Fachbereich Philosophie und Geisteswissenschaften
der Freien Universität Berlin

# Establishing action-perception circuits as a neural basis for meaning-carrying linguistic symbols – the role of frontal speech motor areas and fronto-temporal connectivity

Dissertation

zur Erlangung des akademischen Grades
Doctor of Philosophy (Ph.D.)

vorgelegt von

Malte Schomers

Berlin, 2016

Erstgutachter: Prof. Dr. Dr. Friedemann Pulvermüller

Zweitgutachter: Prof. Dr. Felix Blankenburg

Tag der Disputation: 28. Februar 2017

# Table of Contents

# Abstract

This dissertation investigates how brain language models can explain the neural representation of phonological word forms and their function in speech comprehension. Action-perception theory of language provides a possible explanation based on Hebbian learning principles. In this perspective, word forms are thought to be represented by cell assemblies, or action-perception circuits (APCs), which are distributed across frontal and temporal cortices, as well as inferior parietal areas. These areas are referred to collectively as the perisylvian language areas. While it is generally acknowledged that the perisylvian language areas function as the core regions for language, considerable debate and open questions still exist regarding the precise role of the different perisylvian regions, and in particular concerning the role of long-distance connectivity within the perisylvian areas along the dorsal stream by way of the arcuate fasciculus (AF). Strong AF connectivity is present only in humans, but not in non-human primates, and therefore has been proposed as the specific substrate underlying human word learning abilities. However, a neurobiological explanation why AF connectivity is so important remains to be found. A neuroanatomically grounded and neurophysiologically plausible computational model simulating the emergence of phonological action-perception circuits was used to investigate this. Results suggest that the AF is critical for reverberating activity across the perisylvian cortex, which in turn functions as the substrate for verbal working memory. A further question is what the function of this auditory-motor link is for speech comprehension. One perspective claims that speech comprehension relies on a single lexical interface in temporal cortex, and that any influence of motor systems on speech processing is restricted to ‘artificial’ phonological tasks or particularly noisy listening conditions. In contrast, action-perception theory of language predicts that as a result of Hebbian-based distributed APCs, fronto-parietal systems, in particular motor systems, can take an additional causal role in speech comprehension. In line

with this, there is compelling evidence for a contribution of fronto-parietal systems to speech comprehension, even in noise-free and passive tasks, as demonstrated by an extensive review of functional magnetic resonance imaging (fMRI) studies, especially recent ones employing multivariate pattern analysis (MVPA), as well as transcranial magnetic stimulation (TMS) studies. Specifically, the question whether motor systems causally contribute not only to phonological tasks using isolated, meaningless syllables, but also to comprehension of meaningful words, is addressed in a TMS study where phonemes are placed in the context of whole words, hence allowing them to serve their 'normal' function as meaning-discriminating units.

In summary, the results of this dissertation demonstrate that the neural representation of word forms can be explained in terms of APCs distributed across perisylvian cortex. Strong long-distance connectivity between the perisylvian areas is an important prerequisite for building robust APCs, by enabling verbal working memory, which in turn is important both for word learning and language comprehension. The brain mechanisms for word learning and speech comprehension are thus widely distributed across the perisylvian areas and also crucially rely on long-distance structural connectivity between these areas.

# Zusammenfassung

Die vorliegende Arbeit stellt die Frage, wie neurobiologische Sprachmodelle die neuronale Repräsentation von Wörtern und die Funktion dieser Repräsentationen für Sprachverständnis erklären können. Die Handlungs-Wahrnehmungs-Theorie (engl. *action perception theory)* für Sprache stellt eine mögliche, auf Hebb'schen Lernprinzipien basierende Erklärung in Aussicht. In dieser Sichtweise werden Wörter durch Aktions-Perzeptions-Neuronenverbünde (APNVs) repräsentiert, welche über die frontalen und temporalen Hirnlappen sowie den inferioren Parietallappen verteilt sind. Zusammengefasst werden diese Areale auch als perisylvische Areale bezeichnet. Dass die perisylvischen Areale die Kernregionen für Sprachfunktion sind, ist zwar gemeinhin akzeptiert, beträchtliche Meinungsverschiedenheit und offene Fragen bestehen aber in Bezug auf die genaue Rolle der verschiedenen perisylvischen Regionen, und ferner in Bezug auf die Rolle von weitreichenden Verbindungen innerhalb der perisylvischen Areale durch den dorsalen Pfad, dessen direkte Verbindung durch ein starkes Bündel aus Nervenfasern, dem *Fasciculus arcuatus* (engl. *arcuate fasciculus,* AF) hergestellt wird. Besonders starke AF-Konnektivität besteht nur bei Menschen, nicht aber bei nicht-menschlichen Primaten; daher sieht man dies als mögliche Grundlage für die menschliche Fähigkeit, Wörter zu erlernen. Eine neurobiologische Erklärung, warum AF-Konnektivität so wichtig ist, steht jedoch noch aus. Um dieser Frage nachzugehen, wurde ein neuroanatomisch fundiertes und neurophysiologisch plausibles Computermodell benutzt, welches die Entstehung von phonologischen Aktions-Perzeptions-Schaltkreisen simuliert. Die Ergebnisse legen nahe, dass der AF entscheidend ist für das Entstehen von lang anhaltender, „nachhallender" Aktivität im gesamten perisylvischen Kortex, was wiederum als neuronale Grundlage für phonologisches Arbeitsgedächtnis dient. Eine weitere Frage betrifft die Funktion dieser auditorisch-motorischen Verbindung entlang des dorsalen Pfades. Eine Sichtweise postuliert, dass Sprachverständnis auf einer

„lexikalische Schnittstelle“ im Temporallappen basiert, und dass mögliche Einflüsse des motorischen Systems sich lediglich auf künstliche phonologische Aufgabenstellungen oder besonders geräuschvolle Hörumgebungen beschränken. Im Gegensatz hierzu sagt die Handlungs-Wahrnehmungs-Theorie der Sprache, dass aufgrund der APNVs, die durch Hebb'sches Lernen entstanden sind, fronto-parietale Hirnregionen, insbesondere motorische Regionen, eine zusätzliche kausale Rolle beim Sprachverständnis spielen können. Im Einklang damit konnten überzeugende Beweise dafür gefunden werden, dass fronto-parietale Hirnregionen zum Sprachverständnis beitragen, sogar ohne aktive Aufgabenstellung und in geräuschlosen Umgebungen. Dies konnte durch einen Überblick über Studien mit funktioneller Magnetresonanztomographie gezeigt werden, insbesondere durch neuere Studien, welche multivariate Analyse der Aktivitätsmuster benutzten, sowie zusätzlich durch Studien, die mit Hilfe der transkraniellen Magnetstimulation (TMS) durchgeführt wurden. Die spezifische Fragestellung, ob motorische Systeme eine kausale Rolle nicht nur auf phonologische Aufgaben mit isolierten, bedeutungslosen Silben ausüben, sondern auch auf das Verständnis bedeutungsvoller Wörter, wurde mit einer TMS-Studie untersucht, in der Phoneme im Kontext ganzer Wörter präsentiert wurden, sodass sie ihre „normale“ Rolle als bedeutungsunterscheidende Einheiten erfüllten.

Zusammenfassend lässt sich sagen: die Ergebnisse dieser Dissertation legen dar, dass die neuronale Repräsentation von Wörtern durch Aktions-Perzeptions-Neuronenverbünde erklärt werden kann, welche über den gesamten perisylvischen Kortex verteilt sind. Eine starke direkte Verbindungsbahn dieser Areale durch den AF ist eine wichtige Voraussetzung, um starke APNVs zu bilden, indem phonologisches Arbeitsgedächtnis ermöglicht wird, was wiederum sowohl für das Erlernen als auch für das Verstehen von gesprochenen Wörtern wichtig ist. Die Hirnmechanismen für das Erlenen und Verstehen gesprochener Sprache sind also weit über die perisylvischen Areale verteilt und basieren außerdem entscheidend auf direkten Nervenfaserverbindungen zwischen diesen Arealen.

# List of abbreviations

AF – arcuate fasciculus

APC – action perception circuit

BA – Brodmann area

CA – cell assembly

CV – consonant-vowel

DTI – diffusion tensor imaging

DWI – diffusion weighted imaging

ECoG – electrocorticography

EEG – electroencephalography

fMRI – functional MRI

IFG – inferior frontal gyrus

MEG – magnetoencephalography

MEP – motor-evoked potential

MRI – magnetic resonance imaging

MTG – posterior middle temporal gyrus

MTSP – motor theory of speech perception

MVPA – multivariate pattern analysis

PMC – premotor cortex

dPMC, vPMC – dorsal/ventral PMC

PoA – place of articulation

RSA – representational similarity analysis

SMA – supplementary motor area

SMG – supramarginal gyrus

STG – superior temporal gyrus

STS – superior temporal sulcus

TMS – transcranial magnetic stimulation

rTMS – repetitive TMS

VWM – verbal working memory

# 1. Introduction

Language is one of the most powerful of all human cognitive abilities, and the human language faculty differs strikingly from those of non-human primates and other animals. One important feature that makes language such a powerful tool is its capacity for creating a vast repertoire of words which serve as symbols that can arbitrarily be associated with meaning (de Saussure, 1964). This dissertation explores the question as to which brain mechanisms underly the process of learning words as a basis for meaning-carrying linguistic symbols. The focus will be on the neuronal representation of word forms itself, reflecting the sound of a word (acoustic-perceptual properties) as well as knowledge about how to produce it (articulatory-motor properties).

Specifically, this dissertation deals with neurobiological theories that attempt to explain how humans can learn word forms, what their underlying neuronal representation is and what role these representations play in language understanding, i.e. in the process of perceiving and understanding a known spoken word. Aspects of semantics and syntax (e.g., recursion) undoubtedly also contribute to making language such a powerful tool. However, this dissertation will focus solely on words in the sense of articulatory-acoustic symbols, which constitute the fundamental ‘building block’ on which other language abilities rest.

I will start by introducing classic 19th century neurological models of language as well as alternative theories, in particular action-perception theory of language. I will then highlight some open questions about the neurobiological representation of word forms and their role in speech comprehension, which will be addressed in the studies presented in Chapters 2 through 4.

## Classic neurological models of language

Since the earliest neuropsychological investigations on the localization of language in the brain, most notably by Broca (1861) and Wernicke (1874), it has been known that the areas that are most important for language are, on the one hand, the inferior frontal cortex (containing Broca's area) as well as neighboring motor and premotor areas, and on the other hand, superior temporal cortex including Wernicke's area[1]. Based on early patient evidence, it initially seemed that there was a clear demarcation between speech production and perception. Lichtheim's model (Lichtheim, 1885) therefore proposed that language was represented in the brain in 3 "centers": a motor language center ("Wortbewegungsbilder", M) for production, an acoustic/auditory language center ("Wortklangbilder", A) and concept center ("Begriffe", B).

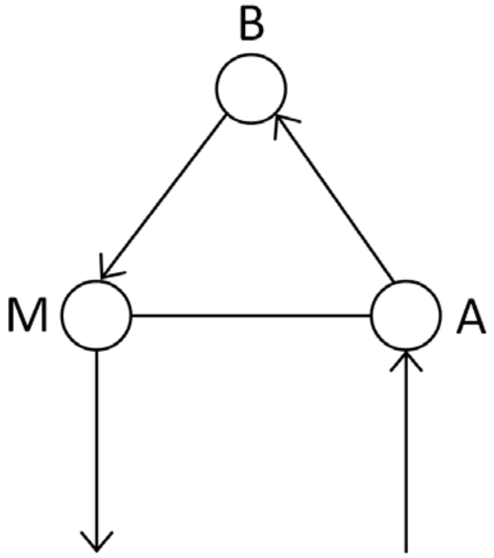


**Figure 1.1.** Lichtheim's model of language representation, proposing a motor movement center ("Wortbewegungsbilder", M) for speech production, an acoustic/auditory language center ("Wortklangbilder", A) and a concept center ("Begriffe", B). Note that connections between the centers are believed to be unidirectional only. Adapted from Lichtheim (1885)

---

[1] There is no universally accepted precise definition of either Broca's area or Wernicke's area (Bogen and Bogen, 1976; Damasio and Damasio, 1980; Mesulam, 1990; Amunts et al., 1999). However, in contrast a recent proposal to completely abandon these terms (Tremblay and Dick, 2016), I still consider them useful as broad terms when discussing neurobiological models of language representation.

Note that the different areas were thought to be connected, but with unidirectional flow of information only. An influence of the production center (M) on auditory processing (A) is not viewed as possible by this model. This model was later revived by Geschwind (1970) and is now often referred to as the Wernicke-Lichtheim-Geschwind (WLG) model, which continues to be popular today, being taught in many textbooks. However, this model is most likely incorrect, certainly in a strict sense. It is true that frontal lesions tend to produce impairments in speech production rather than perception and temporal lesions tend to produce impairments in speech perception rather than production. However, there are also cases where speech production deficits are caused by damage to Wernicke's area (Fridriksson et al., 2007, 2015) and conversely, speech perception or comprehension deficits caused by damage to Broca's area (Basso et al., 1977; Miceli et al., 1980; Shewan, 1980; Utman et al., 2001; Moineau et al., 2005). Furthermore, the damage in Broca's aphasia patients often significantly exceeds Broca's area (Mohr et al., 1978), even in Broca's original two patients, whose data have been reanalyzed by Dronkers et al. (2007), suggesting that Broca's area damage alone is not sufficient for Broca's aphasia. Computational models (Pulvermüller and Preissl, 1991) as well as, more recently, voxel-based lesion symptom mapping across a large dataset of patient data (Bates et al., 2003) were able to confirm that in general, lesions in either frontal or temporal areas can produce deficits both in production and perception, as well as in speech fluency and comprehension. Neuroimaging evidence is also consistent with this view. Early positron emission tomography (PET) studies could already demonstrate that both speech perception and production activate temporal and frontal areas (Petersen et al., 1988; Zatorre et al., 1992; Paus et al., 1996). A meta-analysis of 129 functional magnetic resonance imaging (fMRI) studies by Vigneau et al. (2006; see Figure 1.2) also clearly shows that language processing, across a range of tasks (phonological, semantic and sentence-related) causes widely distributed activation across large parts of the temporal and frontal cortices, as well as parts of the inferior parietal lobe (notably the angular gyrus and supramarginal gyrus).

