output.	Perception of one's own imagined speech production.

59

60 A DS-BCI consists of several important stages (see Figure 1). The stages depicted in Figure
61 1B-G have each been extensively covered in the literature (Blakely et al. 2008, Guenther et al.
62 2009, reviewed in Bocquelet et al. 2017). However, there is relatively little consideration of
63 the difficulty in modelling the first of these stages (Figure 1A), namely imagined speech
64 production, during which a participant articulates words internally without any motor
65 movement. Neurolinguistics research is providing insight into the cognitive function,
66 phenomenology and neurobiology of speech production in general (Hickok, 2014), and
67 imagined speech in particular (Alderson-Day and Fernyhough, 2015; Perrone-Bertolotti et al.,
68 2014) and it is our view that these insights should be utilised within DS-BCI research. We
69 concur with the arguments expressed by Iljina et al. (Iljina et al., 2017) that, given the
70 complexity of speech production processes, combining research from the fields of BCI and
71 neurolinguistics must be seen as an important approach for those seeking to capture and decode
72 the phenomena.

73 Imagined speech is the internal pronunciation of words without any motor movement or
74 acoustic output (Torres-García et al., 2016) (see Section 3). Related, and overlapping,
75 terminology for imagined speech includes self-talk, sub-vocal/covert speech, internal
76 dialogue/monologue, sub-vocalisation, utterance, self-verbalisation and self-statement (Morin
77 and Michaud, 2007). However, for the purposes of performing controlled experiments in the
78 field of DS-BCI, it is necessary to maintain a consistent terminology and description of the
79 phenomena (see Section 3). Although not identical, there is overlap between imagined and
80 overt speech production, and imagined speech has become an alternative neuro-paradigm for
81 communicative BCI (D'Zmura et al., 2009; DaSalla et al., 2009; Deng et al., 2010). Such a
82 system differs from other types of communicative BCIs (Chaudhary et al., 2017; Pandarinath

83 et al., 2017), in that it relies on tapping directly into a person's speech production processes,
84 rather than using some unrelated neural activity as the method of communication.

85 Several DS-BCI studies have used neurolinguistics approaches within their experimental
86 procedures (González-Castañeda et al., 2017; Kim et al., 2013; Wang et al., 2011; Zhao and
87 Rudzicz, 2015). In general, the approaches used have been to design a constrained dictionary
88 of words categorised according to their relative semantic or phonological relationships. The
89 basic principle underpinning this approach is that the categorical features of a word may aid
90 decoding accuracy in imagined speech. There is some evidence that this is a valid approach to
91 take, particularly in relation to semantic categorisation, which has received greater attention in
92 the literature. Studies examining the feasibility of decoding semantic information from neural
93 signals have shown that semantic category can be predicted from brain activity (Kim et al.,
94 2013; Wang et al., 2011). However, further research is required to determine the true potential
95 of neurolinguistics research in relation to the neurobiology of imagined speech and the
96 structured processes underlying speech production, to inform DS-BCI research.

97 Here, we review trends in DS-BCI research, and the current understanding of speech
98 production processes, with an emphasis on imagined speech. We consider the potential
99 implications of attempting to harness neurolinguistics concepts and the limitations of working
100 directly with imagined speech. An argument is presented, that effective research in the field of
101 DS-BCI should incorporate neurolinguistics research and a thorough understanding of
102 imagined speech where possible to aid the development of a naturalistic mode of
103 communication.

(A) Imagined Speech

"I am thirsty!"

(B) Neural Recordings

EEG, MEG, ECoG etc.

(C) Signal Processing

"I"   "am"   "thirsty"

(D) Feature Extraction

(E) Classification

{ 'I' ; 'am' ; 'thirsty' }

(F) Textual Representation

(G) Auditory Feedback

"I am thirsty"

104

105   **Figure 1 Seeking a Naturalistic form of communication through direct-speech BCI.**

106   2   Trends in Direct-Speech BCI

107   The development of a 'silent' interface has long been an active area of research to enable users

108   to communicate without audible articulation of their speech. Several modalities have been

109   developed to facilitate such communication through movement-independent BCI, including

110   BCI-spellers (e.g. D'albis et al. 2012), BCIs based on steady-state visually evoked potential

111   (SSVEP) (e.g. Bin *et al.*, 2009) and BCIs based on motor imagery (e.g. Tabar and Halici,

112   2017a), (see   AlSaleh et al., 2016; Tabar and Halici, 2017b see for reviews). There are

113   numerous forms which these silent interfaces have taken to provide a more naturalistic,

114   language-based mode of communication, including ultrasound imaging of lip profiles (Denby

115   et al., 2006) and word recognition using magnetic implants and sensors (Gilbert et al., 2010).

116   However, approaches such as these require active motor skills that can be readily utilised as

117   the communicative modality and are therefore not movement-independent BCIs. The utility of

118   BCI as a mode for language-based communication has been noted by researchers for many

119   years (Denby et al., 2006; Donchin et al., 2000), with the concept for a DS-BCI being a

120   movement-independent BCI based on neural activity corresponding directly to imagined

121   speech production processes. However, the possibility of developing a BCI predicated purely

on imagined speech has only recently begun to gather momentum (Ikeda et al. 2014, Yoshimura et al. 2016, Nguyen, Karavas and Artemiadis 2017) as researchers have revealed promising results in attempts to classify units of imagined speech (González-Castañeda et al., 2017; Martin et al., 2014; Pei et al., 2011a; Yoshimura et al., 2016; Zhao and Rudzicz, 2015). There have been several incarnations of DS-BCIs, including a wireless BCI for real-time speech synthesis (Guenther et al., 2009) and a concept for continuous speech recognition (Herff et al., 2017). The current stream of DS-BCI research indicates a trend towards improved classification of imagined speech units for decoded brain activity (González-Castañeda et al., 2017; Martin et al., 2014) and the development of methodologies for continuous decoding of imagined speech (Brumberg et al., 2016). There have also been recent developments in classification of the neural correlates of speech perception (Di Liberto et al., 2015; Wang et al., 2018), one of which demonstrates real-time classification of auditory sentences from neural activity (Moses et al., 2018). Although this research is vital for the implementation of a closed-loop DS-BCI, it is important that results from speech perception studies are assessed independently of speech production studies as the neural activity corresponding to each cannot be assumed to have similar properties.

There have been notable successes in attempts to improve the decoding of language content directly from neural activity. The neural correlates of vowels and consonants (Idrees and Farooq, 2016; Pei et al., 2011b; Yoshimura et al., 2016), phonemes (Brumberg et al., 2011; Leuthardt et al., 2011), syllables (Deng et al., 2010), whole words (González-Castañeda et al., 2017; Martin et al., 2016) and even sentences (Herff et al., 2015) have all been evaluated using advanced decoding algorithms. Decoding of discrete units of speech, single vowels for example, has been a popular experimental paradigm in DS-BCI to date (Ikeda et al., 2014). Sereshkeh et al. (Rezazadeh Sereshkeh et al., 2017a) presented evidence suggesting that it is possible to classify units of imagined speech from electroencephalogram (EEG), presenting 63.2% ± 6.4 accuracy for pairwise classification tasks. Other studies have shown that decoding accuracies of vowels and consonants were similar for both overt and imagined speech (Pei et al., 2011a). Elsewhere, linguistic content has been harnessed to aid discrimination of both overt and imagined speech, with phonology (Zhao and Rudzicz, 2015), semantics (Kim et al., 2013) and syntax (Herff et al., 2015) each showing some potential to aid classification in DS-BCI. Figure 2, and the corresponding data in Table 2 categorise DS-BCI studies according to recording technique and the type of speech being investigated. The time-period for this analysis begins with the study of Blakely et al. (Blakely et al., 2008), due to this being the first study based on

the BCI paradigm depicted, and runs through to 2018. Criteria for inclusion in this analysis are those studies using typical recording techniques (EEG, electrocorticogram (ECoG), Micro-arrays, functional magnetic resonance imagining (fMRI) and functional near-infrared spectroscopy (fNIRS) to decode speech production (overt, imagined, intended), but not speech perception, directly from neural activity. Studies utilising speech imagery or imagined hearing have been excluded as we do not consider these modalities to be representative of the speech production required of a DS-BCI. The cross-sectional data (Figure 2A) indicates that studies have favoured two recording techniques and two types of speech. Clearly, EEG and ECoG are the most dominant recording techniques, having been cited in 16 and 20 studies respectively (Figure 2B). The likely reason being the high temporal resolution (milliseconds) they both possess, particularly in comparison to imaging techniques such as fMRI (with temporal resolution in the order of seconds). This high temporal resolution is required to capture the dynamic processes associated with speech production (Herff et al., 2016). As a non-invasive recording technique, EEG makes recruitment of experimental participants easier, but the greater spatial resolution of ECoG render it a better candidate for decoding imagined speech signals when participants are made available due to treatment for pre-existing medical conditions (e.g. epilepsy) (Martin et al., 2016). Although they have shown good performance in fields such as neuromotor prostheses (e.g. Hochberg et al. 2012), relatively few studies have utilised microelectrode arrays for recording the spiking activity of single or multiple units (SU or MU) i.e., neurons, during imagined speech. However, the SU or MU offer the required signal specificity to improve imagined speech decoding processes given its success in movement and movement intention decoding (Bouton et al., 2016).
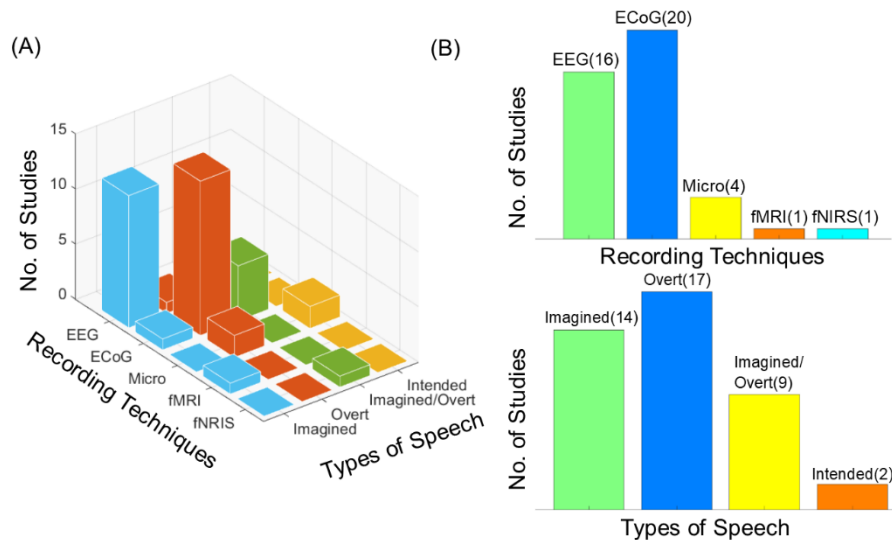
Figure 2 Direct-speech BCI studies categorised according to recording techniques and types of speech.

It is clear from the data presented in Figure 2 that overt speech production is heavily-utilised in experimental trials. Overt speech is included in a total of 26 studies (17 solely overt and 9 alongside imagined speech) (Figure 2B). There are several reasons for this trend, including the lack of behavioural verification associated with imagined speech, whereby it is difficult to confirm whether experimental tasks have been performed correctly, and the lower amplitude of EEG/ECoG signals it produces (Palmer et al., 2001; Shuster and Lemieux, 2005). Despite lower amplitude signals, there is evidence to suggest that EEG can provide considerable information on imagined speech that can be utilised for a DS-BCI (D'Zmura et al., 2009). Attempts to decode continuous overt speech have been made (Herff et al., 2015), and it is anticipated that further developments may enable adaptation of this approach for imagined speech. As stated, the use of overt speech is prevalent in DS-BCI research. However, if a truly naturalistic form of communication is to be achieved using imagined speech, then a thorough understanding of the phenomena is required.

193   **Table 2 Overview of DS-BCI studies attempting to decode speech from neural activity.**

Reference	Recording Technique	Type of Speech	Experimental Paradigm
Blakely et al., 2008	Micro-electrode	Overt	Phoneme pronunciation.
D'Zmura et al., 2009	EEG	Imagined	Imagined speech of two syllables spoken in one of three rhythms.
Guenther et al., 2009	Micro-electrode	Intended	Vowel production involving movement from a central vowel location to one of three peripheral vowel locations.
Porbadnigk et al., 2009	EEG	Imagined	Five words, presented in block, sequential or random order.
Brigham and Kumar, 2010	EEG	Imagined	Imagined speech of two syllables, /ba/ and /ku/ at two rhythms.
Deng et al., 2010	EEG	Imagined	Imagined speech of two syllables spoken in one of three rhythms.
Kellis et al., 2010	Micro-electrode	Overt	Repetition of one of ten words.
Brumberg et al., 2011	Micro-electrode	Intended	Intended production of 38 American English phonemes.
Chi et al., 2011	EEG	Imagined	Generation of five types of phonemes that differ in their manner vocal articulation.
Leuthardt et al., 2011	ECoG	Overt/ Imagined	Overt and imagined phoneme articulation.
Pei et al., 2011a	ECoG	Overt/ Imagined	Overt and imagined repetition of 36 monosyllabic words.
Wang et al., 2011	ECoG	Overt	Three language tasks based on picture-naming.
Pei et al., 2011b	ECoG	Overt/ Imagined	Word repetition using overt or covert speech in response to visual or auditory stimuli.
Derix et al., 2012	ECoG	Overt	Spontaneous speech in non-experimental setup.

Herff et al., 2012	fNIRS	Overt/ Imagined	Utterances produced in auditory, silent and imagined speech.
Zhang et al., 2012	ECoG	Overt	Articulation of Chinese sentences.
Kim et al., 2013	EEG	Overt/ Imagined	Speech of monosyllabic Korean words representing two categories of meaning (number and face).
Bouchard and Chang, 2014	ECoG	Overt	Reading of consonant-vowel syllables.
Derix et al., 2014	ECoG	Overt	Spontaneous speech in non-experimental setup.
Ikeda et al. 2014	ECoG	Imagined	Imagined speech production of three Japanese vowels.
Kanas et al., 2014	ECoG	Overt	Two syllable repetition tasks.
Martin et al., 2014	ECoG	Overt/ Imagined	Overt and covert reading of short-stories.
Mugler et al., 2014a	ECoG	Overt	Overt speech used to identify different phonemes by where they place in different words.
Mugler et al., 2014b	ECoG	Overt	Overt speech used to identify different phonemes by where they place in different words.
Song and Sepulveda, 2014	EEG	Overt/ Imagined	High tone production in overt, inhibited and imagined speech.
Herff et al., 2015	ECoG	Overt	Reading from well-known texts.
Iqbal et al., 2015a	EEG	Imagined	Imagined speech of vowels /a/ and /u/, and no action.
Iqbal et al., 2015b	EEG	Imagined	Imagined speech of vowels /a/ and /u/, and no action.
Lotte et al., 2015	ECoG	Overt	Reading from well-known texts.

Zhao and Rudzicz, 2015	EEG	Overt/ Imagined	Imagined speech production of seven phonemes and two pairs of phonologically-similar words.
Herff et al., 2016	ECoG	Overt	Recitation of a presented sentence.
Martin et al., 2016	ECoG	Overt/ Imagined	Overt and imagined speech production of words selected to maximise variability of number of syllables and semantic category.
Yoshimura et al., 2016	EEG/fMRI	Imagined	Imagined speech production of Japanese vowels /a/ and /i/.
González-Castañeda et al., 2017	EEG	Imagined	Imagined speech production of five Spanish words.
Nguyen et al., 2017	EEG	Imagined	Imagined speech of short words, long words and vowels.
Ramsey et al., 2017	ECoG	Overt	Overt speech production of four phonemes.
Rezazadeh Sereshkeh et al., 2017a	EEG	Imagined	Imagined speech repetition of the words "yes" or "no".
Rezazadeh Sereshkeh et al., 2017b	EEG	Imagined	Imagined speech repetition of the words "yes" or "no".
Fargier et al., 2018	EEG	Overt	Overt word production corresponding to presented pictures.
Hashim et al., 2018	EEG	Imagined	Imagined speech word production.
Ibayashi et al., 2018	ECoG	Overt	Overt speech of 15 Japanese syllables.
Livezey et al., 2018	ECoG	Overt	Overt speech of 57 different consonant-vowel syllables.

194

3 Imagined Speech: A Special Case of Speech

3.1 The phenomena of Imagined Speech

As mentioned above, many definitions for imagined speech are present in the literature (Alderson-Day and Fernyhough, 2015; Hirshorn and Thompson-Schill, 2006), one of which refers to it as the internal pronunciation of words without emitting sounds or making facial movements (Torres-García et al., 2016). Research has demonstrated that imagined speech involves many cognitive functions including learning (Alderson-Day and Fernyhough, 2015), task-production (Dolcos and Albarracin, 2014) and memory (Perrone-Bertolotti et al., 2014).

Despite its central position in everyday life, imagined speech has been the subject of relatively little research. Behavioural evidence has indicated that imagined speech is provided by the motor system's prediction of sensory actions (corollary discharge) (Scott et al., 2013) and it has been suggested that imagined speech is produced in much the same way as overt speech, without the motor-based articulation which generates auditory output (Oppenheim and Dell, 2010). Martínez-Manrique and Vicente (Martínez-Manrique and Vicente, 2015) support an "activity" view of imagined speech, in which the phenomena does not have a "proper function" in cognition but has simply inherited its suite of functions from overt speech.

Other studies have characterised imagined speech as the basis for rehearsal in short-term memory (Baddeley et al., 1975) and as having a phonological influence in reading and writing (Oppenheim and Dell, 2008). Further studies concur with these findings, suggesting that inner rehearsal is a central tenet of imagined speech within the phonological loop, i.e. the temporary storage of information in short-term memory (Perrone-Bertolotti et al., 2014), and that imagined speech may interact with working memory to enhance the encoding of new material (Marvel and Desmond, 2012). It has been suggested that imagined speech serves a regulatory role in social speech communication, meaning that it is utilised in overt speech communications (speaking and listening), as well as being implicated as part of a covert articulatory planning process within the speech-motor processing paradigm (see Price (2012) for review).

It has been proposed that imagined speech may be used to generally represent, maintain, and organise task-relevant information and conscious thoughts (Dolcos and Albarracin, 2014). Although not normally associated with executive control processes, the role of imagined speech in task switching, for example, switching attention across multiple arithmetic problems, has been studied (Emerson and Miyake, 2003). The difficulties associated with studying imagined

226 speech in experimental research has led to the use of overt speech as a proxy for the phenomena
227 in DS-BCI research (e.g. Martin et al., 2014; Pei et al., 2011b). Therefore, it is useful to have
228 a clear picture of the relationship between the two types of speech.

229 **3.2 The relationship between overt and imagined speech production**

230 The relationship between overt speech and imagined speech has been extensively debated
231 (Brocklehurst and Corley, 2011; Corley et al., 2011; Oppenheim and Dell, 2010, 2008), though
232 at present there is no definitive position on the precise nature of this relationship. Here, we
233 present the evidence for a close relationship between overt and imagined speech, before
234 considering the ways in which the two differ. Finally, we discuss the implications of this
235 relationship for DS-BCI research.

236 It has been posited that imagined speech is a truncated form of overt speech, in that the stages
237 of production are the same for both, prior to the articulatory effects associated with overt speech
238 (Oppenheim and Dell, 2010). Subjective accounts of imagined speech indicate that it resembles
239 overt speech in tempo, pitch and rhythm (MacKay and others, 1992) and studies have found
240 that imagined speech retains deep-lying features such as lexical and semantic information
241 (Oppenheim and Dell, 2008). The *motor simulation* hypothesis places overt and imagined
242 speech on a continuum, on which linguistic mechanisms and physiological correlates are shared
243 (Perrone-Bertolotti et al., 2014), albeit with features attenuated in imagined speech (Alderson-
244 Day and Fernyhough, 2015). Importantly, the motor simulation hypothesis assumes that
245 imagined speech necessarily includes fully-specified articulatory detail (e.g. Levelt, 1989),
246 merely lacking observable sound and movement.

247 Phonemic-similarity (in which mistaken phonemes are replaced with similar phonemes) has
248 been observed with similar magnitudes for both overt and imagined speech production
249 (Brocklehurst and Corley, 2011) and further findings suggest that imagined speech is specified
250 at the sub-phonemic level, and that its process of production must be similar to that of overt
251 speech (Corley et al., 2011). The implication here is that imagined speech does contain much
252 of the featural richness associated with overt speech, a view fully compatible with evidence
253 that phonological representations are fully-encoded in imagined speech. Imagined speech has
254 been considered part of an overall speech production system, in which it is used for predictive
255 simulation or "forward models" of linguistic representations, suggesting that it is produced in
256 much the same way as overt speech, minus overt articulation (Levelt et al., 1999).

There is considerable overlap between the neurobiology of overt and imagined speech (Marvel and Desmond, 2012), with neural activations in typical left-hemispheric language regions, in general, being associated with both (Basho et al., 2007; Huang et al., 2002; McGuire et al., 1996a; Palmer et al., 2001) (see section 3.3, below). Activation of Broca's area during imagined speech indicates that this typical language region is associated with its production, and is consistent with results from functional imaging studies examining silent articulation (Paulesu et al., 1993). fMRI results have shown activation of the supplementary motor area (SMA), inferior frontal gyrus (IFG) and insula during phonological processing of imagined and overt speech (Aleman et al., 2005). Furthering current understanding of the neuroanatomy and neural correlates of imagined speech production is an important aspect of research in this field.

Although they suggest that there is significant overlap between overt and imagined speech, Oppenheim and Dell (Oppenheim and Dell, 2008), also advise that imagined speech is impoverished at the featural level and thus abstract and underspecified. It has been suggested that imagined speech is often attenuated at the surface level, lacking phonological (Oppenheim and Dell, 2008) or phonetic (Wheeldon and Levelt, 1995) detail. Countering the view that imagined speech is intrinsically similar to overt speech, the *abstraction hypothesis* contends that imagined speech is produced as a consequence of activation of abstract linguistic representations (e.g. Indefrey and Levelt, 2004). The theory states that imagined speech is activated before the speaker retrieves any articulatory information, and therefore should not require any motor activations. There are several arguments in favour of the abstraction view (summarised in Oppenheim and Dell, 2010), first of which is that imagined speech is produced faster than overt speech, suggesting that imagined speech is abbreviated in some respect (e.g. MacKay and others, 1992), and thus lacks the articulatory properties associated with overt speech. Another argument is that attenuated activity in language-related brain regions during imagined speech indicates that the processes of production are not as complete as in overt speech. The third argument presented is that imagined speech does not require articulatory abilities and so articulation is not required for complete use of imagined speech. The authors also observe that articulatory suppression does not necessarily eliminate imagined speech. Moreover, imagined speech does not (necessarily) translate to overt speech performance. Theoretically, were overt and imagined speech to involve similar planning processes, then it would be reasonable to expect practice of an utterance in one form of speech to improve performance in the other. However, evidence has indicated that this is not the case (Corley et al., 2011).

290   Alternatively, the *flexible abstraction hypothesis* states that there is a single form of imagined

291   speech, which is represented at the phonemic-selection level (Oppenheim and Dell, 2010). The

292   hypothesis states that representations can be modulated by articulation to include more explicit

293   features, and the authors suggest that cases where imagined speech appears to have

294   phonological features may be caused by participants deploying a form of imagined speech

295   involving a greater degree of articulation. The flexible abstraction hypothesis suggests that

296   imagined speech may fail to involve articulatory representations but it can incorporate lower

297   level articulatory planning when speakers silently articulate. The *surface-impoverished*

298   *hypothesis* states that imagined speech is impoverished at the surface level, having weaker

299   lower-level representation (e.g. featural level), and the *deep-impoverished hypothesis* states

300   that imagined speech represents sounds and gestures, but not higher level information

301   (Oppenheim and Dell, 2008). Imagined speech may be formed as a featurally-abstract forward

302   model (Pickering and Garrod, 2013), and phonological features may be experienced due to the

303   sensory prediction created (Scott, 2013). Imagined speech may also vary depending on

304   cognitive and emotional conditions, causing changes between abstract and concrete forms

305   (Fernyhough, 2004).

306   As stated above, neuro-anatomical overlap between regions associated with overt and imagined

307   speech has been observed. Nevertheless, there are significant differences in brain activity

308   between the two processes (e.g. Basho et al. 2007). For example, fMRI has discovered that

309   imagined speech elicits greater activation in several areas of the brain (e.g. Basho et al. 2007)

310   and a lesion symptom mapping (LSM) study of patients with aphasia showed that participants

311   with poor overt speech retained relatively strong imagined speech in comparison (Stark et al.,

312   2017), suggesting a dissociation of the cognitive mechanisms generating overt and imagined

313   speech. Previous work with aphasics, indicating that imagined speech abilities were more

314   effected by lesions to the left pars opercularis than overt speech production, led Geva, Jones et

315   al. (Geva et al., 2011b) to state that imagined speech cannot be assumed to be overt speech

316   without a motor component. For further information on the neuro-biology of imagined speech,

317   see section 3.3.

318   Perrone-Bertolotti et al. (Perrone-Bertolotti et al., 2014) astutely observe that the variance in

319   results between overt and imagined speech experiments may, at least partially, be explained by

320   the different speech tasks involved in the studies. Word repetition, object naming, verb

321   generation, etc., all require different speech production processes and thus engage different

322   areas of the brain. It is also conceivable that differences between the two types of speech could

323   be put down to participants being better able to perceive certain types of error in overt speech.

324   Perrone-Bertolotti et al. (Perrone-Bertolotti et al., 2014) also suggest that differing results may

325   indicate that imagined speech consists of flexible subtypes or levels, and that the experimental

326   paradigm may be partially responsible for the differences observed between the two types of

327   speech.

328   Clearly, there is no definitive description of the precise relationship between overt and

329   imagined speech, and this is a subject that requires further elucidation from neurolinguistics

330   research. We agree with Martínez-Manrique and Vicente (Martínez-Manrique and Vicente,

331   2015), that a comprehensive view of imagined speech will require precise models of linguistic

332   production and comprehension, and a cognitive account will require more data than is currently

333   available. Therefore, we must also agree with Geva, Jones et al. (Geva et al., 2011b) that overt

334   speech cannot simply be assumed to be a reliable substitute for imagined speech. It is our

335   contention, in relation to DS-BCIs, that it is not possible to reliably infer performance in an

336   imagined speech paradigm from results obtained during overt speech experiments. This is not

337   to say that there is no value in overt speech paradigms, and given that there is much overlap in

338   the linguistic theory and neurobiology associated with both, there is certainly a lot to be gained

339   from such experiments. However, as the communicative paradigm for an eventual operational

340   DS-BCI is imagined speech, we must emphasise the importance of utilising this modality, when

341   possible, in experimental protocols.

342   **3.3   The neuroanatomy of imagined speech**

343   Alderson-Day and Fernyhough (Alderson-Day and Fernyhough, 2015) suggest that a *prima*

344   *facie* assumption about the neural correlates of imagined speech might be that they closely

345   resemble an attenuated version of the neural activity associated with overt speech. There is

346   evidence supporting activation in Broca's area, SMA and parts of the prefrontal cortex, having

347   been observed during both overt and imagined speech (see Price, 2012 for review). Studies

348   have shown that overt and imagined speech do produce similar neural activations, with the

349   exception of certain motor-related activity associated with overt speech (Palmer et al., 2001),

350   and that the blood oxygen level-dependent (BOLD) response measured from fMRI recordings

351   was greater during overt than imagined speech (Shuster and Lemieux, 2005). However, the

352   neuro-anatomy of imagined speech has been shown to differ from that of overt speech (e.g.

353   Basho et al. 2007). It is important to identify the regions specifically correlated with imagined

354   speech in the context of development of a DS-BCI that are independent of movement and

therefore not overt speech production, and are independent of stimuli and therefore not speech perception.