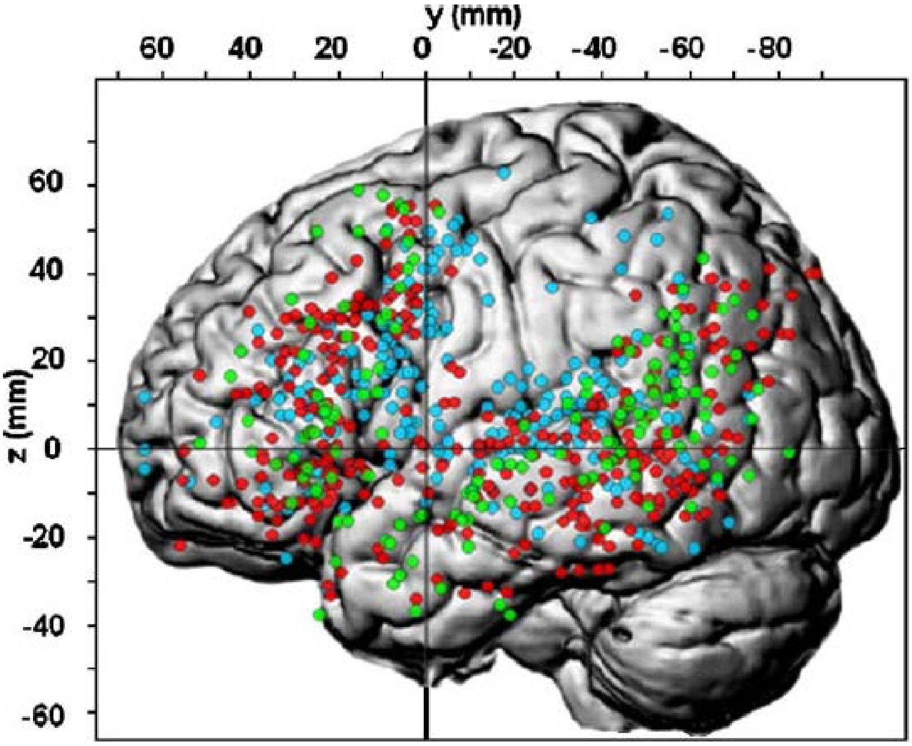


**Figure 1.2:** meta-analysis of 129 fMRI studies by Vigneau et al. (2006). Activation peaks are shown for phonological tasks (blue), semantic tasks (red), and sentence processing (green). Reprinted from "Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing", M. Vigneau, V. Beaucousin, P.Y. Hervé, H. Duffau, F. Crivello, O. Houdé, B. Mazoyer, N. Tzourio-Mazoyer, 30, pp.1414-32, copyright 2005, with permission from Elsevier

Despite clear evidence against this classic ‘modular’ view of language processing which views Broca’s area and inferior frontal cortex as responsible for expressive language and Wernicke’s area as responsible for receptive language processing, this perspective is still prevalent and taught in many textbooks. In particular, one of the most prominent contemporary models of speech processing in the brain (Hickok and Poeppel, 2004, 2007) embraces this modular view of the WLG model. In particular, they propose that “sound to meaning”, or speech recognition, is subserved by the ventral stream with the middle temporal gyrus serving as a “lexical interface, which links phonological and semantic information”. In contrast, the dorsal stream is claimed to play a role only in mapping “sound to action”, i.e. for auditory-motor integration. Crucially, Hickok and Poeppel’s view is that the auditory-motor integration provided by the dorsal stream relate only to functions such as speech segmentation or verbal working memory (referred to as phonological short-term memory by some authors) and play a role in speech development (learning to produce speech sounds) as well as novel word acquisition, but not later on in normal speech understanding. As they state, “although the proposed dorsal stream represents a tight connection between processes involved in speech perception and speech production, it does not appear to be a critical component of the speech perception process under normal (ecologically natural) listening conditions, that is when speech input is mapped onto a conceptual representation” (Hickok and Poeppel, 2004, p. 73).

The ongoing controversy between classic ‘modular’ neurological language models and alternative proposals which view language representations as distributed across temporal, parietal and frontal regions is a major theme of the present dissertation. Therefore, in the next section an alternative brain language theory will be presented which attempts to capture the observation that language function seems to be multimodal and distributed across a wide range of cortical areas.

## Action-perception theory of language

This theory of action-perception circuits as a cortical basis for language processing is based on the well-established principle of Hebbian learning. Donald Hebb postulated that “[w]hen an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased” (Hebb, 1949). This principle was later paraphrased in the short mnemonic “what fires together, wires together” (Shatz, 1992). Conversely, it is also the case that “neurons out of sync, delink” (known as Anti-Hebbian learning) (Artola and Singer, 1993; O’Reilly, 1998). These two learning rules, which are biologically underpinned by long-term potentiation (LTP) and long-term depression (LTD) (Tsumoto, 1992) constitute fundamental principles of brain functioning. Based on Hebbian learning, groups of neurons that are frequently active together will develop a so-called “cell assembly”, a concept already introduced by Hebb himself and later refined by

Valentino Braitenberg (Braitenberg, 1978; Huyck and Passmore, 2013). A cell assembly is a set of neurons which are strongly linked to each other through excitatory connections. Due to the strong internal connections of a cell assembly, reactivating only a small subset of the assembly can trigger (or at least facilitate) ignition of the entire cell assembly. Note that similar proposals for distributed neuronal ensembles supporting language function as well as other cognitive functions (e.g. attention, memory) have also been put forth by other authors (Mesulam, 1990; Fuster, 2003, 2006), and other terms for "cell assemblies" are used by those authors, such as "cognit" by Fuster.

An important point about cell assembly theory is that the resulting assemblies need not be locally restricted, but can be widely distributed across the whole cortex. For example, Fuster and Bressler (2012) write: "The neurons [of a cognit] may be situated within a single cortical area, or dispersed across several contiguous or noncontiguous areas".

Hence, the cognitive function which a distributed cell assembly (CA) subserves cannot be assigned to any single brain area. The facts that (i) cell assemblies have strong internal connections and (ii) can be spatially dispersed across distinct brain regions also entails that they are somewhat robust to damage and hence might be the neuronal basis of the "cognitive degeneracy" (Price and Friston, 2002) observed in many cognitive functions including language, as mentioned earlier.

The theory of cell assemblies as a cortical basis for cognitive functions is also in line with embodied cognition views (Barsalou, 1999, 2008; Glenberg and Kaschak, 2002), which posit that human cognitive functions (of which language is a prime example) traditionally viewed as abstract and amodal, i.e. detached from basic sensory processing, are better conceptualized as being based on, and interlinked with, more basic sensory or motor functions/mechanisms. Evolutionarily, this has been explained as a case of "neural reuse" (Anderson, 2010, 2015), where it is claimed that a previously purely sensory neuron evolved to take on semantic functions, for example.

As a consequence of this view, a given cognitive function cannot necessarily be localized to one single brain area. The neuronal circuit representing and processing the meaningful word "fish" may be spread out across articulatory motor and auditory sensory cortex along with inferior frontal and superior temporal perisylvian 'language areas', encompassing 'semantic hub' areas in anterior temporal and prefrontal cortex and even reaching into sensory visual or motor cortex where knowledge about how to recognize or catch a fish is stored (Pulvermüller, 2013). As will be discussed in more detail later on, such proposals are therefore incompatible with strongly modular or localizationist models of language, which explicitly posit the existence of a single "lexical interface" or a "gateway to understanding" (e.g., Hickok and Poeppel, 2004, 2007; Hickok, 2014).

### *Cell assemblies for spoken word forms*

Where might cell assemblies for spoken language be represented, according to action-perception theories? An important observation that features prominently in such theories (Braitenberg and Pulvermüller, 1992; Pulvermüller, 1999, 2005; Pulvermüller and Fadiga, 2010; Pulvermüller et al., 2014b; Pulvermüller and Fadiga, 2016), is that in (undeprived) humans, there is frequently correlated activity across all perisylvian areas. This is the case because humans typically hear themselves when speaking, i.e. language production yields auditory self-perception, starting with the so-called babbling phase in infants, when they spontaneously start uttering sounds (Locke, 1995). Furthermore, there is also a degree of somatosensory self-stimulation (because of the different articulators touching each other, such as when the lips touch each other to produce a bilabial sound).

All the brain areas active during speech production and perception as well as somatosensory processing should therefore be included in a cell assembly perspective on language (Figure 1.3). They are collectively termed the perisylvian brain areas because they lie close to the Sylvian fissure (or lateral fissure).

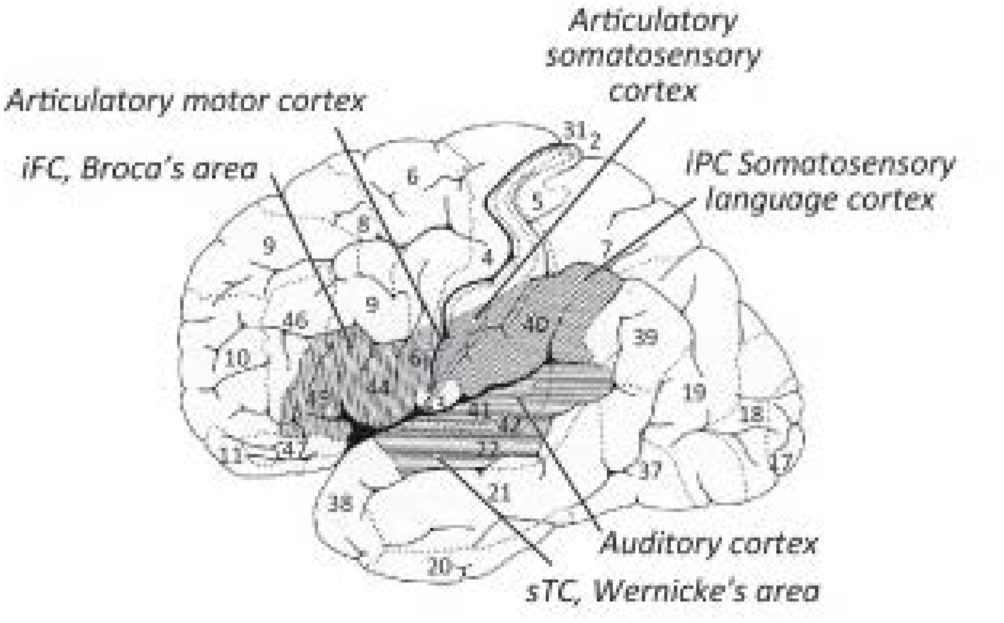


**Figure 1.3:** Overview of perisylvian language areas supporting production and understanding of spoken language. Reprinted from Pulvermüller et al. (2013), published under a Creative Commons license (BY-NC-ND)

These areas necessarily include the primary sensory and primary motor areas that are necessary for speech perception and production. Incoming acoustic speech information is initially processed in superior temporal cortex, which houses the primary auditory and secondary auditory cortices and includes Wernicke's area. In contrast, producing speech sounds requires the articulatory motor cortex which sends afferent signals to the muscles controlling speech articulators such as the tongue and lip. Because speech production also involves a degree of somatosensory self-stimulation of the articulators, the 'somatosensory language cortex' is also part of the perisylvian language areas (note the similarity of the perisylvian brain area model to the pattern of activation observed in the meta-analysis of fMRI studies shown in Figure 1.2).

Higher-order motor areas such as premotor and Broca's area are additionally involved in speech production, e.g. in the planning stages and in sequencing, although the exact role of these secondary motor areas is less clear than that of the primary motor areas[2]. Recent electrocorticography (ECoG) evidence shows that Broca's area is active immediately prior to, but not *during* the actual articulator movements, and therefore it has been proposed that its function in speech production is not to control the speech articulator movements directly, but rather to "coordinate the transformation of information processing across large-scale cortical networks involved in spoken word production, prior to articulation" (Flinker et al., 2015, p. 2871; cf. also Watkins and Paus, 2004). Thus, Broca's area should perhaps not be conceptualized as belonging to the speech production system in the narrow sense. Nevertheless, it is active until immediately prior to articulation, and therefore likely to be causally involved in preparing speech production. Furthermore, even if Broca's area is not active *during* speech production, it is clearly a necessary area for speech production, as "speech arrest" – where the ability to produce speech is temporarily, but severely impaired – can be induced by stimulating Broca's area electrically (Penfield and Roberts, 1959), as well as, more recently, using transcranial magnetic stimulation (TMS) (Pascual-Leone et al., 1991; Epstein et al., 1999; Stewart et al., 2001).

Therefore, starting with the babbling phase, but also throughout their later life, humans frequently experience correlated activity in auditory and motor areas, and to some extent also somatosensory areas. Because of Hebbian learning mechanisms, this predicts the emergence of cell assemblies (or action-perception circuits, APCs) across frontal, temporal and inferior parietal areas. Pulvermüller (1999) refers to these as the word-form circuits in the perisylvian areas. These phonological APCs can be viewed as a neuronal representation of the word form which in itself does not carry any meaning, i.e. the APC for a word form only represents its acoustic, articulatory and somatosensory properties and is therefore confined to the

---

[2] Although beyond the scope of the present dissertation, even the supplementary motor area as well as the cerebellum also play roles in speech production (Ackermann et al., 2007; Hertrich et al., 2016)

perisylvian areas. Crucially, however, an APC for a word form establishes the basis for the representation of meaning, for which additional neurons in extrasylvian areas (i.e. brain areas other than the perisylvian areas) are bound into the overall cell assembly. This again occurs through Hebbian learning, through co-activation of the neurons in perisylvian areas (representing the word form) and neurons in extrasylvian areas (representing meaning) (Pulvermüller 2013), yielding a higher-order APC representing articulatory-acoustic *and* semantic features of that word.

In summary, action-perception theory predicts that APCs for word forms are distributed across temporal, frontal and parietal areas and are strongly linked due to Hebbian learning. As a result of this, a core prediction of this view is that speech perception and comprehension should be subserved by all perisylvian areas.

## Motor theory of speech perception

Another very influential theory pertaining specifically to speech perception, which also placed strong emphasis on auditory-motor links, is the motor theory of speech perception (MTSP) (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000). This theory has certain similarities with action-perception theory of language, but also some important differences. An obvious difference is that the MTSP focused only on the perceptual analysis of speech, whereas action-perception theory also proposes how meaning can be represented in cell assemblies. But, as will be shown, important differences, especially with regard to modularity also exist.

The MTSP was originally proposed in the 1960s by Alvin Liberman. He discovered that there was a lack of invariance between the motor commands (“articulatory gestures”) used to produce a speech sound and the acoustic representation of the resulting speech sound. Due to coarticulation effects, for example, the same articulator movements can result in widely differing acoustic representations. Importantly, however, the listener always perceives the same phoneme despite this lack of invariance in the acoustic representation. In other words, perception follows the (relative) invariance present in the articulatory gestures; it therefore relates more to the (relatively invariant) motor properties of a speech sound rather than its highly variable acoustic properties.