Reports on the anatomical underpinnings of imagined speech have consistently implicated the left inferior frontal gyrus (LIFG) as the anatomical basis for the phenomena (Aleman et al., 2005; McGuire et al., 1996a, 1996b; Shergill et al., 2002) (see Figure 3 (Berwick et al., 2013)). Positron Emission Topography (PET) has attributed LIFG activation to imagined speech during sentence and single-word production (McGuire et al., 1996b) and fMRI was used to observe LIFG activation during imagined sentence production (Shergill and Bullmore, 2001; Shergill et al., 2002). In the second of these fMRI studies (Shergill et al., 2002), the LIFG, along with other regions, was associated with increased activation corresponding to increased rates of imagined speech production. The region has also been associated with increased activation during dialogic, in comparison with monologic, imagined speech (Alderson-Day et al., 2015). Morin and Michaud (Morin and Michaud, 2007) note that the LIFG exhibits functional heterogeneity, observing that its most anterior parts (Brodmann's Area (BA)45) are involved in word retrieval and their associated meanings, while the posterior (BA46/47) specialises in accessing words through an articulatory code (Paulesu et al., 1997). It has been observed that task-elicited imagined speech results in increased activation in the LIFG, in comparison with spontaneous imagined speech (Hurlburt et al., 2016). The authors suggest that activation of LIFG during task-elicited imagined speech may be a reflection of elicitation tasks rather than the speech itself, as the LIFG is thought to be integral to planning and execution of hierarchical sequences.

Among regions most often observed as corresponding to imagined speech production are SMA (Shergill and Bullmore, 2001; Shergill et al., 2002), insula (Aleman et al., 2005), premotor cortex (McGuire et al., 1996a), STG, and middle temporal gyrus (MTG) (Shuster and Lemieux, 2005). The SMA, left precentral gyrus and the right inferior parietal lobe are all associated with increased activation at slower rates of imagined speech production (Shergill et al., 2002). The SMA has also been associated with sentence-repetition tasks (Shergill and Bullmore, 2001) and phonological processing during imagined speech (Aleman et al., 2005). The insula has been implicated in multiple studies reporting on imagined word production (Aleman et al., 2005; Hubbard, 2010; McGuire et al., 1996a; Shergill and Bullmore, 2001) but may not be representative of imagined speech given that it is often associated with imagined hearing (see below) and overt speech. However, Shuster and Lemieux (Shuster and Lemieux, 2005) observed that many studies which have failed to report involvement of the insula in speech

388 production have typically only used imagined or silently-articulated speech (Wildgruber et al.,
389 2001).

390 Increased activation has been observed in the left MTG and STG during the production of
391 multisyllabic words in imagined speech trials (Shuster and Lemieux, 2005) and the posterior
392 STG has been implicated in metric stress evaluation in the phonological loop (Aleman et al.,
393 2005) (see Figure 3). Interestingly, the left MTG and STG are often associated with increased
394 activity during trials involving imagined hearing or dialogic imagined speech (see Alderson-
395 Day and Fernyhough, 2015 for review). This type of task, in which a participant is asked to
396 imagine hearing speech in another person's voice, is thought to rely on memory for
397 phonological information (Alderson-Day and Fernyhough, 2015), and to activate the primary
398 auditory cortex (Heschl's gyrus) (Hurlburt et al., 2016). Other findings indicate that dialogic
399 imagined speech draws from a range of regions beyond a typical left-sided perisylvian language
400 network, including the right IFG, right MTG and the right STG/STS (Alderson-Day et al.,
401 2015). The precuneus, posterior cingulate, left insula and cerebellum are also implicated. The
402 dorsal pathways between BA44 and the posterior superior temporal cortex (pSTC) subserve
403 higher-order hierarchical sequences and thus support core syntactic processes (Friederici,
404 2018), whereas the ventral pathways, including between BA45 and the temporal cortex (TC),
405 support processing of semantic and conceptual information (Berwick et al., 2013).

406 Hurlburt, Heavey and Kelsey (Hurlburt et al., 2013) state that both production, and perception,
407 of imagined speech exhibit activations in regions such as the IFG, SMA, insula and posterior
408 STG (Hubbard, 2010; Price, 2012). Although there certainly appears to be overlap between
409 imagined speech and imagined hearing, they are, in general, anatomically separable. Imagined
410 speech is typically associated with left-hemispheric regions, including the LIFG, insula and
411 STG (McGuire et al., 1996a), whereas imagined hearing corresponds to a bilateral network
412 with activation of SMA, posterior parietal cortex, STG and MTG (Zatorre and Halpern, 2005).
413 It has been suggested that differences between the two conditions may be the result of
414 additional motor elements of imagined speech which involve the deployment of a
415 somatosensory forward model (Tian and Poeppel, 2013).

416
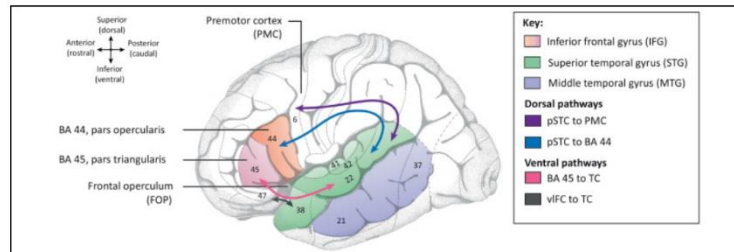
Figure 3 Neuroanatomical regions associated with imagined speech production.

Concerns have been raised surrounding the ecological validity of findings on the neural components of imagined speech (Alderson-Day and Fernyhough, 2015). Paradigms are often simple word or sentence-repetition tasks, ignoring the complexity of imagined speech (Jones and Fernyhough, 2007). Although experiments such as these are a common approach in language studies, it is our view that further studies examining spontaneously-produced speech (Derix et al., 2014, 2012; Ruescher et al., 2013), and imagined speech (Hurlburt et al., 2016), are required to provide greater elucidation of the neural underpinnings of the phenomena. It is also important to note that, as well as general activations associated with imagined speech production, processing of complex lexical, phonological, semantic (Basho et al., 2007) or word retrieval (Hirshorn and Thompson-Schill, 2006) tasks correspond to additional activity in the inferior frontal cortex (IFC) of the left hemisphere. We concur with Bocquelet et al. (Bocquelet et al., 2017) that neuro-anatomical findings indicate high-level processing of imagined speech requires left-lateralisation.

Information on the neuroanatomical regions associated with imagined speech production is enhanced by consideration of the characteristics of the corresponding neural activations, and in particular, the frequency bands that may provide the most discriminable content. Activations

in the beta band above Broca's area and the frontal cortex have been associated with imagined speech production (Rezazadeh Sereshkeh et al., 2017b). In one study, increased activity was observed in EEG channels located close to Broca's area in the frequency range of 20-30 Hz, whereas activity in Wernicke's area appeared primarily below 15 Hz (Nguyen et al., 2017). This may indicate that separate frequency bands contain information relating to different speech production processes. In the same study, the authors use evidence from the classification of short versus long words to suggest that differences in the complexity of words could create discriminative features across frequency bands. In an imagined speech yes/no classification task, no discriminative difference was detected in the delta, theta, alpha and mu rhythms. However, in the higher frequency ranges (beta and gamma), a discriminative pattern was associated with typical left-sided speech regions (Rezazadeh Sereshkeh et al., 2017b).

MEG measurements obtained during a silent reading task showed event-related desynchronization (ERD) in the alpha and beta bands over Broca's area (Goto et al., 2011). The results of an ECoG study into imagined speech vowel articulation suggested that signals in the alpha (8-13 Hz) and beta (14-30 Hz) bands over Broca's area may contain information about the articulatory code of single vowels but not about segmentation of a phoneme sequence (Ikeda et al., 2014). Clearly, the recording technique employed impacts the frequency ranges that can be analysed. For example, filtering imagined speech EEG data between 3 and 20 Hz, (Deng et al., 2010) found considerable energy in the alpha band (8-14 Hz), whereas using ECoG has allowed researchers to obtain features from the high gamma (70-150 Hz) band (Martin et al., 2016), which is useful for its association with spike rate and local field potential, and its reliable tracking of rapid neural fluctuations during speech perception and production (e.g. Pei, Barbour, et al. 2011). It is our view that this information on the important frequency bands associated with imagined speech can aid decoding approaches in future research. However, it is also important that further research in this area is undertaken so that a detailed and accurate picture of the spatial-temporal-spectral correlates of imagined speech is developed.

In section four, we extend our analysis on the neuroanatomical underpinning of imagined speech to include current understanding of speech production processes and the anatomical regions-of-interest they correspond to.

4 How is (Imagined) Speech Produced?

4.1 Models of Speech Production

It is a matter of consensus in psycholinguistic research that speech production is planned across multiple hierarchically organised levels of analysis (Hickok, 2012) and that word production involves at least two stages of processing: a lexical and a phonological stage (Levelt et al., 1999) (Figure 4B). Models of speech production can differ in terms of the number of distinct stages involved (Hickok, 2014, 2012; Levelt, 1999; Levelt et al., 1999), but there is general agreement that it involves a staged, hierarchical process with a temporal structure, as indicated by the models in Figure 4.

According to Levelt (Levelt, 1999), spoken word production includes lexical selection, lemma retrieval, morphological and phonological code retrieval, and is completed with articulation (Figure 4A). Models of speech production typically begin with an input from the conceptual system, i.e. the message to be expressed (Levelt, 1999). This is then mapped to a corresponding lexical representation, encoding properties such as grammatical features but not a phonological form. Following selection of a lemma, the morphological stage bridges the gap between the conceptual domain and the phonological or articulatory domain. Phonetic encoding and articulation, seen in Figure 4A, are stages of the speech production process concerned with acoustic output. The speech production models, as stated here, are based primarily on work in the fields of motor control and psycholinguistics, and it has been noted that linguistic models are currently constrained by the need for further developments in neuroscience (Hickok, 2012). EEG studies have been used to study the time courses associated with the processing stages in word production (see Indefrey 2011 for review). Following analysis of several event-related potential (ERP) studies, Indefrey (Indefrey, 2011) presented the following estimated onset times and durations for overt speech production: conceptual preparation (0-200ms), lemma retrieval (200-275ms), phonological code retrieval (275ms onset), syllabification (355ms onset; 20ms per phonemes, 50-55ms per syllable), phonetic encoding (455ms onset) and articulation (600ms) (Figure 4A). Although this research is based on overt speech, and the articulation stage is not relevant, the estimated timings can be informative for DS-BCI researchers seeking to target a specific stage of the production process during signal decoding.
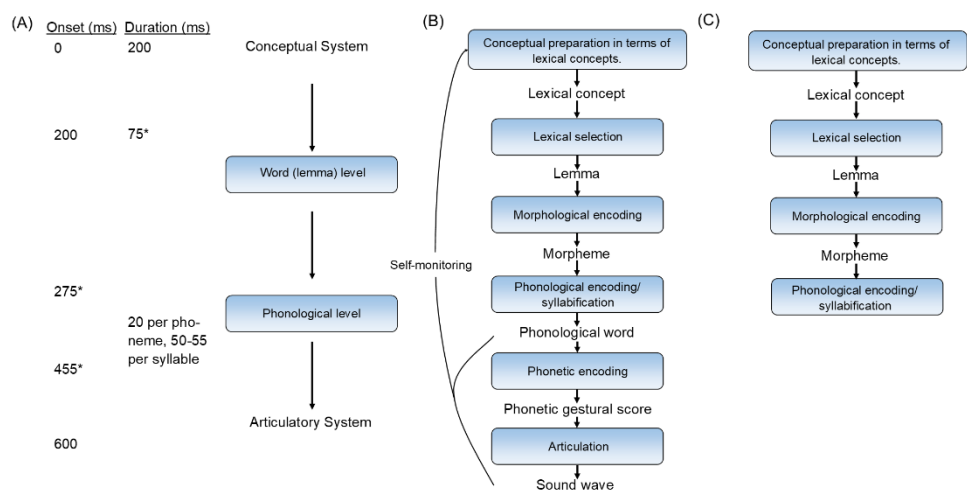
Figure 4 Speech production models with estimated time courses.

Language production involves multiple levels of representation and this modular system incorporates various sub-systems, i.e. semantics, syntax and phonology. Different brain regions in the left and right hemispheres have been identified as supporting these language functions, with syntactic processing supported by networks involving the temporal cortex and inferior frontal cortex, and less lateralised temporo-frontal networks subserving semantic processing (see Friederici, 2011). In discussing Hebbian theory, Pulvermüller (Pulvermüller, 1999) considers whether lexical or semantic distinctions reflect differences that are biologically real, using it to explain the observation that word meanings can be mapped to different cortical regions, for example. This results in words that are distinguished on the basis of linguistic criteria being represented differently in the brain. Investigations into the neural correlates of language function and competence commonly employ functional imaging approaches (see Indefrey and Levelt, 2004), as well as LSM to determine the links between linguistic pathologies and corresponding lesion sites in aphasics (Bates et al., 2003). Linguistic research can be considered within the context of several modular domains, four of which (semantics, lexical access, syntax and phonology) are discussed in the following sections.

4.2 Semantics and the meaning of words

Semantic knowledge has been referred to as the ability to assign and use the meaning of words, relying on both stored semantic knowledge and executive control to enable semantic activation in line with goals and constraints (Whitney et al., 2012). The term semantics refers to the meaning of a word or collection of words. In the models of speech production in Figure 4, semantic information forms part of the conceptual stage in which a message to be expressed is conceived. This conceptual stage precedes lexical selection, syntactic encoding and phonological encoding, with the process leading up to selection of a lexical concept referred to as "conceptual preparation." Mapping between the semantic concept to be expressed and a lexical formulation of this message is not a simple one-to-one process, as there are often multiple ways to refer to a single concept (e.g. a car may be referred to as a vehicle, saloon, motorcar, etc.) (Levelt et al., 1999).

Semantic comprehension studies indicate that semantic operations are normally slower to develop and longer lasting than syntactic operations (Piñango et al., 2006) and thus accommodate slower lexical activation than syntactic dependencies (Love et al., 2008). However, it cannot simply be assumed that the relationship between semantic and syntactic comprehension is mirrored in speech production processes. One study has posited the possibility of an intermediate layer between semantics and phonology due to the arbitrary nature of the mapping from meaning to sounds, i.e. words with similar meanings do not tend to have similar sounds associated (Lambon Ralph et al., 2002), and the Hebbian associationist model predicts that semantic differences between word categories generate patterns of neural activity reflective of those differences (Pulvermüller, 1999). For example, naming of living versus inanimate objects was more strongly correlated with integrity of the middle temporal cortex (MTC), while both categories showed significant overlap in the frontal cortex (Henseler et al., 2014). Additionally, large parts of the IFG appear to be involved in semantic differentiation of verbs versus nouns. Activation in the LIFG is typically exhibited when difficult semantic relationships, such as the meaning of ambiguous words (e.g. words such as break, light and head have multiple meanings) within a sentence, need to be parsed. These difficult relationships may be weak or unusual associations, an increased number of response options or competition among potential targets in a semantic network (Badre et al., 2005). Although many neuroimaging studies have concentrated on the LIFG as the basis for semantic processing and control, other studies show that damage to a wide distribution of brain regions results in impairment of semantic control (Whitney et al., 2012). The orbital IFG exhibited

23

542   higher correlation with the semantic differentiation of nouns, whereas a more posterior,
543   triangular/opercular part of the IFG was associated with the impaired differentiation of verbs.
544   Results from action word studies have indicated that semantic processing can engage many
545   different cortical areas, with Pulvermüller (Pulvermüller, 2005) stating that this contradicts the
546   view that processing of meaning is concentrated in a single cortical location. Moreover, it has
547   been demonstrated that word class-distinctions can be made in relation to different types of
548   action words (Hauk et al., 2004), with different cortical activations associated with the muscles
549   used to perform a given action, the complexity of the movement and the number of muscles
550   involved (Pulvermüller, 1999).

551   **4.3   Lexical access maps meaning to words**

552   Lexical access is the process that facilitates access to the words retained in memory that are
553   required for language production. Dell, Martin and Schwartz (Dell et al., 2007) present a two-
554   step model of lexical access in which a network consists of a semantic layer connected to words,
555   and words connected to a phoneme layer. Word retrieval begins when the semantic features of
556   an intended word are activated. This activation proceeds through the network resulting in the
557   selection of the most active word from a grammatical category. A phonological retrieval stage
558   begins with the activation of this selected word.

559   Lexical access effects the fluency and speed at which speech is produced. For example, it has
560   been shown that function words (i.e. contributing to syntax/grammar) are accessed faster than
561   content words (i.e. contributing to information/meaning), independent of perceptual
562   characteristics (Segalowitz and Lane, 2004). Another factor influencing lexical fluency is the
563   frequency with which a word is used (Mohr et al., 1996). In a picture-naming paradigm,
564   participants displayed quicker response-times in object-naming tasks than they did in action-
565   naming tasks, leading the authors to posit that the process of mapping between the picture and
566   the name itself appears to differ between lexical categories, namely nouns versus verbs
567   (Szekely et al., 2005). Other evidence taken from studies involving patients with aphasia has
568   shown that the mental lexicon distinguishes grammatical classes (Benetello et al., 2016).

569   There are several brain regions associated with word production during lexical selection.
570   Indefrey and Levelt (Indefrey and Levelt, 2004) reviewed 82 functional imaging studies of
571   single word production, identifying 11 regions in the left hemisphere (posterior IFG, ventral
572   precentral gyrus, SMA, mid and posterior STG and MTG, posterior temporal fusiform gyrus,
573   anterior insula, thalamus, and medial cerebellum) and four in the right (mid-STG, medial and

574 lateral cerebellum and SMA) involved in core processes of word production. Other functional
575 imaging studies have demonstrated that lexical-semantic knowledge is stored in the temporal
576 lobe (Vigneau et al., 2006) and that the region can operate as a lexical interface linking
577 phonological and semantic information in a sound-to-meaning interface (Hickok and Poeppel,
578 2007). Elsewhere, the left MTG has been found to associate with lexical selection (Indefrey
579 and Levelt, 2004). The spatiotemporal dynamics of word retrieval, including lexical selection,
580 are not well understood, but Riès et al. (Riès et al., 2017) have shown that activation of word
581 representations and their selection temporally co-occur and that a widespread network of
582 overlapping brain regions is associated. The variety of brain regions implicated in word
583 production suggests that there is potential for exploiting semantics, syntax and phonology to
584 activate different regions during imagined speech production to maximise the separability of
585 brain activations for DS-BCI.

586 **4.4   The hierarchical structure of syntax**

587 Contemporary linguistic theories contend that syntactic and sentential representations are
588 complex sets of hierarchically organised syntactic categories, and that the relationships
589 between categories in this hierarchy determine the different aspects of propositional meaning
590 (see Zaccarella and Friederici (2016) for a neurobiological review of syntactic hierarchies).
591 During syntactic encoding, a conceptual message is linguistically encoded by retrieval of
592 corresponding words from the lexicon, and grammatical ordering of these words (Indefrey et
593 al., 2001). Stored syntactic information, such as word class, are used to compute a structure
594 that specifies the relationships between words in a sentence, e.g. order and inflection.

595 It has been proposed (Frazier, 1987), and countered (Friederici, 2002), that there is an isolated
596 syntactic processing mechanism that has no relation to semantics or other non-syntactic
597 information. It has been stated that syntactic encoding in speech production exhibits close
598 temporal overlap with other processes (Indefrey et al., 2001) and that brain activations in the
599 frontotemporal language network have indicated that syntactic processing occurs prior to
600 semantic processing, but that these processes are not isolated mechanisms (Friederici, 2002).

601 Syntactic processing is specifically associated with BA44, located in the posterior portion of
602 Broca's area in the LIFG, and its white matter connection to the posterior temporal cortex
603 (Friederici, 2018). A functional imaging study has provided evidence that hierarchical syntactic
604 conditions localised in the ventral portion of BA44 (Zaccarella and Friederici, 2015). In
605 contrast, activations corresponding to processing of two-word sentences without syntactic

606 hierarchy were associated with the frontal operculum/anterior insula. Love et al. (Love et al.,
607 2008) provide evidence that the LIFC supports syntactic processing because it sustains the
608 requisite lexical activation speed needed for the real-time formation of a syntactic dependency.
609 Elsewhere, PET has been used to identify both sentence-level and local syntactic encoding of
610 speech in the Rolandic operculum, adjacent to Broca's area (Indefrey et al., 2001).

611 ## 4.5 The internal phonological speech code

612 Within psycholinguistic theory the assumption exists that speech articulation is preceded by an
613 internal abstract speech code (Wheeldon and Levelt, 1995). In speech production, a word can
614 have different intonation, duration and amplitude, leading to the proposal that each linguistic
615 unit has a phonological representation encoding features unique to that unit. Phonological
616 representations are categorical and consist of discrete timeless segments (Wheeldon and Levelt,
617 1995). Models differ as to the timing and order at which phonemes are assigned to a
618 phonological structure. Following the syntactic computation phase, stored information on the
619 sounds of words is retrieved as "phonological codes". These are then transformed to produce
620 an executable code i.e., speech (Indefrey et al., 2001).

621 It has been proposed that phonological word representation is accessed from Broca's area and
622 compiled into segments of syllables (Indefrey and Levelt, 2004). Other studies indicate that the
623 posterior middle and inferior portions of the temporal lobes are linked to phonological and
624 semantic processing (see Hickok and Poeppel, 2007). Another suggestion (Edwards et al., 2010)
625 is that speech production is enabled through verbal/phonological working memory using the
626 dorsal stream areas implicated in speech perception and phonological working memory (e.g.
627 Hickok and Poeppel, 2007). It has been suggested that phonological encoding exhibits
628 correlation with the superior temporal sulcus (STS) (Llorens et al., 2011), while the authors of
629 one study linked the IFG and STS gamma band responses (>40 Hz) to the phonological
630 retrieval processes and imagined speech production, using intracranial EEG recordings (Mainy
631 et al., 2008). Although it is well known that lemma selection begins earlier than phonological
632 encoding it seems that there is some temporal overlap between the two activations (Sedivy,
633 2014) and it is possible that phonologically-similar words are represented by overlapping cell
634 assemblies sharing a single perisylvian region (Pulvermüller, 1999). It is possible for a
635 phonological word form to have two meanings (e.g. the noun/verb dichotomy of the/to beat),
636 and it has been suggested that there must be an underlying mechanism for realising the
637 exclusive-or relationship between the two.

638 The review of the literature presented in sections 2, 3 and 4 provides the basis for our discussion
639 on the role of linguistics within the framework of DS-BCI research. This discussion is
640 presented in section 5.

5 An Enhanced Role for Linguistics in BCI Research

642 Overt speech is a rich tapestry of sound, pitch, rhythm, structure and meaning, and studies have
643 shown that imagined speech retains many of these articulatory characteristics (Alderson-Day
644 and Fernyhough, 2015; Scott et al., 2013). It is one of the great challenges of DS-BCI research
645 to represent this communicative richness through the modality of a BCI. With this goal in mind,
646 improvements to experimental protocol have been suggested, including the use of a vocabulary
647 of words with semantic meaning to improve discrimination between words, and a normalisation
648 of word length to mitigate the high variance of this feature (Porbadnigk et al., 2009). We
649 advocate the use of novel experimental design to enhance effective elicitation of imagined
650 speech and improve discriminability between phonemes, words and sentences. Further
651 investigation into the neurological and neuroanatomical underpinnings of imagined speech
652 production and the development of a more concrete understanding of the information contained
653 within different frequency bands at different brain foci, is also required. The importance of
654 consistency in the way imagined speech is produced by experimental participants, and the
655 effect of providing them with a thorough understanding of what is meant by imagined speech
656 production, are additional areas for investigation that may improve the robustness of
657 experimentation. In the following subsections, we extend the work of Iljina et al. (Iljina et al.,
658 2017) by highlighting three key areas where BCI research can benefit from findings in the field
659 of neurolinguistics.

5.1 Incorporating the structure of speech production processing

661 The sheer complexity of the neural mechanisms underpinning speech is one of the primary
662 factors causing resistance to the development of a DS-BCI. In comparison with many of the
663 previous incarnations of communicative BCI (Chaudhary et al., 2017; Pandarinath et al., 2017),
664 the character of the modality of interaction, i.e. imagined speech, is still a relatively poorly
665 understood phenomenon. In relation to DS-BCIs, the following question has been put forward:
666 when does semantic, phonological, or syntactic processing occur (Iljina et al., 2017)? The
667 analysis of Indefrey (Indefrey, 2011) provides some insight into the relative timings associated
668 with the stages of speech production (see Figure 4) and indicates that it may be possible to
669 target decoding of semantic information at an earlier stage than the phonological representation.

670   The temporal sequence of these processes is an important consideration for BCI researchers
671   seeking to extract meaning from imagined speech, but there are opposing views to navigate.
672   One of these is a sequential model in which word production involves a series of separate stages
673   from semantic concept through word retrieval and phonological articulation (Levelt et al.,
674   1999). Alternative models hypothesize a parallel architecture in which neuro-linguistic
675   processes occur simultaneously (Jackendoff, 2007). Whichever of these models is correct, they
676   must be incorporated into the DS-BCI paradigm.

677   The speech production process as depicted in section 4 offers a staged process with the potential
678   to be mined for more targeted decoding approaches. Models of speech processing, for example,
679   have proposed that accessing the phonological representation of a word releases two kinds of
680   information: a frame which specifies the structure of a word and phonemes to fill slots in this
681   structure (Dell, 1988; Levelt, 1992). An interesting operation referred to as *gap filling* (Love
682   et al., 2008) has been observed in studies of lexical priming where the meaning of a displaced
683   constituent is activated when it is first encountered in a sentence and then reactivated at a site
684   indexed by a trace. Consider the following sentence as an example: "$(The\ boy)_i$ that the horse
685   chased $(t)_i$ is tall." In a case like this, activation is present for "boy" and again at the gap
686   indexed by "t" where there is no phonologically realised word. Crucially, there is no activation
687   before the word "chased", indicating that the activation for "boy" at the gap is not residual
688   activation but the result of reactivation (Love et al., 2008). This may have important
689   implications for the development of a DS-BCI which decodes continuous imagined speech
690   from brain activity, as the neurological basis of syntax requires a complex series of operations
691   not simply based on surface word order. Understanding of the widely distributed brain regions
692   associated with semantic and syntactic processing, and speech production (as discussed in
693   section 4.2 and 4.4) should be harnessed along with enhanced methods for eliciting imagined
694   speech, to improve the decoding accuracy of DS-BCIs.

695   Herff et al. (Herff et al., 2017) have shown that continuous speech is represented as a sequence
696   of phones within the brain and is thus a legitimate target for DS-BCI research. Following this,
697   it seems reasonable to suggest that concatenation of imagined speech units can be used to
698   produce words and sentences. Perrone-Bertolotti et al. (Perrone-Bertolotti et al., 2014) discuss
699   concerns over the way imagined speech manifests itself and how personal agency or lack
700   thereof leads to different forms of imagined speech. The more active form, described as
701   "deliberate covert production of speech", is consciously-generated speech and the target of DS-
702   BCI research. However, a less deliberate manifestation known as "verbal mind wandering" can

703 occur spontaneously. Despite not being the direct target of DS-BCIs, this second state of
704 imagined speech may influence the performance of such a device or even activate
705 communication when none was intended.