In order to account for this observation, the core claim of Liberman's theory was that perceiving speech consists of perceiving the "intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations" (Liberman & Mattingley, 1985, p.2). More specifically, Liberman proposed that in perceiving the intended gestures, the pattern of the underlying phonetic features of a gesture was perceived. A corollary of the claim that perceiving speech is perceiving the intended articulatory gestures is that there should be intimate links between speech perception and production. Note that Liberman himself originally did not make any reference to the underlying brain systems representing this link. However, that aspect of the MTSP is compatible with the proposition of action-perception theory that Hebbian learning leads to strong links between the brain areas mediating speech perception and production (and that, as a result of strong cell assemblies, one cannot clearly delineate speech production from perception areas). An important difference between Liberman's theory and action-perception theory, however, is that Liberman proposed that the production-perception link was achieved by a "phonetic module", which he conceptualized as modular in the sense of Fodor (1983), e.g., exhibiting informational encapsulation. Furthermore, he claimed the module to be innate. This is in sharp contrast to action-perception theory, which states that, firstly, the perception-production link is a result of experience and Hebbian learning (i.e. the links are not already present in a "phonetic module" or some similar way at birth), and secondly, that APCs can be in principal represented in any brain area and even distributed across several brain areas, hence defying the concept of strictly segregated modules. Rather, the APCs for word forms can be built on the general auditory and motor areas of the brain which might also be involved in non-linguistic auditory processing (e.g. of music or non-linguistic sounds) and in non-linguistic motor programming (i.e. motor actions in general). As a consequence of the fact that APCs can be distributed across brain areas (for example auditory and motor areas), action-perception theory also predicts that the distributed circuit as a whole is invoked for perceptual analysis of phonemes, not the motor system on its own, as in Liberman's proposed articulatory-phonetic module.

## The role of fronto-temporal connectivity

As outlined in the previous section, one important difference between the MTSP and action-perception theory of language is that the MTSP posited innate links between speech production and perception, whereas action-perception theory does not. When considering what brain features underlying language function are innate, it is important, however, to distinguish between the predisposition or 'machinery' for establishing action-perception circuits and the links themselves. Although the *links themselves* might not be innate, certain aspects of brain anatomy and connectivity might nonetheless be a prerequisite for establishing these links through experience. Specifically, since the auditory and motor areas in the brain do not lie side by side, strong fronto-temporal connectivity linking these areas might be necessary to provide the potential for distributed APCs to emerge. In line with this, numerous anatomical fiber bundles connecting frontal and temporal regions have been described (Dick and Tremblay, 2012; Gierhan, 2013; Dick et al., 2014). Using electrocorticography (ECoG), Matsumoto et al. (2004) directly investigated functional connectivity between frontal and temporal language regions by using both electrocortical stimulation and recording in patients undergoing awake surgery. They found that cortical electrical stimulation in Broca's area led to evoked potentials in Wernicke's area and adjacent basal temporal area and also vice versa, indicating functional bidirectional links between Broca's and Wernicke's areas.

One of the several fiber tracts linking frontal and temporal areas which appears to be particularly important is the arcuate fasciculus (AF), which links these areas along the dorsal stream, via the parietal lobe. There are several reasons why the AF is believed to be particularly important for building phonological APCs. Glasser and Rilling (2008) view the AF as a specifically phonological pathway because it establishes a direct connection between posterior temporal regions implicated in speech perception (BA 22) and those parts of the frontal lobe most strongly implicated in phonology (BA 6/44). Furthermore, AF strength has been found to correlate with phonological awareness skills (Yeatman et al., 2011; Saygin et al., 2013). But most importantly, in recent years comparative neuroimaging of humans and non-human primates, e.g. chimpanzees and macaque monkeys, has demonstrated that the AF is much more weakly developed or even absent in non-human primates (Rilling et al., 2008; Thiebaut de Schotten et al., 2012; Rilling, 2014) (see Figure 1.4). Thus, whereas the ventral connections between temporal and frontal lobes, including the extreme capsule, do not seem to have changed massively in primate evolution, the AF underwent massive evolutionary changes from macaques to humans, where it is estimated to contain ca. $10^7$ fibers (Schüz and Braitenberg, 2002). Crucially, the AF is strongly left-lateralized – the language-dominant

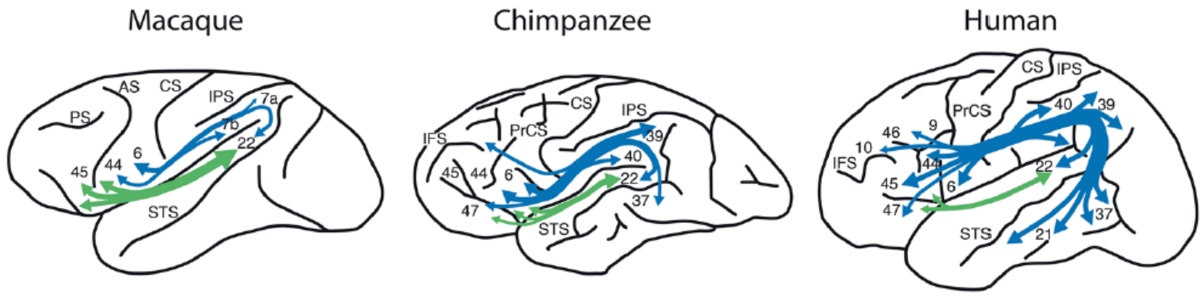


**Figure 1.4:** differences in connectivity strength between frontal and temporal areas along the arcuate fasciculus (AF) in the dorsal stream across macaques, chimpanzees and humans. AF dorsal stream connectivity is shown in blue, connectivity along the ventral stream in green. Figure adapted from Friederici (2016), published under a Creative Commons license (CC-BY); original data from Rilling et al. (2008).

hemisphere in most people – available already shortly after birth (Perani et al., 2011; Dubois et al., 2014), although at birth it is less left-lateralized and less strongly developed than in adults (Skeide and Friederici, 2016).

In summary, a number of reasons suggest that the AF might be a key neurobiological prerequisite for specifically human language abilities, although it is of course important to note that this observation from comparative neuroanatomy merely constitutes correlational evidence. Non-human primates – even our closest ancestors, chimpanzees and macaque monkeys – do not have language, at least not in any way comparable to human language, and it therefore seems likely that this can explained at least in part by the massive differences in AF connectivity between these species. There are some reports in the literature about teaching chimpanzees the meaning of words, but this amounts to the description of a few single cases where a few hundred words could be taught with years of training (Savage-Rumbaugh et al., 1993). This stands in stark contrast to the observation that humans typically acquire the meaning of 40,000 words (Pinker, 1994; Brysbaert et al., 2016). Recent experiments also succeeded in demonstrating that when humans learn novel pseudowords, neurophysiological reflections associated with this learning can be traced extremely rapidly, after only 15-30 minutes of exposure (Shtyrov et al., 2010; Shtyrov, 2011). Kimppa et al. (2015) further found that these neurophysiological reflections occurred for novel words with native-language-like phonology only, but not with non-native phonology. However, a follow-up study (Kimppa et al., 2016) found that this disadvantage for non-native phonology was less pronounced in individuals who had more experience in general in learning novel non-native phonology. In summary, these studies suggest the existence of mechanisms which are particularly suitable to rapid learning of novel phonological word forms and are also modulated by individual language learning experience.

One recent study also directly investigated the role of the arcuate fasciculus for word learning in humans. López-Barroso et al. (2013) employed a word learning task, creating artificial words composed of three syllables. Subjects heard these words and were later tested on recognition, where they had to indicate whether they had previously heard the word or not. López-Barroso et al. correlated performance on this word learning task with both structural and functional connectivity measure of the AF. Structural properties of the AF, in particular the radial diffusivity of water molecules perpendicular to fiber direction, predicted learning success of individual subjects (the less radial diffusivity and the greater the integrity of the AF, the better the word learning outcome). In addition to structural connectivity, López-Barroso et al. also looked at functional connectivity between Broca's and Wernicke's area. Functional connectivity is a measure for how strongly correlated (over time) activity in different brain regions is (Friston, 1994). López-Barroso et al. found that, just like structural connectivity, the strength of functional connectivity also correlated with word learning performance (measured as the word recognition rate, i.e. the ability to distinguish previously heard from non-heard pseudowords).

In summary, there is already abundant evidence indicating that structural connections between frontal and temporal areas by way of the arcuate fasciculus provide an important neuroanatomical basis for human language abilities. Moreover, these areas also functionally connected during word learning. It is therefore highly likely that the innate predisposition of strong AF connectivity, only present in humans, is a crucial neuroanatomical prerequisite enabling the building of APCs across perisylvian language areas which act as neuronal representations of word forms and can serve as basis for linguistic symbols.

## Unresolved questions and the overall objective of this dissertation

From the perspective of action-perception integration theory of language, one would predict that strong fronto-temporal connectivity is a crucial part in establishing APCs for syllables and word forms (prediction 1), and that as a result of this, speech comprehension is subserved not only by temporal areas, but equally by speech production areas in frontal cortex (prediction 2). I will highlight some important open questions that still exist regarding the empirical evidence for these predictions.

### *The role of fronto-temporal connectivity in building action-perception circuits (APCs)*

Regarding prediction 1, the existing evidence is largely correlational; no conclusive evidence showing that AF connectivity is indeed a crucial (or the crucial) neuroanatomical prerequisite for the emergence of human language function, has been found. AF lesions in humans have been shown to lead to verbal working memory impairments (Benson et al., 1973; Damasio and Damasio, 1980) and structural features of the AF correlate with word learning abilities (López-Barroso et al., 2013). Obviously, such correlational evidence can only suggest that structural AF properties drive functional connectivity of the AF, which, in turn leads to better word learning performance. However, this correlational observation sets the stage for putative causal explanations: What are the precise mechanisms that explain this correlational link?

To address this question, the study presented in Chapter 2 employs a novel methodological approach, which is becoming increasingly popular and is referred to by some as "computational comparative neuroprimatology" (Arbib, 2016) or "comparative connectomics" (van den Heuvel et al., 2016). The key advantage of that approach is that different model architectures can be set up which are identical in every way but differ only in the crucial feature of interest (in contrast, when comparing humans to non-human primates, there are likely many other differences apart from the massive differences in AF connectivity). Hence, any resulting functional differences between these architectural 'twins' can then with certainty be attributed to that feature. Neuroanatomically structured and neurophysiologically plausible models of the perisylvian language cortex implementing Hebbian learning principles were used as a basis (cf. Garagnani et al., 2008). The two model architectures differed only in the presence or absence of higher-order ‘jumping’ links between non-adjacent perisylvian areas (representing connectivity provided by the AF). The emergence of auditory-articulatory APCs was simulated by inducing coherent patterns of motor and auditory activity in the networks leading to Hebbian correlation-based learning. After learning, the auditory components of previously learned stimuli were presented to the trained networks. In order to investigate the functional consequence of AF connectivity, structural properties and dynamics of cell assembly activation for the two network architectures was compared.

### *The functional relevance of APCs distributed across fronto-temporo-parietal for speech perception and understanding*

Regarding prediction 2, there is ongoing controversy between proponents of different models (Hickok, 2014; Arsenault and Buchsbaum, 2015; Glenberg, 2015; Hickok, 2015a, 2015b; Kemmerer, 2015b; Rizzolatti and Sinigaglia, 2015). As outlined in the introduction, action-perception theory differs in important ways from the classic 19th century language models as well as more recent models (Hickok and Poeppel, 2004, 2007) and makes clearly differing predictions. The action-perception theory can be labeled as action-perception-integration model, because it predicts that (i) there should be activity in frontal and motor areas during speech perception and that causal effects on speech processing can originate not only in temporal, but also in inferior frontal and motor areas. In contrast, the 19th century models as well as Hickok and Poeppel’s model predict that speech recognition is achieved solely by the temporal cortex. These models can therefore be described as “local fractionated circuit models” (see Chapter 3 for details) because they view speech perception and production as two separate (hence “fractionated”) processes.

One major open question therefore is the role of activation of speech motor areas during passive speech perception and comprehension. There is controversy both about whether activation of speech motor areas can be consistently observed during passive speech perception, and also about what this activation signifies. It is clear that mere activation of these areas does not in itself prove that the area is functionally involved in speech processing (cf. Mahon and Caramazza, 2008). However, if the speech motor system is in principle capable of mediating the perceptual analysis of phonetic features, then this activation of motor systems in speech perception would need to demonstrate some kind of specificity with respect to articulatory features. Indeed, Pulvermüller et al. (2006) was the first to demonstrate in an fMRI experiment that during speech perception of stop consonants differing in place of articulation (PoA) – either bilabial or alveolar – there was differential activation of lip and tongue motor areas of motor cortex. In other words, lip and tongue motor regions were differentially activated as a function of the place of articulation (PoA) of the perceived phoneme. Therefore, motor areas contained and processed phonological information about perceived speech sounds, proving that in principle they could play a functional role in speech perception. Since this initial demonstration, several other fMRI studies have also investigated similar questions, in particular using a novel analysis method called multivariate pattern analysis, which, in contrast to univariate fMRI, allows testing if local brain activity patterns contain information about phonetic and phonemic features which might differ in their neural patterns *within* a given motor cortex sector. A detailed review of these recent fMRI studies on this topic is one focus of Chapter 3.

However, even with more advanced analysis methods providing converging evidence for phonological information in frontal areas during speech processing, this is still not conclusive evidence as to what the functional role of this activity is. To investigate this, a neuropsychological research strategy is required, which investigates whether neurophysiological changes in the sensorimotor cortex impact on speech perception. Unfortunately, patient evidence is rather limited and inconclusive on this count. As alluded to in the introduction, there have been descriptions of Broca’s aphasics who were not only impaired in speech production, but also in speech perception tasks involving explicit discrimination or identification of phonemes (Basso et al., 1977; Blumstein et al., 1977, 1994; Miceli et al., 1980; Caplan et al., 1995; Utman et al., 2001; Moineau et al., 2005). On the other hand, there are some reports of patients with impaired speech production but still relatively impaired speech perception abilities (Rogalsky et al., 2011; Stasenko et al., 2015). Another issue with lesion studies is that critical lesions are often rather extensive, and it is therefore very hard to find a set of patients with lesions pertaining exactly to the area of interest. In particular, the articulatory motor cortex is rather small compared to Broca's area, and therefore it is unlikely that patients with selective lesions to articulatory motor cortex or even only a specific sub-part of motor cortex controlling only one articulator could be found. To my knowledge, no such confined lesion has ever been documented in the literature. Therefore, transcranial magnetic stimulation (TMS) has become a popular method to address the question of the causal influence of frontal areas on speech perception and comprehension. TMS allows specifically targeting different subregions of the articulatory motor cortex. The spatial resolution of single-pulse TMS is quite high, such that regions about 0.5cm apart can be differentially stimulated (Walsh and Cowey, 2000). Since lip and tongue regions of the articulatory motor cortex lie about 2 cm apart (Pulvermüller et al., 2006), they can be differentially targeted with TMS (D'Ausilio et al., 2009), which is a great improvement over lesion studies, as discussed above. A review of recent TMS studies investigating whether stimulation of motor/premotor areas and inferior frontal areas has a causal effect on speech processing is therefore the second focus of Chapter 3.