5.2 Leveraging neurolinguistics concepts to improve discriminability

707 The ability to effectively discriminate between neural recordings is an essential component of
708 any BCI, and it is a particularly complicated challenge in relation to DS-BCIs, given the
709 complex and dynamic processes of speech production. Decoding brain activity corresponding
710 to imagined speech, given the dense vocabulary and the volume of potential semantic
711 combinations that humans possess is an exceptional challenge. In Section 4.2, evidence is
712 presented linking different semantic categories to different lesion foci, and semantic
713 categorisation of words appears to be a promising method for improving classification from a
714 constrained lexicon. Content words i.e., words with rich semantic meaning (e.g. words
715 referring to tastes, sensations, sounds, motor activities etc.) have been associated with distinct
716 regions of the brain and may enable classification of words based on semantic criteria
717 (Pulvermüller, 1999). Although this may appear to be a somewhat contrived method for
718 improving accuracy, this approach can help elucidate the degree to which semantic
719 categorisation contributes to differentiation between words (Wang et al., 2011). Categorical
720 differences between words can induce significantly different brain activity and this variance
721 may be an aid to classification. For example, action words (e.g. kick, throw, blink) can have
722 the effect of activating brain regions actually involved in carrying out the activity (Hauk et al.,
723 2004). Similarly, words corresponding to touch may include significant activation in the
724 somatosensory cortices and sound words may cause increased activation in bilateral auditory
725 cortices (Pulvermüller, 1999).

726 Imagined speech's close association with working memory (Marvel and Desmond, 2012), the
727 range of articulatory forms it can take (Alderson-Day et al., 2015; Deng et al., 2010) and the
728 different neural activations it exhibits in relation to overt speech (Basho et al., 2007), contribute
729 to making imagined speech extremely difficult to decode effectively. Methods employed in
730 neurolinguistics can help DS-BCI researchers improve cuing and elicitation techniques,
731 making it easier to determine precisely what is being decoded from brain activity. This may
732 take the form of semantic or phonological priming, as suggested above, or improvement of
733 experimental protocols to ensure participants are clear on what is expected from them. It may

734 also be possible to protect against unwanted noise in the data, for example, via articulatory
735 suppression.

736 The previously-stated proposal that each linguistic unit has a unique phonological
737 representation (Section 4.5) is a potential avenue for improving imagined speech
738 discriminability (Zhao and Rudzicz, 2015). Clearly, if the assertion of a unique phonological
739 code is correct, this would be a primary target of DS-BCI decoding approaches, as a single
740 representation corresponding to a single word or phoneme would make those approaches easier
741 to implement, given that the prior stages in the speech production process may not be required.
742 It is the recommendation of this review that further investigation into the potential phonological
743 discriminability of units of imagined speech is pursued.

744 Although much of the research to date into a possible DS-BCI has focused on discrete linguistic
745 units, i.e. vowels, consonants etc., it has been suggested that the neural substrates responsible
746 for the representation of phonemes may differ depending on whether they are processed as part
747 of a sequence or processed alone (Ikeda et al., 2014). Di Liberto, O'Sullivan and Lalor (Di
748 Liberto et al., 2015) lament the lack of research present in the literature regarding the parsing
749 and processing of continuous speech. However, the difficulty of experimentation with
750 imagined speech and the impracticality of attempting to decode continuous speech, at a time
751 when decoding discrete units of speech is still enormously challenging, has meant that to-date
752 the majority of studies have focused on discrete units of speech in the development of decoding
753 strategies.

754 If progress is to be made using these approaches, the anatomical information summarised in
755 Sections 3 and 4 will be important for informing decoding strategies. Targeting regions-of-
756 interests specific to speech production may be a promising approach to the development of a
757 DS-BCI (Guenther et al., 2009), particularly considering that speech processing is a highly-
758 distributed operation with semantics, lexical access, syntax and phonology all correlated to
759 different regions. Although we agree with Bocquelet et al. (Bocquelet et al., 2017) that the
760 LIFG is clearly implicated in imagined speech production, and a promising candidate for DS-
761 BCI research, we think it is important to consider a wider, and probably bilateral, network
762 where the distributed connectivity predicted by Hebbian theory is accounted for. The evidence
763 presented here indicates a wide cortical network associated with different linguistic categories
764 and stages of the speech production process. It is our assertion that a complete picture of the
765 neuro-anatomical correlates of imagined speech will provide greater opportunities for effective
766 discriminability.

5.3 Mitigating the limitations of experimental methodology

Progress towards a DS-BCI is dependent on the effectiveness of future research methodologies and on novel approaches to system development. It has been noted that researchers seeking to distinguish word classes from neural activation should consider the effect of word length, word frequency, emotional properties of the stimuli, word repetition, priming and syntactic and semantic context when designing experiments (Pulvermüller, 1999). The same author also warns of the possible unintended effects of presenting words in sentences or word-strings, due to the neurophysiological response being a complex blend of the semantic and syntactic interactions of the given words. One of the difficulties associated with development of a DS-BCI is inferring from experimental participants that the required tasks have been performed (Geva et al., 2011b). The lack of behavioural output from participants has meant that researchers have been faced with a choice of whether to accept assertions that a given task has been correctly undertaken, to design their experimental procedure in a manner that will elicit the required imagined speech activity (Geva et al., 2011a), or to merge their imagined speech protocols with an overt action in an attempt at cross-verification (Oppenheim and Dell, 2008). Limitations to the scope of empirical study in the case of imagined speech has induced the development of methods for indirect study of the phenomenon (Filik and Barber, 2011; Oppenheim and Dell, 2008). Alderson-Day and Fernyhough (Alderson-Day and Fernyhough, 2015) present recent methodological advances in the field, including imagined speech inducement and inhibition as a means of studying its effects.

Neuroimaging studies into the nature of imagined speech have often asked participants to simply articulate some words or sentences in imagined speech, or to imagine speech with different characteristics. A danger associated with these studies is the lack of ecological validity in eliciting imagined speech (Alderson-Day and Fernyhough, 2015) and the failure of researchers to acknowledge the possibility that imagined speech is present during baseline assessments (Jones and Fernyhough, 2007). A technique known as articulatory suppression might provide some assistance in ameliorating this issue (Miyake et al., 2004). The evidence presented in Section 3.1 indicated variation in the phenomena of imagined speech, both in terms of how it is activated and how it is perceived. Studies have shown that imagined speech is not generally understood in the same way by participants and can vary widely in its phenomenology (Alderson-Day and Fernyhough, 2015). It is the job of the DS-BCI researcher to ensure that each participant is well-informed prior to engaging in experimentation. The methodology employed by Geva, Jones, et al. (Geva et al., 2011b) may be an interesting avenue

for exploration in DS-BCI research. Their use of rhyming words and/or homophones is commonly applied in linguistics (Badre et al., 2005; Filik and Barber, 2011) to allow researchers to know whether participants are using imagined speech or resorting to other linguistic/cognitive strategies. For example, 'might' and 'mite' are homophones, while 'ear' and 'oar' are not. These are tasks that could not be solved by orthography alone and thus require the use of imagined speech.

Research methodology using overt speech to represent imagined speech within experimental paradigms is flawed, at least to some degree. Overt speech-trained models for example, are an active research area but it must be understood that neural representations of overt and imagined speech are not identical (Chakrabarti et al., 2015). Hubbard (Hubbard, 2010) reflects that differences in experimental results between overt and imagined speech may simply be a function of a participant's ability to self-monitor and report accurately. There is general agreement that overt speech engages greater activation across a broader network of the brain than imagined speech, with areas including the mesial temporal lobe and sub-cortical structures (Kielar et al., 2011). Due to some notable differences observed from neural responses in overt and imagined conditions, inferences drawn from language processing studies should be considered with caution (Llorens et al., 2011). However, Iljina et al. (Iljina et al., 2017) believe that the body of research presented on both overt and imagined speech supports the premise of being able to decode expressive language from neuronal processes as well as translation of findings from overt to imagined speech.

Experimental results can be negatively affected by experimental conditions and an alternative approach to improving the robustness of results in relation to speech production and communicative interaction is the use of non-experimental, "real-world" speech (Derix et al., 2014, 2012; Ruescher et al., 2013). Spontaneous language can reflect mental states and thus constitutes a fundamental link between externally-observable behaviour and internal cognitive processes (Derix et al., 2014). Using their methodology, in which simultaneous ECoG and digital video recordings are used to identify periods of spontaneous communication between interlocutors, the group cited above has conducted studies based on concepts developed in psycholinguistic research into spontaneously-spoken language. The authors highlight the importance of study paradigms in which real-world situations can be investigated in a way not possible under strict experimental procedures. They present the use of stimuli such as naturalistic texts, recordings of interacting individuals and virtual reality simulations as associated methods being employed elsewhere (Derix et al., 2014).

833 In a series of studies, the research team used their methodology to study the neuronal processes
834 related to real-life communication in a non-experimental scenario (Derix et al., 2014, 2012;
835 Ruescher et al., 2013). This involved a technique for identifying time-periods in which patients
836 were involved in conversation with either partners or physicians (Derix et al., 2012). Extracted
837 epochs consisted of periods of natural, uninstructed conversation, with the results indicating
838 that the choice of linguistic and non-linguistic behaviours depends on whom a person is
839 speaking with. The authors suggest that such meta-information may have utility in BCI
840 applications aimed at restoration of expressive speech. Although non-experimental conditions
841 do facilitate the study of spontaneous speech, it is important to acknowledge, as the research
842 team has, that participants' behaviour may be moderated by the knowledge that they are under
843 surveillance, and therefore not completely natural (Derix et al., 2014). However, we agree with
844 Iljina et al. (Iljina et al., 2017) that a thorough understanding of brain activity during real-world
845 speech is required for the development of truly naturalistic DS-BCI.

846 As indicated throughout this review, there are several ways in which DS-BCI research can
847 benefit from neurolinguistics research advances. Understanding the phenomena of imagined
848 speech and individual speech processes is crucial, but looking towards neurolinguistics to
849 enhance experimental methodology and interpretation of results is also advocated here. Other
850 avenues exist for exploration of improvements to the performance of DS-BCIs, including signal
851 acquisition and advanced classification algorithms, but it would be wrong to ignore the
852 potential utility of cross-disciplinary research in neurolinguistics and DS-BCI.

6 Concluding Remarks

854 Development of a DS-BCI is an extremely challenging undertaking. It is the assertion of this
855 review that a cross-disciplinary approach must be taken to advance the field towards a
856 naturalistic form of communication. Here, we advocate the integration of neurolinguistics
857 within the DS-BCI paradigm for the improvement of experimental methodology and to aid
858 approaches to the decoding of neural signals. Insights into the nature of imagined speech, and
859 speech production processes, can inform research practices, while methodological approaches
860 common in linguistics can help improve procedural robustness in studies involving imagined
861 speech.

862 Clearly, there is no definitive description of the phenomena of imagined speech. Independently
863 depicted as a truncated form of overt speech, as showing greater activation in several brain
864 regions than overt speech and as having attenuated features in comparison with overt speech,

865 imagined speech is still relatively poorly understood. Continuing research into imagined speech
866 from a neurolinguistics perspective will be vital for DS-BCI. Imagined speech manifests itself
867 in different forms, whether that be through active or passive generation of imagined speech,
868 through accent, rhythm or pitch, or through conversational or single-speaker scenarios. That
869 being the case, future research in this field must make it abundantly clear to experimental
870 participants precisely what is being asked of them. The field of neurolinguistics can help inform
871 DS-BCI research on methods for targeting the imagined speech content required. Not unrelated
872 to this is the potential for additional information to be encoded in the neural recordings
873 extracted during periods of imagined speech production. Working memory and imagined
874 speech appear to be intrinsically linked and imagined speech trials are susceptible to influence
875 from the auditory or visual cues presented. It is therefore important that experimental
876 methodologies and decoding approaches mitigate against this unwanted content where possible.

877 This review has shown that not only is DS-BCI concerned with the phenomena of imagined
878 speech and how it differs from overt speech, but also with the neuroanatomy and specific
879 processes involved in the production of speech. Speech production is a temporal process with
880 a hierarchical structure and it is clear that it cannot be considered a single function localised in
881 a single brain region. Evidence has been presented from neurolinguistics research to indicate
882 that different systems of speech production, such as semantics and syntax, operate at distinct
883 time periods (sometimes overlapping) across a distributed network of brain regions, and that
884 these systems activate patterns of brain activity which may be useful for approaches to
885 decoding imagined speech.

886 A fully-functioning DS-BCI may, at present, seem a long way off and it may appear that there
887 are more pressing concerns, such as improving signal acquisition, for the field to be focused
888 on at present. However, it is our contention that it would be remiss to ignore the field of
889 neurolinguistics in DS-BCI research, given the potential benefits it can offer in the short-term
890 and the high-probability that it will be required in the longer-term development of a naturalistic
891 mode of communication.