### Speech perception vs. speech comprehension/speech recognition

Although, as reviewed in Chapter 3¸ a growing number of TMS studies suggests that motor systems can play a causal role in speech perception, one important caveat to these findings is that in those previous TMS studies, phonemes were presented in the context of isolated meaningless syllables. Thus, the tasks did not entail the comprehension of the meaning of speech and, additionally, also used somewhat unnatural tasks involving the explicit categorization or discrimination of phonemes. There is also evidence that speech perception performance in aphasia patients can be dissociated from their ability to understand the meaning of single words (Miceli et al., 1980). Therefore, a deficit in a speech perception task does not imply a comprehension failure and most of the reviewed TMS studies in Chapter 3 do not address whether the influence of articulatory motor areas extends to normal speech comprehension and semantic understanding.

The distinction between the *perception* of phonemes and the *comprehension* of meaningful language also features prominently in the model of Hickok and Poeppel (2007, p. 394): "[T]he implicit goal of speech perception studies is to understand sublexical stages in the process of speech recognition (auditory comprehension). This is a perfectly reasonable goal, and the use of sublexical tasks would seem to be a logical choice for assessing these sublexical processes, except for the empirical observation that speech perception and speech recognition doubly dissociate. The result is that many studies of speech perception have only a tenuous connection to their implicit target of investigation, speech recognition". As Chapters 3 and 4 show, this is indeed correct, and the fact that artificial phonological tasks were used in most TMS studies arguing for a role of speech motor systems in speech processing has significantly contributed to the widespread skepticism regarding this evidence. For example, McGettigan et al. (2010) write: "We have previously argued that […] unambiguous evidence supporting an obligatory role for motor representations in the perception of normal speech remains hard to find. This debate is unlikely to move forward when active tasks like phoneme monitoring are commonly used to measure basic speech perception". Many authors critical of motor involvements of speech perception have reiterated this argument, in virtually identical form, in many other critical publications (Lotto et al., 2009; Scott et al., 2009; Rogalsky et al., 2011; Hickok, 2014).

Chapter 4 therefore provides an experiment which investigates, using TMS, the role of the articulatory motor cortex for comprehension of single, meaningful, spoken words, i.e. where phonemes serve their normal function as meaning-discriminating units. To this end, instead of artificial phoneme discrimination tasks, the standard psycholinguistic test of spoken language comprehension, the word-to-picture-matching task, was used. Naturally spoken words, which were minimal pairs differing only in their initial phoneme were used. All critical phonemes were either bilabial (lip-related) or alveolar (tongue-related) stop consonants. Immediately before onset of the spoken word, TMS was applied to left articulatory motor cortex, either to the lip or tongue representation. For word-to-picture matching, two images appeared immediately after the word and subjects had to indicate which of them corresponded to the meaning of the word. One of the images depicted was always the spoken target word while the distractor image differed in the initial phoneme only (e.g. “Panne” vs. “Tanne”). Therefore, in this task, the initial phonemes served their normal role as meaning-discriminating units. Recognizing the word-initial sounds as either a bilabial or alveolar stop consonant was not the task per se but rather was implicit to understanding the whole spoken word and mapping it onto its meaning, as it is typical for natural language processing. Therefore, this task overcomes the problem voiced by Hickok and Poeppel (2007) that “many studies of speech perception have only a tenuous connection to their implicit target of investigation, speech recognition”.

# 2. Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex

---

This chapter is based on:

Schomers, M. R., Garagnani, M., Pulvermüller, F. (2017). Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. ***Journal of Neuroscience*** 37(11):3045**,** https://doi.org/10.1523/jneurosci.2693-16.2017

## Abstract

The human brain sets itself apart from that of its primate relatives by specific neuroanatomical features, especially the strong linkage of left perisylvian language areas (frontal and temporal cortex) by way of the arcuate fasciculus (AF). AF connectivity has been shown to correlate with verbal working memory—a specifically human trait providing the foundation for language abilities— but a mechanistic explanation of any related causal link between anatomical structure and cognitive function is still missing. Here, we provide a possible explanation and link, by using neurocomputational simulations in neuroanatomically structured models of the perisylvian language cortex. We compare networks mimicking key features of cortical connectivity in monkeys and humans, specifically the presence of relatively stronger higher-order "jumping links" between nonadjacent perisylvian cortical areas in the latter, and demonstrate that the emergence of working memory for syllables and word forms is a functional consequence of this structural evolutionary change. We also show that a mere increase of learning time is not sufficient, but that this specific structural feature, which entails higher connectivity degree of relevant areas and shorter sensorimotor path length, is crucial. These results offer a better understanding of specifically human anatomical features underlying the language faculty and their evolutionary selection advantage.

## Introduction

One of the key questions about human nature addresses the brain mechanisms underlying the language faculty. In sharp contrast to their closest relatives, humans learn novel words effortlessly and extremely rapidly (Shtyrov et al., 2010; Kimppa et al., 2015), and build vocabularies of tens of thousands of words (Pinker, 1994; Brysbaert et al., 2016), which can be stored in verbal working memory (VWM), an apparently specific human trait. We here ask which neural mechanisms and features of brain-structural connectivity might enable these uniquely human abilities.

Comparative neuroanatomical investigations in humans vs. other primates using diffusion tensor imaging (DTI) and diffusion weighted imaging (DWI) along with invasive tracer studies in nonhuman primates have greatly advanced the search for the specific structural features of the human brain. Lesion evidence shows that inferior frontal (including Brodmann areas (BAs) 44/45) and superior-temporal areas (BAs 42/22) of the left perisylvian cortex are most crucial for language, as lesions therein lead to aphasias involving both language production and comprehension (Bates et al., 2003). These core language areas are connected by a dorsal fiber bundle, the arcuate fasciculus (AF) (Schüz and Braitenberg, 2002), providing a bidirectional link (Matsumoto et al., 2004). Whereas the ventral connections between these areas do not seem to have changed massively in primate evolution, this dorsal bundle via the AF is rich and strong in humans (Rilling et al., 2008), available already shortly after birth and strongly lateralized to the left hemisphere (Dubois et al., 2009, 2014) – the language-dominant hemisphere in most people. Invasive tracing studies of macaque brains revealed a similar dorsal link between temporal and prefrontal areas (Petrides and Pandya, 2009), but parallel DTI/DWI and tractography in humans and macaques indicate relatively richer direct connections between inferior prefrontal and temporal parabelt areas in humans. In addition to this quantitative statement, specific qualitative difference appear to be present within the AF, where some area-specific long-distance connections seem to have strengthened massively or may even have newly emerged in the evolution from macaque and chimpanzee to human. Whereas comparative neuroanatomical DTI studies show connections between prefrontal cortex and temporal areas in the auditory parabelt in both humans and monkeys (Thiebaut de Schotten et al., 2012; their Fig. 3), the additional links between prefrontal and auditory belt and between premotor and auditory parabelt areas are well documented with DTI/DWI in humans but not so in macaques or chimpanzees (Fig. 2.1A; Rilling et al., 2012; Thiebaut de Schotten et al., 2012); as these connections introduce shortcuts to what can be described as a 5-step next-neighbor architecture (Fig 2.1C/D), we call them “jumping links.” Although not implying a complete absence of jumping links in nonhuman primates (Romanski et al., 1999; Smiley et al., 2007; Scott et al., 2015), the DTI-documented evolutionary change in dorsal connectivity leads to a shorter path length (defined as minimal number of synaptic steps) of strong links between auditory and articulatory motor areas. The AF appears crucially important for language, not only because of this evolutionary change, but also because its strength correlates with numerous human language abilities (Yeatman et al., 2011; López-Barroso et al., 2013; Saygin et al., 2013). However, a neuromechanistic explanation for why, among other factors, the quantitative topological differences in connectivity may be vital for the emergence of human-like language is still missing.

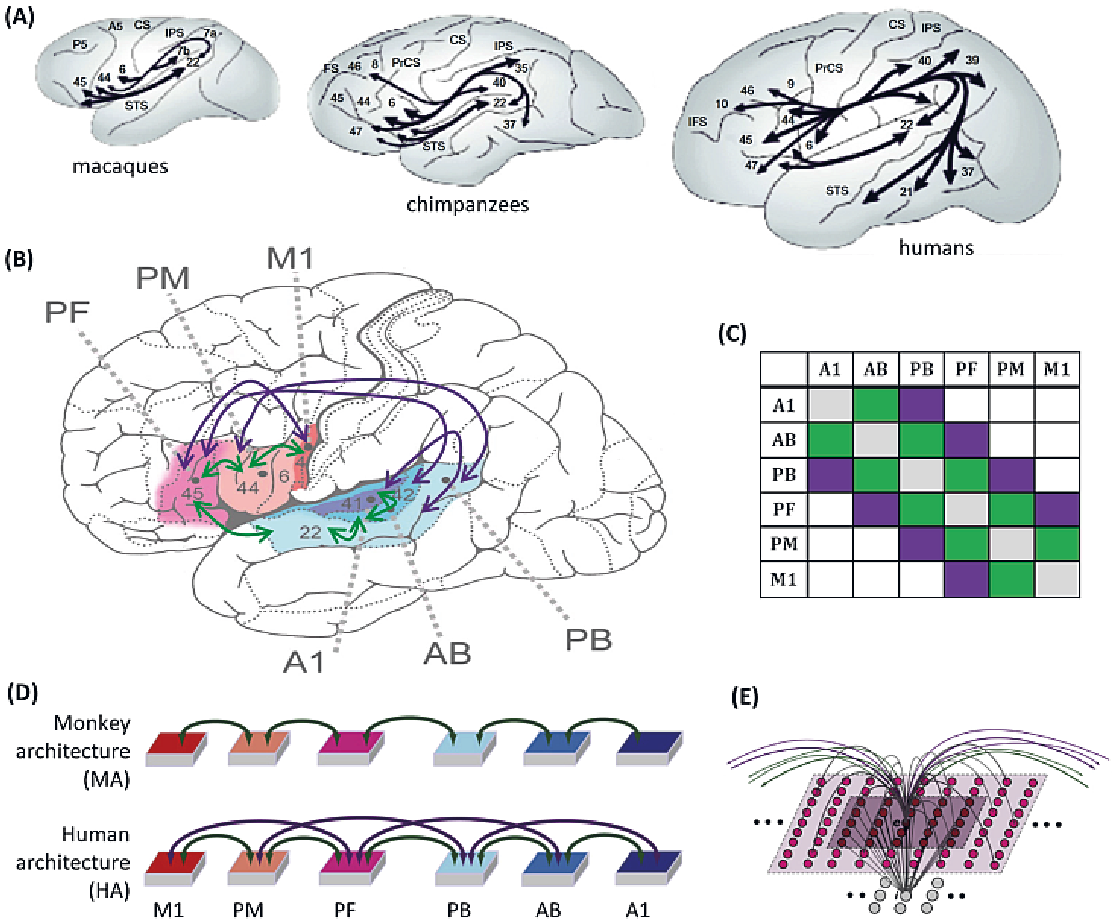
We here address this question using a novel approach of neurocomputational modeling, which has key advantages over both comparative studies and correlational evidence linking AF strength to language abilities. In those studies, a range of alternative features also distinguishing between monkey and human brains (including cortical area size and fiber diameters) could partly explain the observed performance differences. In contrast, models can be specifically designed to differ only in their connectivity structure, so that any functional change between them allows for definitive causal conclusions. We asked whether word learning or VWM abilities of humans could be causally linked to the presence of relatively stronger jumping links in human perisylvian cortex, as suggested by DTI/DWI data.

## Methods

### ***Network structure and function***

We used a neurocomputational model of the perisylvian language cortex. These networks were composed of graded response cells thought to represent average activity of a local pool of neurons (Eggert and Van Hemmen, 2000). Networks were subdivided into model areas of 25 × 25 = 625 excitatory and the same number of inhibitory neurons each (Figure 2.3A). One model area was established for each of the following perisylvian areas (Figure 2.1B; see Garagnani et al., 2008): superior-temporal primary auditory cortex (A1), auditory belt (AB), and parabelt areas (PB) and inferior frontal articulatory (M1), premotor (PM) and prefrontal cortex (PF). Adjacent areas in all models were connected, based on reciprocal links documented between the corresponding brain areas (green-colored arrows, Fig. 2.1B, e.g., A1 to AB, AB to PB etc. (Pandya and Yeterian, 1985; Braitenberg and Pulvermüller, 1992; Pandya, 1994; Young et al., 1994, 1995; Kaas and Hackett, 2000; Rauschecker and Tian, 2000).

As outlined in the Introduction, the rationale for this study was to investigate the functional consequence of qualitative and quantitative differences in connectivity between temporal and frontal regions along the dorsal stream. We therefore implemented two model architectures, a monkey architecture (MA) and human architecture (HA). In creating these architectures, we focused on major differences in the connectivity structure between monkey and human perisylvian regions that have been suggested by DTI/DWI-based tractography. This method currently offers the only prospect for comparative neuroanatomy of cortical long distance connectivity, as invasive tracer studies are not possible in humans. We did not aim at modeling the full complexity of the connectivity structures in each species, because even tractography data of exceptionally high quality are not as accurate as neuroanatomical tracing data (Thomas et al., 2014) and therefore may not allow one to uncover all functionally relevant links in a given species. However, DTI/DWI tractography studies converge on showing stronger left frontotemporal connections in humans compared with nonhuman primates and more specifically the unique presence of strong jumping links (see Introduction). Therefore, we focus on modeling these differences between species, rather than complete architectures. Whereas the MA included only next-neighbor connections between adjacent areas, the HA included additional 'jumping' links (see Figure 2.1B-D, purple). The strengths of all links were identical.



| | A1 | AB | PB | PF | PM | M1 |
|---|---|---|---|---|---|---|
| A1 | ▨ | 🟩 | 🟪 | | | |
| AB | 🟩 | ▨ | 🟩 | 🟪 | | |
| PB | 🟪 | 🟩 | ▨ | 🟩 | 🟪 | |
| PF | | 🟪 | 🟩 | ▨ | 🟩 | 🟪 |
| PM | | | 🟪 | 🟩 | ▨ | 🟩 |
| M1 | | | | 🟪 | 🟩 | ▨ |

**Figure 2.1**

**(A)** Illustration of perisylvian connectivity structure in macaques, chimpanzees and humans as revealed by tractography studies [Adapted by permission from Macmillan Publishers Ltd: Nature Neuroscience (Rilling et al., 2008), copyright 2008]. Note the strong frontotemporal connectivity of the latter, especially through the dorsal AF curving around the Sylvian fissure, and the presence of ventral connections in both.