894 **AUTHOR CONTRIBUTIONS**

895 Conceptualization, C.C., R.F., and D.C.; writing: original draft, C.C; writing: review and
896 editing R.F. and D.C.

897 **References**

898 Alderson-Day, B., Fernyhough, C., 2015. Inner speech: Development, cognitive functions,
899   phenomenology, and neurobiology. Psychol. Bull. 141, 931–965.
900   https://doi.org/10.1037/bul0000021

901 Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., Fernyhough, C.,
902   2015. The brain's conversation with itself: Neural substrates of dialogic inner speech. Soc.
903   Cogn. Affect. Neurosci. 11, 110–120. https://doi.org/10.1093/scan/nsv094

904 Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., De Haan, E.H.F., Kahn, R.S., 2005.
905   The functional neuroanatomy of metrical stress evaluation of perceived and imagined
906   spoken words. Cereb. Cortex 15, 221–228. https://doi.org/10.1093/cercor/bhh124

907 AlSaleh, M.M., Arvaneh, M., Christensen, H., Moore, R.K., 2016. Brain-computer interface
908   technology for speech recognition: a review. Asia-Pacific Signal Inf. Process. Assoc.
909   Annu.   Summit   Conf.   (APSIPA   ASC   2016).
910   https://doi.org/10.1109/APSIPA.2016.7820826

911 Baddeley, A.D., Thomson, N., Buchanan, M., 1975. Word length and the structure of short-
912   term   memory.   J.   Verbal   Learning   Verbal   Behav.   14,   575–589.
913   https://doi.org/10.1016/S0022-5371(75)80045-4

914 Badre, D., Poldrack, R.A., Par??-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable
915   controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal
916   cortex. Neuron 47, 907–918. https://doi.org/10.1016/j.neuron.2005.07.023

917 Basho, S., Palmer, E.D., Rubio, M.A., Wulfeck, B., Müller, R.A., 2007. Effects of generation
918   mode in fMRI adaptations of semantic fluency: Paced production and overt speech.
919   Neuropsychologia   45,   1697–1706.
920   https://doi.org/10.1016/j.neuropsychologia.2007.01.007

921 Bates, E., Wilson, S.M., Saygin, A.P., Dick, F., Sereno, M.I., Knight, R.T., Dronkers, N.F.,
922   2003. Voxel-based lesion–symptom mapping. Nat. Neurosci. 6, 448–450.
923   https://doi.org/10.1038/nn1050

924 Benetello, A., Finocchiaro, C., Capasso, R., Capitani, E., Laiacona, M., Magon, S., Miceli, G.,
925   2016. The dissociability of lexical retrieval and morphosyntactic processes for nouns and
926   verbs: A functional and anatomoclinical study. Brain Lang. 159, 11–22.
927   https://doi.org/10.1016/j.bandl.2016.05.005

928   Berwick, R.C., Friederici, A.D., Chomsky, N., Bolhuis, J.J., 2013. Evolution, brain, and the
929   nature of language. Trends Cogn. Sci. 17, 98. https://doi.org/10.1016/j.tics.2012.12.002

930   Bin, G., Gao, X., Yan, Z., Hong, B., Gao, S., 2009. An online multi-channel SSVEP-based
931   brain-computer interface using a canonical correlation analysis method. J. Neural Eng. 6.
932   https://doi.org/10.1088/1741-2560/6/4/046002

933   Blakely, T., Miller, K.J., Rao, R.P.N., Holmes, M.D., Ojemann, J.G., 2008. Localization and
934   classification of phonemes using high spatial resolution electrocorticography (ECoG)
935   grids. 2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 4964–4967.
936   https://doi.org/10.1109/IEMBS.2008.4650328

937   Bocquelet, F., Hueber, T., Girin, L., Chabardès, S., Yvert, B., 2017. Key considerations in
938   designing a speech brain-computer interface. J. Physiol. 110, 392–401.
939   https://doi.org/10.1016/j.jphysparis.2017.07.002

940   Bouchard, K.E., Chang, E.F., 2014. Neural decoding of spoken vowels from human sensory-
941   motor cortex with high-density electrocorticography. Conf. Proc. ... Annu. Int. Conf.
942   IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2014, 6782–6785.
943   https://doi.org/10.1109/EMBC.2014.6945185

944   Bouton, C.E., Shaikhouni, A., Annetta, N. V., Bockbrader, M.A., Friedenberg, D.A., Nielson,
945   D.M., Sharma, G., Sederberg, P.B., Glenn, B.C., Mysiw, W.J., Morgan, A.G.,
946   Deogaonkar, M., Rezai, A.R., 2016. Restoring cortical control of functional movement in
947   a human with quadriplegia. Nature 533, 247–250. https://doi.org/10.1038/nature17435

948   Brigham, K., Kumar, B.V.K.V., 2010. Imagined speech classification with EEG signals for
949   silent communication: A preliminary investigation into synthetic telepathy. 2010 4th Int.
950   Conf. Bioinforma. Biomed. Eng. iCBBE 2010 1–4.
951   https://doi.org/10.1109/ICBBE.2010.5515807

952   Brocklehurst, P.H., Corley, M., 2011. Investigating the inner speech of people who stutter:
953   Evidence for (and against) the Covert Repair Hypothesis. J. Commun. Disord. 44, 246–
954   260. https://doi.org/10.1016/j.jcomdis.2010.11.004

955   Brumberg, J.S., Krusienski, D.J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A.L.,
956   Schalk, G., 2016. Spatio-temporal progression of cortical activity related to continuous
957   overt and covert speech production in a reading task. PLoS One 11, 1–21.
958   https://doi.org/10.1371/journal.pone.0166872

959 Brumberg, J.S., Wright, E.J., Andreasen, D.S., Guenther, F.H., Kennedy, P.R., 2011.
960   Classification of intended phoneme production from chronic intracortical microelectrode
961   recordings in speech-motor cortex. Front. Neurosci. 5, 1–12.
962   https://doi.org/10.3389/fnins.2011.00065

963 Chakrabarti, S., Sandberg, H.M., Brumberg, J.S., Krusienski, D.J., 2015. Progress in speech
964   decoding from the electrocorticogram. Biomed. Eng. Lett. 5, 10–21.
965   https://doi.org/10.1007/s13534-015-0175-1

966 Chaudhary, U., Xia, B., Silvoni, S., Cohen, L.G., Birbaumer, N., 2017. Brain–Computer
967   Interface–Based Communication in the Completely Locked-In State. PLoS Biol. 15, 1–
968   25. https://doi.org/10.1371/journal.pbio.1002593

969 Chi, X., Hagedorn, J.B., Schoonover, D., Zmura, M.D., 2011. EEG-Based Discrimination of
970   Imagined Speech Phonemes. Int. J. Bioelectromagn. 13, 201–206.

971 Corley, M., Brocklehurst, P.H., Moat, H.S., 2011. Error biases in inner and overt speech:
972   Evidence from tongue twisters. J. Exp. Psychol. Learn. Mem. Cogn. 37, 162–175.
973   https://doi.org/10.1037/a0021321

974 D'albis, T., Blatt, R., Tedesco, R., Sbattella, L., Matteucci, M., 2012. A predictive speller
975   controlled by a brain-computer interface based on motor imagery. ACM Trans. Comput.
976   Interact. 19, 1–25. https://doi.org/10.1145/2362364.2362368

977 D'Zmura, M., Deng, S., Lappas, T., Thorpe, S., Srinivasan, R., 2009. Toward {EEG} Sensing
978   of Imagined Speech. Hum. Comput. Interact. 40–48.

979 DaSalla, C.S., Kambara, H., Sato, M., Koike, Y., 2009. Single-trial classification of vowel
980   speech imagery using common spatial patterns. Neural Networks 22, 1334–1339.
981   https://doi.org/10.1016/j.neunet.2009.05.008

982 Dell, G.S., 1988. The retrieval of phonological forms in production: Tests of predictions from
983   a connectionist model. J. Mem. Lang. 27, 124–142. https://doi.org/10.1016/0749-
984   596X(88)90070-8

985 Dell, G.S., Martin, N., Schwartz, M.F., 2007. A case-series test of the interactive two-step
986   model of lexical access: Predicting word repetition from picture naming. J. Mem. Lang.
987   56, 490–520. https://doi.org/10.1016/j.jml.2006.05.007

988 Denby, B., Oussar, Y., Dreyfus, G., Stone, M., 2006. Prospects for a Silent Speech Interface
989   using Ultrasound Imaging. Proc. 2006 IEEE Int. Conf. Acoust. Speed Signal Process. Proc.

990   1, I-365-I-368. https://doi.org/10.1109/ICASSP.2006.1660033

991   Deng, S., Srinivasan, R., Lappas, T., D'Zmura, M., 2010. EEG classification of imagined
992   syllable rhythm using Hilbert spectrum methods. J. Neural Eng. 7, 046006.
993   https://doi.org/10.1088/1741-2560/7/4/046006

994   Derix, J., Iljina, O., Schulze-Bonhage, A., Aertsen, A., Ball, T., 2012. "Doctor" or "darling"?
995   Decoding the communication partner from ECoG of the anterior temporal lobe during
996   non-experimental, real-life social interaction. Front. Hum. Neurosci. 6, 1–14.
997   https://doi.org/10.3389/fnhum.2012.00251

998   Derix, J., Iljina, O., Weiske, J., Schulze-Bonhage, A., Aertsen, A., Ball, T., 2014. From speech
999   to thought: the neuronal basis of cognitive units in non-experimental, real-life
1000   communication investigated using ECoG. Front. Hum. Neurosci. 8, 1–17.
1001   https://doi.org/10.3389/fnhum.2014.00383

1002   Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to
1003   speech reflects phoneme-level processing. Curr. Biol. 25, 2457–2465.
1004   https://doi.org/10.1016/j.cub.2015.08.030

1005   Dolcos, S., Albarracin, D., 2014. The inner speech of behavioral regulation: Intentions and task
1006   performance strengthen when you talk to yourself as a You. Eur. J. Soc. Psychol. 44, 636–
1007   642. https://doi.org/10.1002/ejsp.2048

1008   Donchin, E., Spencer, K.M., Wijesinghe, R., 2000. The mental prosthesis: Assessing the speed
1009   of a P300-based brain- computer interface. IEEE Trans. Rehabil. Eng. 8, 174–179.
1010   https://doi.org/10.1109/86.847808

1011   Edwards, E., Nagarajan, S.S., Dalal, S.S., Canolty, R.T., Kirsch, H.E., Barbaro, N.M., Knight,
1012   R.T., 2010. Spatiotemporal imaging of cortical activation during verb generation and
1013   picture naming. Neuroimage 50, 291–301.
1014   https://doi.org/10.1016/j.neuroimage.2009.12.035

1015   Emerson, M.J., Miyake, A., 2003. The role of inner speech in task switching: A dual-task
1016   investigation. J. Mem. Lang. 48, 148–168. https://doi.org/10.1016/S0749-
1017   596X(02)00511-9

1018   Fargier, R., Bürki, A., Pinet, S., Alario, F.X., Laganaro, M., 2018. Word onset phonetic
1019   properties and motor artifacts in speech production EEG recordings. Psychophysiology
1020   55, 1–10. https://doi.org/10.1111/psyp.12982

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

Filik, R., Barber, E., 2011. Inner speech during silent reading reflects the reader's regional accent. PLoS One 6. https://doi.org/10.1371/journal.pone.0025782

Frazier, L., 1987. Sentence processing: A tutorial review., in: Attention and Performance 12: The Psychology of Reading. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, pp. 559–586.

Friederici, A.D., 2018. The neural basis for human syntax: Broca's area and beyond. Curr. Opin. Behav. Sci. 21, 88–92. https://doi.org/10.1016/j.cobeha.2018.03.004

Friederici, A.D., 2011. The Brain Basis of Language Processing: From Structure to Function. Physiol. Rev. 91, 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

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. https://doi.org/10.1080/02687038.2010.511236

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, 3071–3082. https://doi.org/10.1093/brain/awr232

Gilbert, J.M., Rybchenko, S.I., Hofe, R., Ell, S.R., Fagan, M.J., Moore, R.K., Green, P., 2010. Isolated word recognition of silent speech using magnetic implants and sensors. Med. Eng. Phys. 32, 1189–1197. https://doi.org/10.1016/j.medengphy.2010.08.011

González-Castañeda, E.F., Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L., 2017. Sonification and textification: Proposing methods for classifying unspoken words from EEG signals. Biomed. Signal Process. Control 37, 82–91. https://doi.org/10.1016/j.bspc.2016.10.012

Goto, T., Hirata, M., Umekawa, Y., Yanagisawa, T., Shayne, M., Saitoh, Y., Kishima, H., Yorifuji, S., Yoshimine, T., 2011. Frequency-dependent spatiotemporal distribution of cerebral oscillatory changes during silent reading: A magnetoencephalograhic group analysis. Neuroimage 54, 560–567. https://doi.org/10.1016/j.neuroimage.2010.08.023

1052 Guenther, F.H., Brumberg, J.S., Joseph Wright, E., Nieto-Castanon, A., Tourville, J.A., Panko,
1053   M., Law, R., Siebert, S.A., Bartels, J.L., Andreasen, D.S., Ehirim, P., Mao, H., Kennedy,
1054   P.R., 2009. A wireless brain-machine interface for real-time speech synthesis. PLoS One
1055   4. https://doi.org/10.1371/journal.pone.0008218

1056 Hashim, N., Ali, A., Mohd-Isa, W.-N., 2018. Word-Based Classification of Imagined Speech
1057   Using EEG, in: Alfred, R., Iida, H., Ag. Ibrahim, A.A., Lim, Y. (Eds.), Computational
1058   Science and Technology. Springer Singapore, Singapore, pp. 195–204.

1059 Hauk, O., Johnsrude, I., Pulvermüller, F., 2004. Somatotopic Representation of Action Words
1060   in Human Motor and Premotor Cortex. Neuron 41, 301–307.
1061   https://doi.org/10.1016/S0896-6273(03)00838-9

1062 Henseler, I., Regenbrecht, F., Obrig, H., 2014. Lesion correlates of patholinguistic profiles in
1063   chronic aphasia: Comparisons of syndrome-, modality-and symptom-level assessment.
1064   Brain 137, 918–930. https://doi.org/10.1093/brain/awt374

1065 Herff, C., de Pesters, A., Heger, D., Brunner, P., Schalk, G., Schultz, T., 2017. Towards
1066   continuous speech recognition for {BCI}. Brain-computer interface Res. a state-of-the-art
1067   Summ. 5 1–9.