**(B)** A human brain schematic is used to illustrate the area subdivision of the primate frontotemporal perisylvian cortex into frontal articulatory motor, premotor and prefrontal (M1, PM, PF) and temporal auditory, auditory belt and parabelt areas (A1, AB, PB) (Garagnani et al., 2008). Green arrows give the connections available in both the human and monkey architecture (HA, MA); purple arrows giveconnections unique to the human architecture. The purple arrows present only in the HA are meant to reflect the additional connectivity strength available only in humans, as shown by comparative DTI/DWI studies (see main text for detailed discussion).

**(C)** Connectivity matrix schematizing the connections according to next-neighbor (green) and indirect, jumping links (purple) skipping one intermediate area.

**(D)** Schematic depiction of the neural network architectures, equivalent to panel B.

**(E)** Microstructure of the connectivity of one single excitatory cell (labelled "e"). Local (lateral) inhibition is implemented by an underlying cell "i" (representing a cluster of inhibitory interneurons situated within the same cortical column), which receives excitatory input from all cells situated within a local (5 x 5) neighborhood (dark-colored area) and projects back to e, inhibiting it. Within-area sparse excitatory links (in grey) to and from e are limited to a (19 x 19) neighborhood (light-colored area); between-area excitatory projections (green and purple arcs) are topographic and target 19 x 19 neighborhoods in other areas (not depicted).

Panel B has been adapted from Garagnani and Pulvermüller (2013); panels D and E have been adapted from Cortex, 57, Pulvermüller, F. and Garagnani, M., "From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment", pp. 1–21, copyright 2014, with permission from Elsevier.

In addition to the between-area connectivity which differed between the network architectures, both architectures were designed so as to mimic a range of biologically realistic properties and therefore included the following features:

1. within-area connectivity, which was random, sparse (thus realizing only a small fraction of all possible connections), patchy and topographic (Gilbert and Wiesel, 1983; Amir et al., 1993; Braitenberg and Schüz, 1998), and such that local connection probability fell off with distance (Braitenberg and Schüz, 1998; Perin et al., 2011);
2. local and area-specific inhibition mechanisms (cf. Figure 2.3A, caption; Palm 1982; Bibbig et al. 1995; Wennekers et al. 2006) which act as a means to regulate and control activity in the network (Braitenberg, 1978; Palm, 1982; Garagnani et al., 2008);
3. synaptic modification by way of Hebb-type learning including both long-term potentiation (LTP) and long-term depression (LTD) (Artola et al., 1990; Artola and Singer, 1993);
4. constant presence of uniform uncorrelated white noise during all phases of learning and retrieval in all parts of the network (Rolls and Deco, 2010).

The implementation of the computational model follows that used in previous publications (Garagnani et al., 2008, 2009; Garagnani and Pulvermüller, 2013; Pulvermüller and Garagnani, 2014). Details about the underlying computations are also given in the section on *full model specification*.

### *Simulation procedures*

Simulations consisted of the following two phases: the learning phase and the testing phase. Twelve pairs of network instances were built, with each pair consisting of one MA and one HA network (i.e. 24 networks in total). In each instance, we first initialized an HA network, which entailed (1) randomizing all synaptic links (and corresponding weights) between cells in neighboring areas (and within areas) and (2) randomly generating 14 sensorimotor patterns ("words") to be used during training. Following this HA initialization, the network was copied, preserving the initial random links and the set of to-be-learned patterns, and the additional 'jumping' links (purple connections in Figure 2.1C) were removed, resulting in an initialized MA network. Both network architectures were then trained separately, but in exactly the same way (see below).

While each of the 12 different pairs of network instances had its own initial randomization of synaptic links and its own set of to-be-learned patterns, these features were identical to both pair members. Thus, there was some degree of 'between-subject' variability among the 12 network instances (because of randomly generated patterns and weight initializations for each pair), but there was parallelism with respect to these features between the two instances of each MA-HA pair. This ensured that the only difference between each MA-HA pair was in their long-distance connectivity – our variable of interest – while keeping all other factors identical. One may see this as the simulation of the same brain with human and monkey architecture, and thus as a 'within-subject' manipulation.

### *Training phase*

The 14 different acoustic-articulatory patterns were generated for each pair of network instances including 17 specific cells in A1 and another 17 in M1, equaling 2.72% of the neurons in each respective 25 × 25 area. These neurons were thought to represent abstract articulatory and acoustic phonological features (including articulatory and acoustic phonological distinctive features and coarticulatory information) about spoken word forms. Selection of neurons was random and (again) identical between HA-MA pairs. When producing a spoken word form, specific articulatory movements yield acoustic signals which, in turn, stimulate, with only minimal delay, the auditory system. To model this undeniable correlation of sensorimotor neuronal activity related to speech, which also receives support from recent electrocorticography studies (Cheung et al., 2016; Leonard et al., 2016), stimulus patterns were presented to sensory and motor areas A1 and M1 of each network. By "presenting a stimulus pattern", we mean that its 2 × 17 cells were activated together for 16 time steps. We wanted to avoid any possibly contaminating activity related to the previously presented stimulus pattern, and hence an inter-stimulus interval (ISI) followed each stimulus presentation. This ISI lasted for at least 30 time steps, until network activity had returned to the baseline value. During these ISIs the only input to the network was uniform white noise, simulating the spontaneous baseline neuronal firing observed in real neurons. Note that all parts of the network were subjected to the same amount of noise. The trial-to-trial presentation sequence of the different patterns was random. Hebbian learning was effective throughout learning trials, both during stimulus presentation and ISIs.

After stimulation to M1 and A1, activation spread throughout the model areas. As a consequence of activation spreading and resultant co-activation of neurons across the network, associative learning led to the formation of circuits interlinking the articulatory and auditory patterns, as documented in several previous studies (Garagnani et al., 2008; Garagnani and Pulvermüller, 2013, 2016; Pulvermüller and Garagnani, 2014; Tomasello et al., 2016). Due to sensorimotor activation, neural activity was present in specific neurons in A1 and M1, which partly activated further neural elements connected to these stimulated ones. Correlated activity and Hebbian learning mechanisms led to synaptic strengthening so that, eventually, sensorimotor stimulation led to increasingly stronger activation spreading to specific neuron sets throughout the network, which finally led to formation of a distributed LTM trace, or cell assembly (for detailed description and analyses of cell assembly formation in this type of network, Garagnani et al., 2008, 2009; Pulvermüller and Garagnani, 2014).

### *Testing phase*

The functionality of the circuits developing in the HA and MA was then compared in the testing phase, where all previously learnt auditory patterns were presented once again, in random order. Auditory stimulation (without articulatory pattern stimulation) was chosen to simulate speech perception. Stimulation was for two time steps; network responses were recorded during stimulation and the 30 subsequent time steps (i.e., 32 time steps in total).

### *Data analysis*

#### Structural network properties: cell assembly sizes

To assess whether articulatory-acoustic learning led to cell assembly formation, the presence and sizes of these circuits were assessed in each network instance. To identify the neurons forming cell assemblies across the different network areas, the activity of all 3750 excitatory network cells was monitored in response to one specific stimulation pattern. For each area, we calculated the maximum firing rate occurring across all 625 excitatory cells of a given area at any time during the 30 time steps following sensory stimulation. A cell was considered a member of a given cell assembly if and only if at any time step its firing rate reached at least 50% of the firing rate of the maximally responsive cell in the given area at that time step (provided that such maximum firing rate was at least 0.2). These procedures and thresholds were chosen on the basis of simulation results obtained with the present and previous networks (Garagnani et al., 2008, 2009).

## Dynamics of network activation

We also analyzed neural dynamics within each area separately in response to learnt patterns. To quantify differences in activation dynamics, we first calculated, for each area, the time point at which the firing rate was highest ($T_{max}$). This value was then used to quantify the area-specific duration of sustained activity (which we interpret as a measure of verbal working memory; Fuster and Bressler, 2012), defined as the length of the interval (in simulation time-steps) during which activity in an area remained significantly above the pre-stimulation average (≥ 2 SDs of the average firing rate in the 10 time steps immediately before stimulation). We refer to this quantity as the (area-specific) “sustained memory period” (SMP). For both $T_{max}$ and SMP data, we conducted repeated-measures ANOVAs with factors model architecture (MA/HA) and area (six areas), both as within-subjects factor (see section Simulation procedures).

## *Full model specification*[3]

Each model area consists of two layers of 625 excitatory and 625 inhibitory cells (see Fig. 2.1D). Each excitatory cell represents a cluster of cortical neurons (pyramidal cells), and the underlying inhibitory cell models the cluster of inhibitory interneurons situated within the same cortical column (Wilson and Cowan, 1972; Eggert and Van Hemmen, 2000). The state of each cell x is uniquely defined by its membrane potential V(x,t), representing the average of the sum of all (excitatory and inhibitory) postsynaptic potentials acting upon neural pool (cluster) x at time t, and governed by the following equation:

$$\tau \cdot \frac{dV(x,t)}{dt} = -V(x,t) + k_1(V_{In}(x,t) + k_2\eta(x,t)) \qquad \text{(A1)}$$

where $V_{In}(x,t)$ is the net input to cell x at time t (sum of all inhibitory and excitatory

---

[3] this section has been adapted from Garagnani and Pulvermüller (2013)

postsynaptic potentials – I/EPSPs; inhibitory synapses are given a negative sign), $\tau$ is the membrane's time constant, $k_1$, $k_2$ are scaling constants and $\eta(x,t)$ is a white noise process with uniform distribution over [-0.5,0.5]. Time is in arbitrary units. Cells produce a graded response that represents the average firing rate of the neuronal cluster; in particular, the output (transformation function) of an excitatory cell x at time t is:

$$O \begin{cases} 0 & \text{if } V(x,t) \leq \varphi \\ (V(x,t) - \varphi) & \text{if } 0 < (V(x,t) - \varphi) \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{A2}$$

O(x,t) represents the average (graded) firing rate (number of action potentials per time unit) of cluster x at time t; it is a piecewise-linear sigmoid function of the cell's membrane potential V(x,t), clipped into the range [0, 1] and with slope 1 between the lower and upper thresholds $\varphi$ and $\varphi$ +1. The output O(x,t) of an inhibitory cell is 0 if $V(x,t)<0$, and V(x,t) otherwise. In excitatory cells, the value of the threshold $\varphi$ in Eq. (A2) varies in time, tracking the recent mean activity of the cell so as to implement a simple version of neuronal adaptation (Kandel et al., 2000) (higher activity leads to a higher threshold). More precisely,

$$\varphi\,(x, t) = \alpha \cdot \omega(x,t) \tag{A3}$$

where $\omega(x,t)$ is the time-average of the cell's recent output and $\alpha$ is the "adaptation strength".

For an excitatory cell x, the approximate time-average $\omega(x,t)$ of its output O(x,t) is estimated by integrating the linear differential equation Eq. (A4.1) below with time constant $\tau_A$, assuming initial average $\omega(x,0)=0$:

$$\tau_A \cdot \frac{d\omega(x,t)}{dt} = -\omega(x,t) + O(x,t) \tag{A4.1}$$

Local (lateral) inhibitory connections (see Fig. 2.1D) and area-specific inhibition are also implemented, realizing, respectively, local and global competition mechanisms (Duncan,

2006) and preventing activation from falling into non-physiological states (Braitenberg and Schüz, 1998). More formally, in Eq. (A1) the input $V_{In}(x,t)$ to each excitatory cell of the same area includes an area-specific ("global") inhibition term $k_S \cdot \omega_S(x,t)$, which is subtracted from the total sum of the I/EPSPs postsynaptic potentials $V_{In}$ in input to the cell, with $\omega_S(x,t)$ defined by:

$$\tau_S \cdot \frac{d\omega_s(x,t)}{dt} = -\omega_s(x,t) + \sum_{x \in area} O(x,t) \qquad \text{(A4.2)}$$

The low-pass dynamics of the cells (Eq. (A1), (A2), (A4.1-2)) are integrated using the Euler scheme with step size $\Delta t$, where $\Delta t = 0.5$ (in arbitrary time units).

Excitatory links within and between (possibly non-adjacent) model areas are random and limited to a local (topographic) neighborhood; weights are initialized at random, in the range [0, 0.1]. The probability of a synapse to be created between any two cells falls off with their distance (Braitenberg and Schüz, 1998) according to a Gaussian function clipped to 0 outside the chosen neighborhood (a square of size n=19 for excitatory and n=5 for inhibitory cell projections). This produces a sparse, patchy and topographic connectivity, as typically found in the mammalian cortex (Amir et al., 1993; Kaas, 1997; Braitenberg and Schüz, 1998; Douglas and Martin, 2004).

The Hebbian learning mechanism implemented simulates well-documented synaptic plasticity phenomena of long-term potentiation (LTP) and depression (LTD), believed to play a key role in experience-dependent plasticity, memory and learning (Rioult-Pedotti et al., 2000; Malenka and Bear, 2004). In particular, the learning rule is an implementation of the Artola-Bröcher-Singer model of LTP/LTD (Artola et al., 1990; Artola and Singer, 1993). In the model, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, $+\Delta w$ and $-\Delta w$ (with $\Delta w \ll 1$ and fixed). We defined as "active" any link from an excitatory cell x such that the output $O(x,t)$ of cell x at time t is larger than $\theta_{pre}$, where

$\theta_{pre} \in ]0,1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP (or LTD) to occur. Thus, given any two cells x and y connected by a synaptic link with weight $w_t(x,y)$, the new weight $w_{t+1}(x,y)$ is calculated as follows:

$$w_{t+1}( \begin{cases} w_t(x,y)+\Delta w & \textit{(LTP)} \quad \text{if } O(x,t) \geq \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y)-\Delta w & \textit{(LTD)} \quad \text{if } O(x,t) \geq \theta_{pre} \text{ and } \theta_- \leq V(y,t) < \theta_+ \\ w_t(x,y)-\Delta w & \textit{(LTD)} \quad \text{if } O(x,t) < \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y) & \textit{(no change)} \quad \text{otherwise} \end{cases} \tag{A5}$$

Parameter values used for the simulations:

| | |
|---|---|
| Eq. (A1) | Excitatory cells: $\tau = 2.5$ (in simulation time steps)<br>Inhibitory cells: $\tau = 5$ (in simulation time steps)<br>Scaling factor: $k_1$=0.01<br>Noise scaling factor (training phase): $k_2 = 15\sqrt{48}$<br>Noise scaling factor (testing phase): $k_2 = 5\sqrt{48}$<br>Global inhibition strength (training phase): $k_s$=95<br>Global inhibition strength (testing phase): $k_s$=60 |
| Eq. (A3) | Adaptation: $\alpha$ =0.026 |
| Eq. (A4.1) | Time constant for computing gliding average of cell activity:<br>$\tau_A = 15$ (in simulation time steps) |
| Eq. (A4.2) | $\tau_S = 8$ |
| Eq. (A5) | Postsynaptic potential thresholds for LTP: $\vartheta_+ = 0.15$<br>Postsynaptic potential thresholds for LTD: $\vartheta_- = 0.15$<br>Presynaptic output activity required for any synaptic change: $\vartheta_{pre} = 0.05$<br>Learning rate: $\Delta$w = 0.0007 |

## Results

### ***Cell assembly sizes***

Before analyzing the dynamics of network activation, we computed the resulting cell assembly sizes obtained after representing the sensory part of a previously learnt pattern to the model again (see Methods section for cell assembly definition). We computed CA sizes for 50, 100, 200, 500, 1,000, 1,500, 2,000, 6,000, and 10,000 learning trials. CA sizes were always larger for the HA than the MA (all $p < 0.001$), regardless of the number of learning trials (presentations per pattern). The resulting CA sizes are shown in Figure 2. During the first 1,000 presentations the number of CA cells grew at a very fast rate (relative ratios of CA sizes at 1,000 vs. 50 presentations were 1.82 for HA and 2.87 for MA). Growth rate fell off after 1,000 presentations for both types of architectures (relative ratios of CA size ratios at 2,000 vs. 1,000 presentations were 1.04 for HA networks and 1.06 for MA networks). These observations were supported by Tukey's HSD tests, which confirmed that CA sizes differed between 50 and 1000 presentations for both architectures (both $p < 0.001$), but did not significantly change between 1,000 and 2,000 presentations for either HA or MA networks (HM: p=0.17; MM: p=0.14). In addition, we approximated the derivative of the CA size changes at 500 and 1,000 learning trials and found that at 500 time steps, the size increase per additional learning trial (i.e., CA growth rate) was larger for MA (0.3 cells/100 learning trials) than HA (0.2 cells/100 learning trials). However, at 1000 time steps, this growth rate was 0.1 cells per 100 learning trials for both architectures. Hence, we assume that at 1,000 learning trials, both networks had become relatively saturated with respect to learning such that additional learning trials produced very small increase in CA sizes. We therefore focused our further analyses on networks trained to 1,000 learning trials.

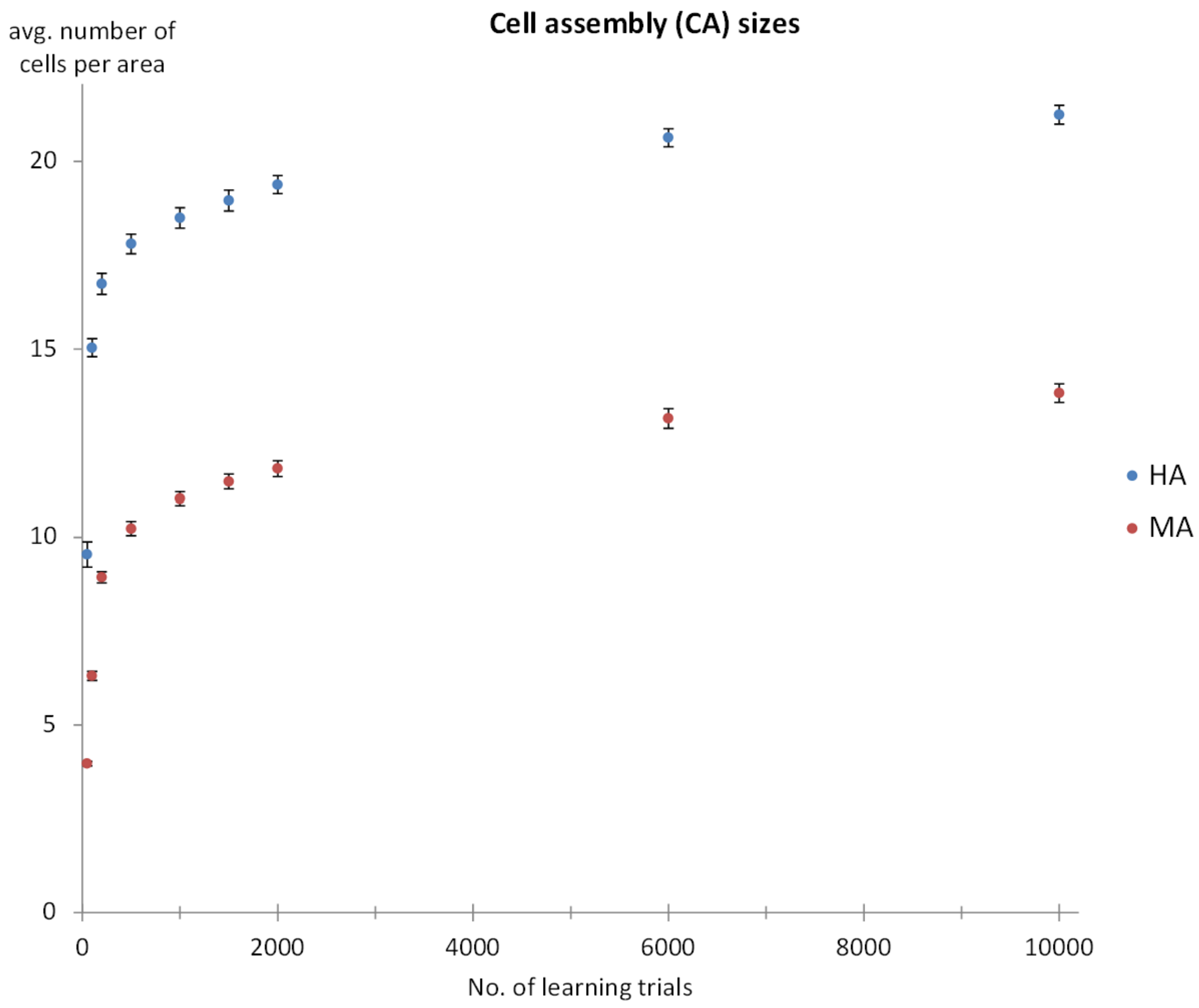


**Figure 2.2: Cell assembly sizes as a function of the number of learning trials**. Data are presented separately for the MA (in red) and the HA (in blue). Each data point represents the average of 12 network instances with 14 patterns per network (N=168). Error bars show SEM after removing between-network variance (Morey, 2008). Note the asymptotic behavior of both architectures with increasing number of learning trials.

### *Dynamics of network activation*

Figure 2.3 shows network dynamics (sum of firing rates as a function of simulation time step) induced by presentation of the sensory part of a previously learnt word pattern to area A1 in for the MA (top) and HA (bottom). Inspection of these plots reveals three qualitative difference in the dynamics of activation:

1. overall sum of firing rates are higher for the HA than for the MA (in part reflecting larger CA sizes; see Fig. 2.2)
2. activation is parallel for the HA, with areas AB/PB and PF/PM activating near-simultaneously; in contrast, in the MA, activation spreads in a serial manner throughout the six areas
3. activation persists for a much longer time in the HA than the MA.

Note that the modeling results of serial versus parallel activation seem to match recent experimental results. Whereas in the auditory system of macaques, “latencies [of auditory-evoked activity sometimes] increase with increasing hierarchical region” (Camalier et al., 2012), a feature that Camalier et al. see as partly “consistent with [serial] anatomical predictions,” recordings from humans have been found to be “not supportive of a strict serial model envisioning principal flow of information” along the A1 to parabelt pathway (Nourski et al., 2014) but were supportive of largely parallel auditory area activation instead. This contrast, although coming from methodologically very different studies and only reflecting some aspects of extremely rich datasets, seems consistent with the tendencies toward serial versus parallel processing implicated by our MA and HA models, respectively (Figs. 3, 4*A*). To investigate aspects 2 and 3 mentioned above (i.e., the seriality and persistence of activation quantitatively), we used the following measures (separately for each area and network type): the time step at which the maximum firing rate in a given area was reached, $T_{max}$ (Figure 4A), and the sustained memory period (SMP; see Methods) (Figure 2.4B).

We also conducted separate ANOVAs on these two measures. For $T_{max}$ data, the ANOVA revealed significant main effects of Type (F(1,11)=446, p<0.001) and Area (F(5,55)=1878, GGe=0.52, p<0.001, Greenhouse-Geisser-(GG)-corrected) and a significant interaction of Type x Area (F(5,55)=466, GGe=0.49, p<0.001, GG-corrected).

We conducted post-hoc Tukey's HSD tests comparing, separately for HA and MA, $T_{max}$ for adjacent areas. For the MA, all pairwise comparisons between adjacent areas were significant (p<0.001), confirming the seriality of activation of adjacent areas. In contrast, for the HA, comparisons were not significant between adjacent areas PB-PF (mean difference=0.19 time steps; p=0.98), between PF-PM (mean difference=0.18 time steps; p=0.99), between PM-M1 (mean difference=0.37; p=0.37), or even between the non-adjacent areas PB-PM (mean difference=0.37; p=0.37). All other comparisons between adjacent areas (A1-AB, AB-PB) for the human architecture were significant (all $p < 0.001$). This indicates that in the HA, initially activation spread serially from A1 via AB to PB, at which point the remaining areas of PF, PM and M1 activated near-simultaneously (Figure 2.4A).

For SMP data, the ANOVA revealed significant main effects of Type (F(1,11)=1388, p<0.001) and Area (F(5,55)=186, GGe=0.63, p<0.001, GG-corrected) and a significant interaction of Type x Area (F(5,55)=62, GGe=0.54, p<0.001, GG-corrected).

We conducted post-hoc-Tukey's HSD tests on the SMP data, comparing HA and MA at for each area separately. These comparisons showed that the SMP was significantly larger for the HA than the MA in all six areas (all $p < 0.001$).

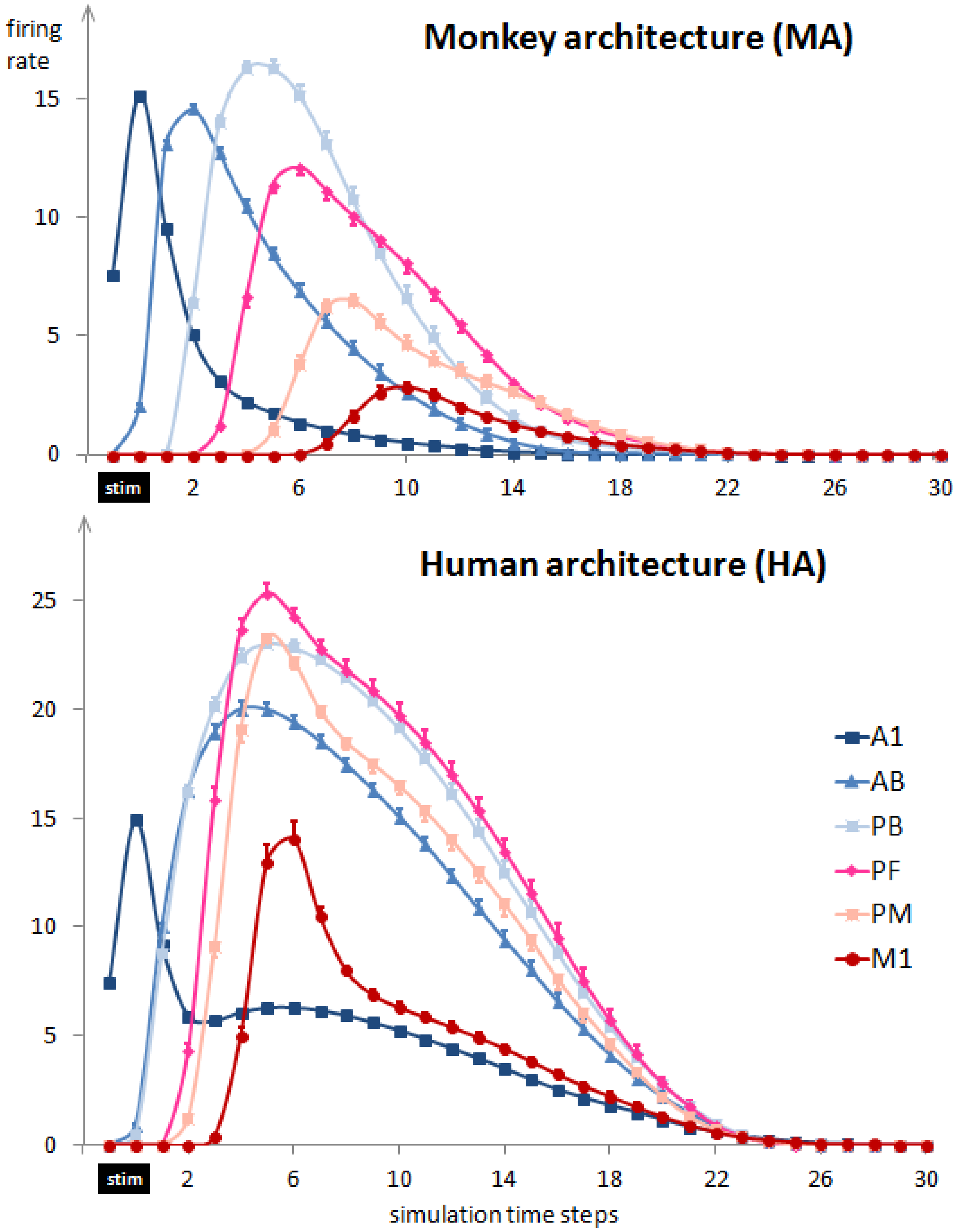


**Figure 2.3: Dynamics of network activation after sensory stimulation;** The figures show the sum of firing rates after presenting the sensory components of previously learnt patterns to A1. Stimulation was for the first two time steps (marked by a black bar), and following this, firing rates were recorded for 30 time steps. Each data point represents the average of 12 network instances with 14 patterns per network (N=168). Error bars show SEM after removing between-network variance (Morey, 2008).