1068 Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., Schultz, T., 2015.
1069   Brain-to-text: Decoding spoken phrases from phone representations in the brain. Front.
1070   Neurosci. 9, 1–11. https://doi.org/10.3389/fnins.2015.00217

1071 Herff, C., Johnson, G., Diener, L., Shih, J., Krusienski, D., Schultz, T., 2016. Towards direct
1072   speech synthesis from ECoG: A pilot study. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.
1073   Soc. EMBS 2016–Octob, 1540–1543. https://doi.org/10.1109/EMBC.2016.7591004

1074 Herff, C., Putze, F., Heger, D., Guan, C., Schultz, T., 2012. Speaking mode recognition from
1075   functional Near Infrared Spectroscopy, in: Proceedings of the Annual International
1076   Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.
1077   https://doi.org/10.1109/EMBC.2012.6346279

1078 Hickok, G., 2014. The architecture of speech production and the role of the phoneme in speech
1079   processing. Lang. Cogn. Neurosci. 29, 2–20.
1080   https://doi.org/10.1080/01690965.2013.834370

1081 Hickok, G., 2012. Computatinal neuroanatomy of speech production. Nat. Rev. Neurosci. 13,
1082   135–145. https://doi.org/10.1038/nrn3158

1083 Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev.
1084   Neurosci. 8, 393–402. https://doi.org/10.1038/nrn2113

1085 Hirshorn, E.A., Thompson-Schill, S.L., 2006. Role of the left inferior frontal gyrus in covert
1086   word retrieval: Neural correlates of switching during verbal fluency. Neuropsychologia
1087   44, 2547–2557. https://doi.org/10.1016/j.neuropsychologia.2006.03.035

1088 Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin,
1089   S., Liu, J., Cash, S.S., Van Der Smagt, P., Donoghue, J.P., 2012. Reach and grasp by
1090   people with tetraplegia using a neurally controlled robotic arm. Nature 485, 372–375.
1091   https://doi.org/10.1038/nature11076

1092 Huang, J., Carr, T.H., Cao, Y., 2002. Comparing Cortical Activations for Silent and Overt
1093   Speech Using Event-Related fMRI. Hum. Brain Mapp. 15, 26–38.
1094   https://doi.org/10.1002/hbm.XXXX

1095 Hubbard, T.L., 2010. Auditory imagery: Empirical findings. Psychol. Bull. 136, 302–329.
1096   https://doi.org/10.1037/a0018436

1097 Hurlburt, R.T., Alderson-Day, B., Kuhn, S., Fernyhough, C., 2016. Exploring the Ecological
1098   Validity of Thinking on Demand: Neural Correlates of Elicited vs. Spontaneously
1099   Occurring Inner Speech. PLoS One 11, 1–16.
1100   https://doi.org/10.1371/journal.pone.0147932

1101 Hurlburt, R.T., Heavey, C.L., Kelsey, J.M., 2013. Toward a phenomenology of inner speaking.
1102   Conscious. Cogn. 22, 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

1103 Ibayashi, K., Kunii, N., Matsuo, T., Ishishita, Y., Shimada, S., Kawai, K., Saito, N., 2018.
1104   Decoding speech with integrated hybrid signals recorded from the human ventral motor
1105   cortex. Front. Neurosci. 12. https://doi.org/10.3389/fnins.2018.00221

1106 Idrees, B.M., Farooq, O., 2016. EEG based vowel classification during speech imagery. 2016
1107   3rd Int. Conf. Comput. Sustain. Glob. Dev. 1130–1134.

1108 Ikeda, S., Shibata, T., Nakano, N., Okada, R., Tsuyuguchi, N., Ikeda, K., Kato, A., 2014. Neural
1109   decoding of single vowels during covert articulation using electrocorticography. Front.
1110   Hum. Neurosci. 8, 1–8. https://doi.org/10.3389/fnhum.2014.00125

1111 Iljina, O., Derix, J., Schirrmeister, R.T., Schulze-Bonhage, A., Auer, P., Aertsen, A., Ball, T.,
1112   2017. Neurolinguistic and machine-learning perspectives on direct speech BCIs for
1113   restoration of naturalistic communication. Brain-Computer Interfaces 4, 186–199.

1114   https://doi.org/10.1080/2326263X.2017.1330611

1115   Indefrey, P., 2011. The spatial and temporal signatures of word production: a critical update.
1116   Front. Psychol. 2. https://doi.org/10.3389/fpsyg.2011.00255

1117   Indefrey, P., Brown, C.M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R.J., Hagoort, P., 2001.
1118   A neural correlate of syntactic encoding during speech production. Proc. Natl. Acad. Sci.
1119   98, 5933–5936. https://doi.org/10.1073/pnas.101118098

1120   Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production
1121   components. Cognition 92, 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

1122   Iqbal, S., Khan, Y.U., Farooq, O., 2015a. EEG based classification of imagined vowel sounds.
1123   2015 2nd Int. Conf. Comput. Sustain. Glob. Dev. 1591–1594.

1124   Iqbal, S., Shanir, P.P.M., Khan, Y.U.K., Farooq, O., 2015b. Time Domain Analysis of EEG to
1125   Classify Imagined Speech. Proc. Second Int. Conf. Comput. Commun. Technol. Adv.
1126   Intell. Syst. Comput. 380, 793–800. https://doi.org/10.1007/978-81-322-2523-2

1127   Jackendoff, R., 2007. A Parallel Architecture perspective on language processing. Brain Res.
1128   1146, 2–22. https://doi.org/10.1016/j.brainres.2006.08.111

1129   Jones, S.R., Fernyhough, C., 2007. Neural correlates of inner speech and auditory verbal
1130   hallucinations: A critical review and theoretical integration. Clin. Psychol. Rev. 27, 140–
1131   154. https://doi.org/10.1016/J.CPR.2006.10.001

1132   Kanas, V.G., Mporas, I., Benz, H.L., Sgarbas, K.N., Bezerianos, A., Crone, N.E., 2014. Joint
1133   spatial-spectral feature space clustering for speech activity detection from ecog signals.
1134   IEEE Trans. Biomed. Eng. 61, 1241–1250. https://doi.org/10.1109/TBME.2014.2298897

1135   Kellis, S., Miller, K., Thomson, K., Brown, R., House, P., Greger, B., 2010. Decoding spoken
1136   words using local field potentials recorded from the cortical surface. J. Neural Eng. 7,
1137   056007. https://doi.org/10.1088/1741-2560/7/5/056007

1138   Kielar, A., Milman, L., Bonakdarpour, B., Thompson, C.K., 2011. Neural correlates of covert
1139   and overt production of tense and agreement morphology: Evidence from fMRI. J.
1140   Neurolinguistics 24, 183–201. https://doi.org/10.1016/j.jneuroling.2010.02.008

1141   Kim, T., Lee, J., Choi, H., Lee, H., Kim, I.Y., Jang, D.P., 2013. Meaning based covert speech
1142   classification for brain-computer interface based on electroencephalography. Int.
1143   IEEE/EMBS Conf. Neural Eng. NER 53–56. https://doi.org/10.1109/NER.2013.6695869

1144 Lambon Ralph, M.A., Moriarty, L., Sage, K., Ackerman, T., Arul, K., Bird, H., Crisp, J.,
1145   Davies, S., Ellis, A., Harold, R., Heaney, P., Jones, M., Knapp, S., Luff, H., McKinnon,
1146   R., Morris, J., Newton, A., Raynor, R., Sharpe, S., Southwell, L., Swinson, J., Taylor, L.,
1147   Walker, E., Whitton, E., 2002. Anomia is simply a reflection of semantic and phonological
1148   impairments: Evidence from a case-series study. Aphasiology 16, 56–82.
1149   https://doi.org/10.1080/02687040143000448

1150 Leuthardt, E.C., Gaona, C., Sharma, M., Szrama, N., Roland, J., Freudenberg, Z., Solis, J.,
1151   Breshears, J., Schalk, G., 2011. Using the electrocorticographic speech network to control
1152   a brain–computer interface in humans. J. Neural Eng. 8, 036004.
1153   https://doi.org/10.1088/1741-2560/8/3/036004

1154 Levelt, W.J.M., 1999. Models of word production. Trends Cogn. Sci. 3, 223–232.
1155   https://doi.org/10.1016/S1364-6613(99)01319-4

1156 Levelt, W.J.M., 1992. Accessing words in speech production: Stages, processes and
1157   representations. Cognition 42, 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

1158 Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production.
1159   Behav. Brain Sci. 22, 1–75. https://doi.org/10.1017/S0140525X99001776

1160 Livezey, J.A., Bouchard, K.E., Chang, E.F., 2018. Deep learning as a tool for neural data
1161   analysis: speech classification and cross-frequency coupling in human sensorimotor
1162   cortex 1–23.

1163 Llorens, A., Trébuchon, A., Liégeois-Chauvel, C., Alario, F.X., 2011. Intra-cranial recordings
1164   of brain activity during language production. Front. Psychol. 2, 1–12.
1165   https://doi.org/10.3389/fpsyg.2011.00375

1166 Lotte, F., Brumberg, J.S., Brunner, P., Gunduz, A., Ritaccio, A.L., Guan, C., Schalk, G., 2015.
1167   Electrocorticographic representations of segmental features in continuous speech. Front.
1168   Hum. Neurosci. 09, 1–13. https://doi.org/10.3389/fnhum.2015.00097

1169 Love, T., Swinney, D., Walenski, M., Zurif, E., 2008. How left inferior frontal cortex
1170   participates in syntactic processing: Evidence from aphasia. Brain Lang. 107, 203–219.
1171   https://doi.org/10.1016/j.bandl.2007.11.004

1172 MacKay, D.G., others, 1992. Constraints on theories of inner speech. Audit. Imag. 121–149.

1173 Mainy, N., Jung, J., Baciu, M., Kahane, P., Schoendorff, B., Minotti, L., Hoffmann, D.,
1174   Bertrand, O., Lachaux, J.P., 2008. Cortical dynamics of word recognition. Hum. Brain

1175   Mapp. 29, 1215–1230. https://doi.org/10.1002/hbm.20457

1176 Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N.E., Rieger, J., Schalk, G., Knight,
1177   R.T., Pasley, B.N., 2014. Decoding spectrotemporal features of overt and covert speech
1178   from the human cortex. Front. Neuroeng. 7. https://doi.org/10.3389/fneng.2014.00014

1179 Martin, S., Brunner, P., Iturrate, I., Millán, J. del R., Schalk, G., Knight, R.T., Pasley, B.N.,
1180   2016. Word pair classification during imagined speech using direct brain recordings. Sci.
1181   Rep. 6. https://doi.org/10.1038/srep25803

1182 Martínez-Manrique, F., Vicente, A., 2015. The activity view of inner speech. Front. Psychol.
1183   6, 1–13. https://doi.org/10.3389/fpsyg.2015.00232

1184 Marvel, C.L., Desmond, J.E., 2012. From storage to manipulation: How the neural correlates
1185   of verbal working memory reflect varying demands on inner speech. Brain Lang. 120, 42–
1186   51. https://doi.org/10.1016/j.bandl.2011.08.005

1187 McGuire, P.K., Silbersweig, D.A., Murray, R.M., David, A.S., Frackowiak, R.S.J., Frith, C.D.,
1188   1996a. Functional anatomy of inner speech and auditory verbal imagery. Psychol. Med.
1189   26, 29. https://doi.org/10.1017/S0033291700033699

1190 McGuire, P.K., Silbersweig, D.A., Wright, I., Murray, R.M., Frackowiak, R.S.J., C.D., F.,
1191   McGuire, P.K., Silbersweig, D.A., Wright, I., Murray, R.M., Frackowiak, R.S., Frith,
1192   C.D., 1996b. The neural correlates of inner speech and auditory verbal imagery in
1193   schizophrenia: Relationship to auditory verbal hallucinations. Br. J. Psychiatry 169, 148–
1194   159.

1195 Miyake, A., Emerson, M.J., Padilla, F., Ahn, J.C., 2004. Inner speech as a retrieval aid for task
1196   goals: The effects of cue type and articulatory suppression in the random task cuing
1197   paradigm. Acta Psychol. (Amst). 115, 123–142.
1198   https://doi.org/10.1016/j.actpsy.2003.12.004

1199 Mohr, B., Pulvermüller, F., Mittelstädt, K., Rayman, J., 1996. Multiple simultaneous stimulus
1200   presentation facilitates lexical processing. Neuropsychologia 34, 1003–1013.

1201 Morin, A., Michaud, J., 2007. Self-awareness and the left inferior frontal gyrus: Inner speech
1202   use during self-related processing. Brain Res. Bull. 74, 387–396.
1203   https://doi.org/10.1016/J.BRAINRESBULL.2007.06.013

1204 Moses, D.A., Leonard, M.K., Chang, E.F., 2018. Real-time classification of auditory sentences
1205   using evoked cortical activity in humans. J. Neural Eng. 15, 036005.

1206   https://doi.org/10.1088/1741-2552/aaab6f

1207 Mugler, E.M., Goldrick, M., Slutzky, M.W., 2014a. Cortical encoding of phonemic context
1208   during word production. Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE
1209   Eng. Med. Biol. Soc. Annu. Conf. 2014, 6790–6793.
1210   https://doi.org/10.1109/EMBC.2014.6945187

1211 Mugler, E.M., Patton, J.L., Flint, R.D., Wright, Z.A., Schuele, S.U., Rosenow, J., Shih, J.J.,
1212   Krusienski, D.J., Slutzky, M.W., 2014b. Direct classification of all American English
1213   phonemes using signals from functional speech motor cortex. J. Neural Eng. 11, 035015.
1214   https://doi.org/10.1088/1741-2560/11/3/035015

1215 Nguyen, C.H., Karavas, G., Artemiadis, P., 2017. Inferring imagined speech using EEG signals:
1216   a new approach using Riemannian Manifold features. J. Neural Eng.
1217   https://doi.org/10.1088/1741-2552/aa8235

1218 Oken, B.S., Orhan, U., Roark, B., Erdogmus, D., Fowler, A., Mooney, A., Peters, B., Miller,
1219   M., Fried-Oken, M.B., 2014. Brain–Computer Interface With Language Model–
1220   Electroencephalography Fusion for Locked-In Syndrome. Neurorehabil. Neural Repair.
1221   https://doi.org/10.1177/1545968313516867

1222 Oppenheim, G.M., Dell, G.S., 2010. Motor movement matters: The flexible abstractness of
1223   inner speech. Mem. Cognit. 38, 1147–1160. https://doi.org/10.3758/MC.38.8.1147

1224 Oppenheim, G.M., Dell, G.S., 2008. Inner speech slips exhibit lexical bias, but not the
1225   phonemic similarity effect. Cognition 106, 528–537.
1226   https://doi.org/10.1016/j.cognition.2007.02.006

1227 Palmer, E.D., Rosen, H.J., Ojemann, J.G., Buckner, R.L., Kelley, W.M., Petersen, S.E., 2001.
1228   An event-related fMRI study of overt and covert word stem completion. Neuroimage 14,
1229   182–193. https://doi.org/10.1006/nimg.2001.0779

1230 Pandarinath, C., Nuyujukian, P., Blabe, C.H., Sorice, B.L., Saab, J., Willett, F.R., Hochberg,
1231   L.R., Shenoy, K. V., Henderson, J.M., 2017. High performance communication by people
1232   with paralysis using an intracortical brain-computer interface. Elife 6, 1–27.
1233   https://doi.org/10.7554/eLife.18554

1234 Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal
1235   component of working memory. Nature 361, 342–345.