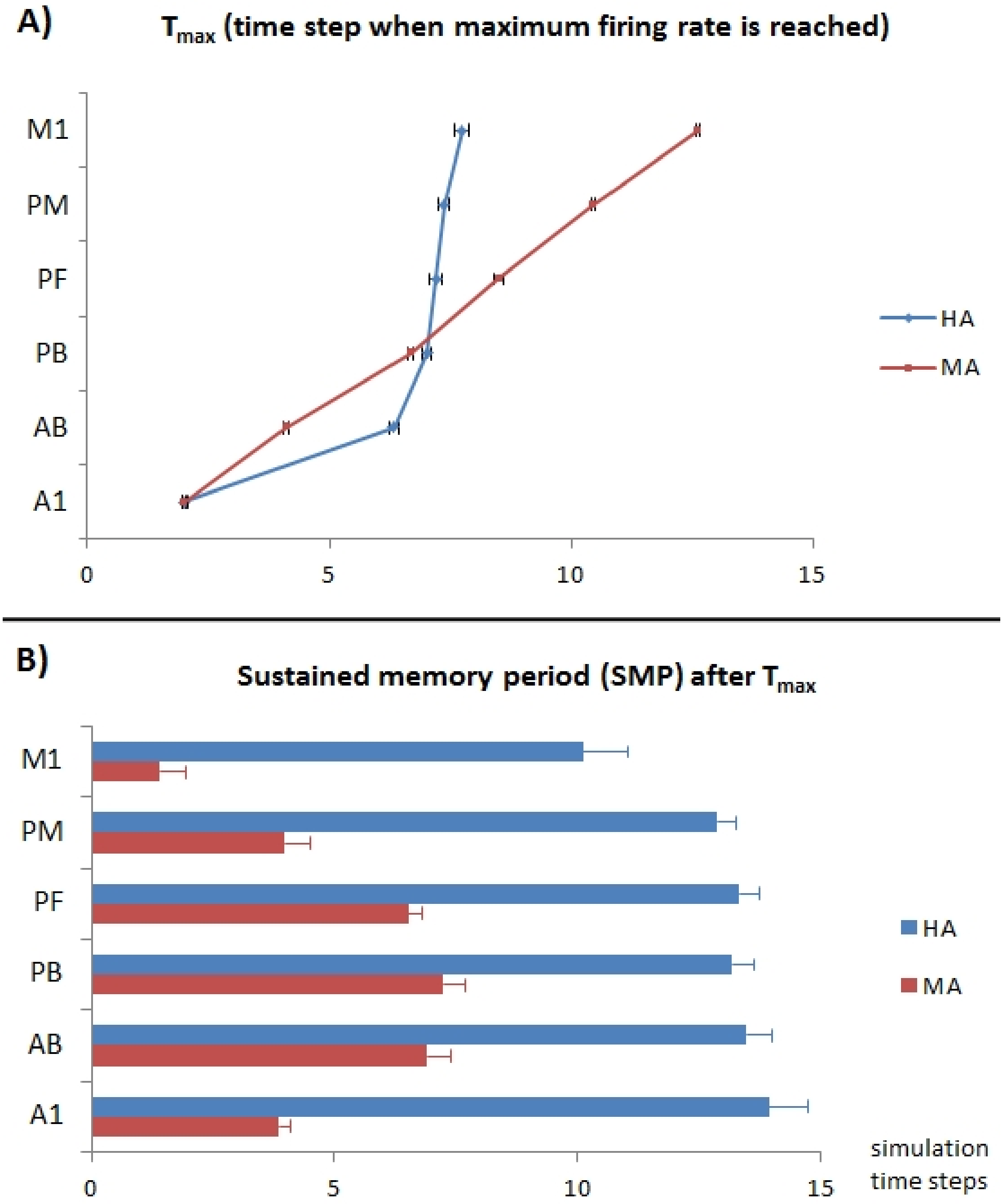


**Figure 2.4: Quantitative analyses of the dynamics of network activation in the MA (in red) and the HA (in blue)**

**(A)** Time step when the maximum firing rate is reached, $T_{max}$ (within the 30 post-stimulation time steps only). Note the serial activation of the MA and the near-simultaneous 'ignition' effect of all areas except A1/AB in the HA;

**(B)** Sustained memory period (SMP), defined as the duration (in time steps) starting from $T_{max}$ during which the firing rate remained at ≥ 2 SDs of the average firing rate of the pre-stimulation phase. Note the significantly larger SMP values for the HA architecture in all areas. Each data point represents the average of 12 network instances with 14 patterns per network (N=168). Error bars show SEM after removing between-network variance (Morey, 2008).

Both figures 2.4A and 2.4B are calculated based on the same raw data as depicted in Figure 2.3 (pre-stimulation baseline period not depicted)

### Repetition of analyses comparing HA 1,000 to MA 10,000

Although, as described in the Results section on cell assembly sizes, after 1,000 learning trials both models were at comparable learning stages, we wanted to rule out the possibility that the MA is simply slower in learning but still able to achieve qualitatively similar results. Therefore, we tested whether all statistical analyses obtained on comparing the HA and MA still obtained even when comparing the HA after 1,000 learning trials to the MA after 10,000 learning trials. The ANOVAs for $T_{max}$ and SMP data provided the same significance level for all effects. Thus, even when giving the weaker architecture the benefit of a ten-fold increase in learning trials, fundamental differences remained.

## Discussion

We used a neural-network model mimicking fronto-temporal perisylvian language areas, including primary sensorimotor, secondary and multimodal brain areas, to simulate word learning and examine the network responses to the sensory ("auditory") component of a previously learned pattern, which is akin to perceiving a familiar spoken word. Crucially, we compared the performances of two types of architectures, MA and HA, implementing differences in the connectivity of perisylvian areas suggested by DTI/DWI tractography in monkeys/apes and humans. Our results showed the following advantages of the HA over the MA:

1. Larger overall size of cell assemblies, or action-perception circuits (APCs) (Fig. 2.2), and thus stronger and more robust circuit activation (Fig. 2.3),
2. Parallel rather than serial activation reflecting cell assembly ignition (Fig. 2.3 and 2.4A),
3. Long-lasting activity in the network, reflecting cell assembly reverberation, and hence, emergence of verbal working memory (VWM) only in the HA (Fig. 2.3 and 2.4B).

Crucially, the disadvantages of the MA could not be compensated by longer training (up to 10,000 learning trials; see Results).

### *What are the linguistic implications of the observed functional changes?*

These results suggest that a change in neuroanatomical connectivity structure emerging in primate evolution underlies the build-up of a functionally robust lexicon of neuronal memory traces for multimodal articulatory-auditory patterns, which, in turn, can serve as carriers of meaning. Although the present simulations did not implement semantics, the emergent neuronal assemblies with long-lasting reverberating activity can be seen as a prerequisite for building a cortical lexicon of meaningful words. As complementary simulation studies show, such semantic learning is possible based on the same mechanisms of correlation mapping as those functional in the current model (Garagnani and Pulvermüller, 2016; Tomasello et al., 2016). In contrast to the large and fast-activating cell assemblies in the HA, the smaller and functionally sluggish circuits in the MA activated in a serial fashion, area by area, and, in spite of this prolonged activation process, there was little to no reverberatory activity (sustained memory period, SMP; see Figure 2.4B). In contrast, the HA yielded longer-persisting activity in its action-perception circuits (APCs), which we interpret as signifying verbal (phonological) working memory (Zipser et al., 1993; Fuster and Bressler, 2012).

The evolutionary change in neuroanatomical structure may provide a partial explanation partial explanation for why nonhuman primates have extremely weak auditory working memory, not only compared to humans but also compared to primates' working memory abilities in other sensory modalities (Fritz et al., 2005; Scott et al., 2012, 2014), and why, even after extensive training, nonhuman primates achieve vocabularies of only a fraction of those seen in humans (Savage-Rumbaugh et al., 1993; Call and Tomasello, 2007).

### *The functional relevance of the motor system for verbal working memory (VWM)*

It is widely agreed that a main function of the arcuate fasciculus is to map acoustic to articulatory representations. In our models, when presenting learned auditory patterns to A1, sustained activation in motor areas (M1 and PM) – those areas most distant from the sensory input – was observed only in the HA, and it occurred much earlier than in the MA (Figures 3 and 4B). This motor activity in our model can be viewed as reflecting (subvocal) articulation or rehearsal processes in a 'phonological loop' (Baddeley, 2003). Our results thus support the idea that VWM is not subserved by any dedicated module, but rather consists in reverberating activity between frontal and temporoparietal areas, in line with patient, neuroimaging and transcranial magnetic stimulation (TMS) evidence demonstrating the importance of speech perception and production areas in VWM (Belleville et al., 1992; Paulesu et al., 1993; Wilson, 2001; Nixon et al., 2004; Buchsbaum et al., 2005; Jacquemot and Scott, 2006; Romero et al., 2006; Buchsbaum and D'Esposito, 2008; Strand et al., 2008; Acheson et al., 2011; Liao et al., 2014). Hence, an obvious explanation of why AF strength is important for VWM is that AF enables quick and efficient sensory-to-motor coupling along the dorsal stream for retrieving word form representations when listening and thereby activates bidirectional auditory-to-motor and motor-to-auditory loops for activity maintenance in reverberating working memory circuits (Pulvermüller and Garagnani, 2014).

Individual differences in the degree of motor systems recruitment during speech perception could also contribute to differential working memory abilities. Correlations between verbal working memory performance and speech motor system activations in speech perception have been demonstrated, both in fMRI (Szenkovits et al., 2012) and using motor-evoked potentials (Murakami et al., 2015). Hence, one can speculate that higher verbal working memory abilities are driven by stronger motor systems recruitment, although the existing studies do not allow definite conclusions about the causality of this relationship.

We note that one prediction emerging from the present account is that the producibility of incoming auditory stimuli should influence working memory. Producibility of speech sounds influences the activation of motor areas (Wilson and Iacoboni, 2006) and auditory-to-motor connectivity (Londei et al., 2010) during perception. If this motor activation is also functionally relevant for verbal working memory, then producibility should similarly influence learning of novel word forms. Indeed, producibility has been shown to influence recognition accuracy in word learning (Schulze et al., 2012). Furthermore, neurophysiological memory traces for newly learned word forms differ depending on whether they exhibit native-like – and hence pronounceable – phonology (Kimppa et al., 2015) and also depending on whether they are actually articulated during learning (Pulvermüller et al., 2012).

### *The relation between working memory and language learning*

Just like large vocabularies, VWM is a unique feature of humans, and even across human individuals there seem to be intrinsic relationships between VWM and language learning abilities (Baddeley et al., 1988; Baddeley, 1993; Papagno and Vallar, 1995). Furthermore, speech production deficits can lead to reduced vocabulary size, presumably due to impairments in overt or covert repetition of novel pseudowords (Bishop et al., 1990). Finally, the perisylvian areas implicated in articulatory rehearsal have been shown to also be important for word recognition memory by fMRI (Wagner et al., 1998; Davachi et al., 2001; Clark and Wagner, 2003; Paulesu et al., 2009) and non-invasive brain stimulation experiments (Karabanov et al., 2015; Savill et al., 2015a).

In essence, current theory and data strongly support that word learning in humans requires and relies on VWM. Human (anterior and posterior) perisylvian cortex provides the substrate for VWM and the perisylvian dorsal connection by way of the AF plays a crucial role in word learning (López-Barroso et al., 2013), likely in concert with the extreme capsule (López-Barroso et al., 2011). This is not to say that perisylvian connectivity is the only relevant factor, as other structures, notably the hippocampus (Breitenstein et al., 2005; Sederberg et al., 2007) and the amygdala (Ripollés et al., 2014), play important complementary roles in word learning.

### *What are the critical variables and benefits of the evolutionary network-topological change?*

Although there is agreement that the human AF is important for language (Wernicke, 1874; Rilling et al., 2008) and experimental evidence demonstrates its importance for verbal working memory (Benson et al., 1973; Damasio and Damasio, 1980; Catani et al., 2007; Rauschecker et al., 2009; Buchsbaum et al., 2011), the precise reason and underlying cortical mechanisms for these structure-function relationships had long remained unclear. Carefully controlled comparison of neural architectures may help here, as these can be exactly parallelized so that any functional difference between architectural ‘twins’ can be uniquely attributed to the one and only manipulated structural feature. In our present case, this specific feature was the implementation of strong corticocortical ‘jumping link’ connections, which, as suggested by comparative neuroanatomical studies using DTI/DWI tractography (see Introduction), may constitute an important structural difference between human and nonhuman primates’ brains. These connections, provide ‘shortcuts’ in the auditory-articulatory pathway in left-perisylvian cortex, leading to shorter sensorimotor path length. In general, path length is an important feature of cortical neuroanatomy which can be used to characterize functionally relevant differences (Kaiser and Hilgetag, 2006; van den Heuvel and Sporns, 2013). Furthermore, as more connections were present in the HA, multiple parallel links became available for projecting acoustic and articulatory phonological information onto each other. Rapid activation flow between articulatory and auditory regions appears necessary for building ‘actively’ reverberating loops, providing the rehearsal mechanism in human verbal working memory, and it is precisely this active memory component that nonhuman primates lack (Scott and Mishkin, 2016). Hence, we propose that these two features taken together – the more numerous connections and their shorter path lengths – are the crucial variables determining the more robust ‘word representations’ and the emergence of verbal (phonological) working memory in humans. Moreover, short sensorimotor path length may offer a mechanism not only for verbal working memory, but also for the coupling of auditory and motor information related to speech (for a related computational model, see Westermann and Miranda, 2004). This coupling could also explain why auditory-articulatory interactions are pervasive in speech perception and comprehension (Schomers and Pulvermüller, 2016; Skipper et al., 2017).

## Conclusion

Our results suggest that the AF plays a critical role in word learning because its rich connectivity in humans allows for efficient binding of auditory and articulatory information about speech into persistently active neuronal circuits carrying VWM functions. As VWM is necessary for acquiring a vast repertoire of meaningful words, such strongly reverberating circuits may be essential for explaining human language. We believe that the present comparative-neurocomputational research approach may open new and exciting pathways for explanatory evolutionary neuroscience.