1236 Paulesu, E., Goldacre, B., Scifo, P., Cappa, S.F., Gilardi, M.C., Castiglioni, I., Perani, D., Fazio,

F., 1997. Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. Neuroreport 8, 2011–2017. https://doi.org/10.1097/00001756-199705260-00042

Pei, X., Barbour, D.L., Leuthardt, E.C., Schalk, G., 2011a. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J. Neural Eng. 8, 046028. https://doi.org/10.1088/1741-2560/8/4/046028

Pei, X., Leuthardt, E.C., Gaona, C.M., Brunner, P., Wolpaw, J.R., Schalk, G., 2011b. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage 54, 2960–2972. https://doi.org/10.1016/j.neuroimage.2010.10.029

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

Piñango, M.M., Winnick, A., Ullah, R., Zurif, E., 2006. Time-course of semantic composition: The case of aspectual coercion. J. Psycholinguist. Res. 35, 233–244. https://doi.org/10.1007/s10936-006-9013-z

Porbadnigk, A., Wester, M., Calliess, J., Schultz, T., 2009. EEG-based speech recognition impact of temporal effects. Int. Conf. Bio-inspired Syst. Signal Process. https://doi.org/10.1.1.157.8486

Price, C.J., 2012. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62, 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Pulvermüller, F., 2005. Brain mechanisms linking language and action. Nat. Rev. Neurosci. 6, 576–582. https://doi.org/10.1038/2201358a0

Pulvermüller, F., 1999. Words in the brain ' s language. Behav. Brain Sci. 22, 253–336. https://doi.org/10.1016/j.bandl.2004.12.007

Ramsey, N.F., Salari, E., Aarnoutse, E.J., Vansteensel, M.J., Bleichner, M.B., Freudenburg, Z.V., 2017. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. Neuroimage 1–11. https://doi.org/10.1016/j.neuroimage.2017.10.011

Rezazadeh Sereshkeh, A., Trott, R., Bricout, A., Chau, T., 2017a. EEG Classification of Covert Speech Using Regularized Neural Networks. IEEE/ACM Trans. Audio Speech Lang.

1268   Process. 25, 2292–2300. https://doi.org/10.1109/TASLP.2017.2758164

1269   Rezazadeh Sereshkeh, A., Trott, R., Bricout, A., Chau, T., 2017b. Online EEG Classification
1270   of Covert Speech for Brain–Computer Interfacing. Int. J. Neural Syst. 27, 1750033.
1271   https://doi.org/10.1142/S0129065717500332

1272   Riès, S.K., Dhillon, R.K., Clarke, A., King-Stephens, D., Laxer, K.D., Weber, P.B., Kuperman,
1273   R.A., Auguste, K.I., Brunner, P., Schalk, G., Lin, J.J., Parvizi, J., Crone, N.E., Dronkers,
1274   N.F., Knight, R.T., 2017. Spatiotemporal dynamics of word retrieval in speech production
1275   revealed by cortical high-frequency band activity. Proc. Natl. Acad. Sci. 114, E4530–
1276   E4538. https://doi.org/10.1073/pnas.1620669114

1277   Ruescher, J., Iljina, O., Altenmüller, D.M., Aertsen, A., Schulze-Bonhage, A., Ball, T., 2013.
1278   Somatotopic mapping of natural upper- and lower-extremity movements and speech
1279   production with high gamma electrocorticography. Neuroimage 81, 164–177.
1280   https://doi.org/10.1016/j.neuroimage.2013.04.102

1281   Scott, M., Yeung, H.H., Gick, B., Werker, J.F., 2013. Inner speech captures the perception of
1282   external speech. J. Acoust. Soc. Am. 133, EL286-EL292.
1283   https://doi.org/10.1121/1.4794932

1284   Sedivy, J., 2014. Language in mind: An introduction to psycholinguistics. Sinauer Associates,
1285   Incorporated, Sunderland, MA.

1286   Segalowitz, S.J., Lane, K., 2004. Perceptual fluency and lexical access for function versus
1287   content words. Behav. Brain Sci. 27, 2018.

1288   Shergill, S., Bullmore, E., 2001. A functional study of auditory verbal imagery. Psychol. Med.
1289   31, 241–253.

1290   Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M., McGuire,
1291   P.K., 2002. Modulation of activity in temporal cortex during generation of inner speech.
1292   Hum. Brain Mapp. 16, 219–227. https://doi.org/10.1002/hbm.10046

1293   Shuster, L.I., Lemieux, S.K., 2005. An fMRI investigation of covertly and overtly produced
1294   mono- and multisyllabic words. Brain Lang. 93, 20–31.
1295   https://doi.org/10.1016/j.bandl.2004.07.007

1296   Song, Y., Sepulveda, F., 2014. Classifying speech related vs. idle state towards onset detection
1297   in brain-computer interfaces overt, inhibited overt, and covert speech sound production
1298   vs. idle state. IEEE 2014 Biomed. Circuits Syst. Conf. BioCAS 2014 - Proc. 568–571.

1299   https://doi.org/10.1109/BioCAS.2014.6981789

1300 Stark, B.C., Geva, S., Warburton, E.A., 2017. Inner Speech's Relationship With Overt Speech
1301   in Poststroke Aphasia. J. Speech, Lang. Hear. Res. 26, 1–10. https://doi.org/10.1044/2017

1302 Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T.,
1303   Arévalo, A.L., Vargha, A., Bates, E., 2005. Timed action and object naming. Cortex 41,
1304   7–25. https://doi.org/10.1016/S0010-9452(08)70174-6

1305 Tabar, Y.R., Halici, U., 2017a. A novel deep learning approach for classification of EEG motor
1306   imagery signals. J. Neural Eng. 14, 016003. https://doi.org/10.1088/1741-
1307   2560/14/1/016003

1308 Tabar, Y.R., Halici, U., 2017b. Brain Computer Interfaces for Silent Speech. Eur. Rev. 25,
1309   208–230. https://doi.org/10.1017/S1062798716000569

1310 Tian, X., Poeppel, D., 2013. The Effect of Imagination on Stimulation: The Functional
1311   Specificity of Efference Copies in Speech Processing. J. Cogn. Neurosci. 25, 1020–1036.
1312   https://doi.org/10.1162/jocn

1313 Torres-García, A.A., Reyes-García, C.A., Villaseñor-Pineda, L., García-Aguilar, G., 2016.
1314   Implementing a fuzzy inference system in a multi-objective EEG channel selection model
1315   for imagined speech classification. Expert Syst. Appl. 59, 1–12.
1316   https://doi.org/10.1016/j.eswa.2016.04.011

1317 Vigneau, M., Beaucousin, V., Herv??, P.Y., Duffau, H., Crivello, F., Houd??, O., Mazoyer, B.,
1318   Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: Phonology,
1319   semantics, and sentence processing. Neuroimage 30, 1414–1432.
1320   https://doi.org/10.1016/j.neuroimage.2005.11.002

1321 Wang, W., Degenhart, A.D., Sudre, G.P., Pomerleau, D.A., Tyler-Kabara, E.C., 2011.
1322   Decoding semantic information from human electrocorticographic (ECoG) signals. Proc.
1323   Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 6294–6298.
1324   https://doi.org/10.1109/IEMBS.2011.6091553

1325 Wang, Y.Y., Wang, P., Yu, Y., 2018. Decoding English alphabet letters using EEG phase
1326   information. Front. Neurosci. 12, 1–10. https://doi.org/10.3389/fnins.2018.00062

1327 Wheeldon, L.R., Levelt, W.J.M., 1995. Monitoring the Time Course of Phonological Encoding.
1328   J. Mem. Lang. 34, 311–334.

1329   Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M.A., Jefferies, E., 2012. Executive
1330   Semantic Processing Is Underpinned by a Large-scale Neural Network: Revealing the
1331   Contribution of Left Prefrontal, Posterior Temporal, and Parietal Cortex to Controlled
1332   Retrieval and Selection Using TMS. J. Cogn. Neurosci. 24, 133–147.
1333   https://doi.org/10.1162/jocn_a_00123

1334   Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex,
1335   basal ganglia, and cerebellum to speech motor control: Effects of syllable repetition rate
1336   evaluated by fMRI. Neuroimage 13, 101–109. https://doi.org/10.1006/nimg.2000.0672

1337   Yoshimura, N., Nishimoto, A., Belkacem, A.N., Shin, D., Kambara, H., Hanakawa, T., Koike,
1338   Y., 2016. Decoding of covert vowel articulation using electroencephalography cortical
1339   currents. Front. Neurosci. 10, 1–15. https://doi.org/10.3389/fnins.2016.00175

1340   Zaccarella, E., Friederici, A.D., 2016. The neurobiological nature of syntactic hierarchies.
1341   Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2016.07.038

1342   Zaccarella, E., Friederici, A.D., 2015. Merge in the human brain: A sub-region based functional
1343   investigation in the left pars opercularis. Front. Psychol. 6, 1–9.
1344   https://doi.org/10.3389/fpsyg.2015.01818

1345   Zatorre, R.J., Halpern, A.R., 2005. Mental concerts: Musical imagery and auditory cortex.
1346   Neuron 47, 9–12. https://doi.org/10.1016/j.neuron.2005.06.013

1347   Zhang, D., Gong, E., Wu, W., Lin, J., Zhou, W., Hong, B., 2012. Spoken sentences decoding
1348   based on intracranial high gamma response using dynamic time warping. Proc. Annu. Int.
1349   Conf. IEEE Eng. Med. Biol. Soc. EMBS 3292–3295.
1350   https://doi.org/10.1109/EMBC.2012.6346668

1351   Zhao, S., Rudzicz, F., 2015. Classifying phonological categories in imagined and articulated
1352   speech. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2015–Augus,
1353   992–996. https://doi.org/10.1109/ICASSP.2015.7178118

1354

1355

1356 **Figure Descriptions**

1357 **Figure 5 Seeking a Naturalistic form of communication through Direct-speech BCI. A**
1358 **DS-BCI is a system that decodes neural signals (e.g., electroencephalography (EEG) or**

electrocorticography (ECoG) (B) corresponding to imagined speech (A). Recorded signals are processed to facilitate maximal information extraction and improvement of signal-to-noise ratio (C). The feature extraction (D) and classification (E) stages compute the most discriminative information in the recorded signals and classify them as a part-of-speech. The output of a DS-BCI system is a textual representation of the imagined speech (F) and auditory representation which can be used for both communication and feedback (G).

In this example, the user actively produces the words "I am thirsty!" with imagined speech. The signals acquired are temporally-aligned with each word to facilitate feature extraction and classification. The system produces two outputs: a text print-out of the imagined speech words being produced and a synthesised audio output, i.e. "I am thirsty!"

Figure 6 Direct-speech BCI studies categorised according to recording techniques and types of speech. (A) is a cross-categorisation of DS-BCI studies according to the recording techniques applied and the types of speech being investigated. The time-period for this analysis begins with the study of Blakely et al. (Blakely, Miller, Rao, Holmes, & Ojemann, 2008), due to this being the first study based on the BCI paradigm depicted, and runs to 2018. Criteria for inclusion in this analysis are those studies using said recording techniques to decode speech production (overt, imagined, and intended) directly from neural activity. EEG and ECoG are the most-often used recording approaches. High temporal resolution is an important feature of both. Although micro-electrodes do offer high spatial and temporal resolution, their use is not always possible or appropriate. Overt speech has been used as a proxy for imagined speech, or in comparative studies. The behavioural difficulty of studying imagined speech is, at least in-part, a reason for this trend. The two bar-graphs (B) show the distribution of measurement techniques and of types of speech used across all studies. ECoG is utilised in a total of twenty studies and EEG in a total of sixteen. See Table 2.

Figure 7 Neuroanatomical regions associated with imagined speech production. The diagram depicts brain regions typically associated with language function in the left-hemisphere (Berwick, Friederici, Chomsky, & Bolhuis, 2013), with each of the numbered sections indicating one of Brodmann's Areas (BA). The IFG, which includes BA44 and BA45, is the most common region associated with imagined speech production. Single word and sentence production both activate the IFG, and the region is thought to be associated with word retrieval and associated meanings (BA45). Both the STG and MTG

have been implicated in imagined speech studies as relating to the phonological loop and to production of dialogic imagined speech. The dorsal pathways between BA44 and the posterior superior temporal cortex (pSTC) supports core syntactic processes. The ventral pathways, including between BA45 and the temporal cortex (TC), support processing of semantic and conceptual information. Reprinted with permission from (Berwick et al., 2013), copyright 2013, Elsevier.

Figure 8 Speech production models with estimated time courses. Although models can differ in the number of components, there is general agreement that speech production is a staged, hierarchical process with a temporal structure, as indicated in the diagram. In (A), estimated time courses associated with the stages of production are provided in milliseconds (ms) (Indefrey, 2011) along with a production model containing two major components. These are the word (lemma) level and the phonological level (Hickok, 2012). In (B), a more detailed model depicts several different phases in the production process (Levelt, Roelofs, & Meyer, 1999). The initial stage is conceptual preparation, where a message to be expressed is formulated and a lexical concept produced. Next is lexical selection, in which a word or lemma is retrieved for use. Following selection of a lemma, the morphological stage bridges between the conceptual domain and the phonological, or articulatory domain. A word is then encoded in syllabic form before being encoded in phonetic form, from which the audible output is produced. In (C), a truncated version of the model in (A) is presented to highlight the stages of production corresponding to imagined speech. The estimated time courses end with the phonological encoding/syllabification stage. A is adapted with permission from (Hickok, 2012), copyright 2012, Springer Nature. B is adapted with permission from (Levelt, Roelofs, & Meyer, 1999), copyright 1999, Cambridge University Press. *upper boundary.

1422

1423

1424

1425

1426

1427