# 3. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review

---

This chapter is based on:

Schomers, M.R., Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. ***Frontiers in Human Neuroscience***, 10:435, https://doi.org/10.3389/fnhum.2016.00435

## Abstract

In the neuroscience of language, phonemes are frequently described as multimodal units whose neuronal representations are distributed across perisylvian cortical regions, including auditory and sensorimotor areas. A different position views phonemes primarily as acoustic entities with posterior temporal localization, which are functionally independent from frontoparietal articulatory programs. To address this current controversy, we here discuss experimental results from functional magnetic resonance imaging (fMRI) as well as transcranial magnetic stimulation (TMS) studies. On first glance, a mixed picture emerges, with earlier research documenting neurofunctional distinctions between phonemes in both temporal and frontoparietal sensorimotor systems, but some recent work seemingly failing to replicate the latter. Detailed analysis of methodological differences between studies reveals that the way experiments are set up explains whether sensorimotor cortex maps phonological information during speech perception or not. In particular, acoustic noise during the experiment and 'motor noise' caused by button press tasks work against the frontoparietal manifestation of phonemes. We highlight recent studies using sparse imaging and passive speech perception tasks along with multivariate pattern analysis (MVPA) and especially representational similarity analysis (RSA), which succeeded in separating acoustic-phonological from general-acoustic processes and in mapping specific phonological information on temporal and frontoparietal regions. The question about a causal role of sensorimotor cortex on speech perception and understanding is addressed by reviewing recent TMS studies. We conclude that frontoparietal cortices, including ventral motor and somatosensory areas, reflect phonological information during speech perception and exert a causal influence on language understanding.

## Introduction

Establishing links between the specifically human ability to speak and understand language and the underlying neuronal machinery of the human brain is key to modern cognitive neuroscience. At the level of specific language sounds, or *phonemes*, such links were first suggested by magnetoencephalography (MEG) recordings which showed that neuromagnetic activity differed between vowel types (Diesch et al., 1996). This work was followed by demonstrations of distinct and phoneme-specific local activity patterns in the superior temporal cortex, close to auditory perceptual areas (Obleser et al., 2003, 2006; Obleser and Eisner, 2009). However, phonemes are abstract multimodal units interlinking what is heard with how to produce the acoustic signals, and even visual representations of the articulatory movement play a role in processing speech sounds (McGurk and MacDonald, 1976; Schwartz et al., 2004). Therefore, their neuronal correlates may not be locally represented in the brain in and close to the auditory-perceptual temporal cortex alone, but, instead, may be supported by distributed neuronal circuits that interlink acoustic perceptual and articulatory motor information (Pulvermüller, 1999; Pulvermüller and Fadiga, 2010, 2016; Schwartz et al., 2012; Strijkers and Costa, 2016).

That phonemic perceptual mechanisms link up with articulatory information processing in the mind and brain had long been stated by biological and cognitive models of speech processing. In particular Fry's (1966) early model postulated sensorimotor articulatory-acoustic mechanisms and also the Motor Theory of Speech Perception (Liberman et al., 1967; Liberman and Whalen, 2000) linked phonemic production with perception, although other statements immanent to that theory – about the modularity of speech processing and the primacy of the speech motor module for perception – appear problematic today (Galantucci et al., 2006; Pulvermüller et al., 2006; Schwartz et al., 2012). Contrasting with the cross-modal links suggested by biological and motor theories, a classic position in the neuroscience of language had been that speech motor and speech perception networks are relatively independent from each other (Wernicke, 1874; Lichtheim, 1885), a position also inherited by more recent approaches. As one example, Hickok views the posterior superior temporal sulcus as the locus for phonemes and as "the real gateway to understanding" (Hickok, 2014, p. 181). Today, two diverging positions dominate discussions about the brain basis of phonemes (Figure 3.1). In one view, phonemic speech perception circuits are located in temporal and temporo-parietal cortex and are functionally separate from speech production circuits in inferior frontal and articulatory areas. We call this the 'local fractionated circuit model' of speech perception and production, because, in this view, the temporal speech perception network would realize speech recognition on its own (local fractionation) and speech production circuits in fronto-parietal cortex (or 'dorsal stream') are considered to play "little role in perceptual recognition" (Hickok, 2014, p. 239)[4]. Speech production and perception are thus viewed as independent processes, mapped onto separate brain substrates with no significant interaction between them, hence the term 'fractionated circuit model'. In contrast, the 'action perception integration model' postulates strong reciprocal links between speech perception and production mechanisms yielding multimodal distributed neuronal circuits, which provide the neuronal basis for the production, perception and discrimination of phonemes. These distributed multimodal circuits encompass acoustic perceptual mechanisms in temporal cortex along with articulatory sensorimotor information access in fronto-parietal areas.[5] Thus, in contrast to Liberman's pure motor theory, which viewed articulatory gestures,

---

[4] But note that some other publications of the same author (Hickok et al., 2011; Hickok, 2015b), while still proposing a fractionated model overall, do acknowledge that the speech motor system could exert a *modulatory* influence on speech perception by way of forward predictions, at least under certain specific task or contextual constraints.

[5] Such an action-perception-integration perspective is not restricted to the speech domain, but equally applies to written word processing, where it has been demonstrated that reading letters activates the hand motor areas involved in writing (Longcamp et al., 2003).

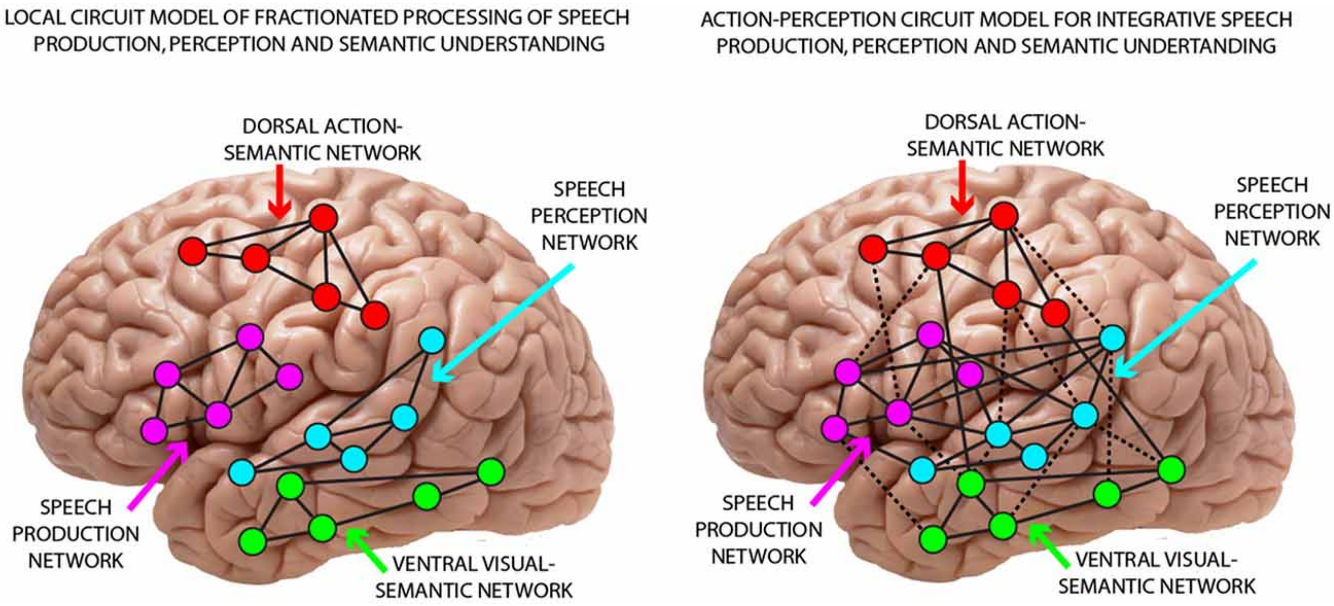


**Figure 3.1. Illustration of two competing theoretical positions regarding the role of temporal and frontal regions in speech perception.**
**(Left)** The local fractionated circuit model implies segregated processes for speech production (in frontal and sensorimotor cortex) and speech perception (in superior temporal cortex). Accordingly, sensorimotor fronto-central speech production networks are not involved in and in particular, do not functionally contribute to phoneme processing.
**(Right)** The action-perception-integration model postulates strong reciprocal links between superior-temporal speech perception and fronto-central production mechanisms yielding multimodal distributed neuronal circuits, which provide an interactive distributed neuronal basis for the production, perception and discrimination of phonemes.

i.e. motor units, as the central unit of speech perception, modern neurobiological theories of speech perception emphasize the interplay between perceptual and motor processes, positing that language processing relies on action-perception circuits distributed across auditory and motor systems (Pulvermüller and Fadiga, 2010, 2016).

From an integrative action perception perspective, the fronto-parietal sensorimotor system appears well suited for processing fine-grained differences between speech sounds, because the muscles and motor movements relevant for the articulation of speech sounds have different and well-investigated cortical loci side by side (Penfield and Rasmussen, 1950; Bouchard et al., 2013). Neighboring body parts are controlled by adjacent locations of the motor and premotor cortex (PMC) and a similar somatotopic relationship holds in the somatosensory cortex, where sensations in adjacent parts of the body are represented side-by-side. Different articulators such as the lips, jaw and tongue are localized from top to bottom in the so-called 'motor strip', thus predicting that a phoneme strongly involving the lips – such as the [+bilabial] phoneme /p/ – is cortically underpinned by relatively more dorsal neuronal assemblies than a tongue related phonological element – such as the [+alveolar] phoneme /t/. Apart from predominant articulator involvement per se (e.g., tongue vs. lips), different actions performed with the same articulator muscles may have their specific articulatory-phonological mappings in the motor system (Kakei et al., 1999; Graziano et al., 2002; Pulvermüller, 2005; Graziano, 2016), thus possibly resulting, for example, in differential cortical motor correlates of different tongue-dominant consonants (/s/ vs. /ʃ/) or vowels (features [+front] vs. [+back] of /i/ vs. /u/). Crucially, in the undeprived language learning individual, (a) phoneme articulation yields immediate perception, so that articulatory motor activity is immediately followed by auditory feedback activity in auditory cortex, and (b) the relevant motor and auditory areas are strongly connected by way of adjacent inferior frontal and superior temporal areas, so that (c) well-established Hebbian learning implies that auditory-motor neurons activated together during phoneme production will be bound together into one distributed neuronal ensemble (Pulvermüller & Fadiga, 2010).

In this *action-perception integration perspective*, speech sounds with different places of articulation have their cortical correlates in different activation topographies across superior-temporal and fronto-parietal areas, including articulatory sensorimotor cortex. If this statement is correct, it should be possible (i) to see motor activity during speech perception, phoneme recognition and language understanding[6], and (ii) phonemes with different places of

---

[6] Note that obviously some suppression mechanisms are necessary to prevent overt motor movements/articulation during perception. In concrete implementations of action perception integration models of language, the obvious differences between speech production and recognition

articulation and articulator involvement should differentially activate subsections of the articulatory motor system. Furthermore, distributed sensorimotor cortical circuits for phonemes imply (iii) that causal effects on speech perception and understanding can originate not only in auditory cortex and adjacent secondary and 'higher' multimodal areas, but also in frontoparietal areas in and close to sensorimotor ones. As the motor and the somatosensory cortex have parallel somatotopies and with every articulator movement (performed under undeprived conditions) there is specific stimulation of the corresponding somatosensory cortex as well, this position predicts not only specific motor cortex activation in speech perception, but, in addition, somatosensory cortex activation. Indeed, there is evidence for a role of somatosensory systems both in speech production (Tremblay et al., 2003; Bouchard et al., 2013) and perception (Möttönen et al., 2005; Skipper et al., 2007; Ito et al., 2009; Nasir and Ostry, 2009; Correia et al., 2015; Bartoli et al., 2016). The motor and somatosensory system may already be important for speech perception early in life, since pacifiers blocking specific articulator movements were shown to affect the discrimination of speech sounds even in the first year (Yeung and Werker, 2013; for review, see Guellaï et al., 2014).

To sum up, a major controversy between the competing models (Figure 3.1) surrounds the involvement of the sensorimotor cortex and adjacent areas in the fronto-parietal cortex (or 'dorsal stream') in speech perception and understanding. While both agree on a role of temporal areas in speech recognition, the 'fractionated' model states independence of speech perception from fronto-parietal circuits, whereas the integrative action-perception perspective predicts interaction, and hence, additional involvement of fronto-parietal including sensorimotor cortices in speech perception and understanding. In this review article, we will evaluate the empirical results that speak to this controversy in an attempt to settle the debate.

---

(overt motor movements vs. open auditory 'gates') are implemented in terms of area-specific cortical regulation processes (Garagnani et al., 2008; Pulvermüller et al., 2014a).

## Auditory/temporal and sensorimotor/fronto-parietal activation in speech perception

When speech sounds embedded in meaningless syllables are presented to the ears, functional magnetic resonance imaging (fMRI) reveals widespread activation in both temporal and frontal areas (for a meta-analysis, see Vigneau et al., 2006). Activation in the auditory cortex and surrounding areas of superior and middle temporal cortex is not surprising because most of the afferent ‘cables’ of the auditory pathway conveying sound information from the ears terminate in superior temporal primary auditory cortex (Brodmann Area (BA) 41), from where activation spreads to adjacent and connected areas. Some of this activity, especially in the left language-dominant hemisphere, but also to a degree in the other one, is specific to speech, as is evident from comparisons of speech-sound elicited activity with that to noise patterns matched to speech (Scott et al., 2000; Uppenkamp et al., 2006). Some discrepancy still exists between data showing that speech specific activity is primarily present in anterior superior temporal cortex (Scott and Johnsrude, 2003; Rauschecker and Scott, 2009) or, alternatively, in posterior superior and middle temporal cortex (Shtyrov et al., 2000, 2005; Uppenkamp et al., 2006). Therefore, a role of both anterior and posterior temporal areas in processing speech sounds needs to be acknowledged (DeWitt and Rauschecker, 2012).

However, in addition to temporal areas, the frontal and sensorimotor cortex seems to equally be activated in speech processing. Early fMRI studies could already demonstrate general activation in the left inferior frontal cortex during passive speech perception (Poldrack et al., 1999; Benson et al., 2001). In a seminal study, Fadiga et al. (2002) applied magnetic stimulation to the articulatory motor cortex and showed that motor-evoked potentials (MEPs) in the tongue muscle are specifically enhanced when subjects listen to speech containing phonemes that strongly involve the tongue – in particular the rolling /r/ of Italian – and are enhanced even more if these speech sounds are embedded into meaningful words (but see

Roy et al., 2008). As this evoked-potential enhancement is likely due to increased activity in tongue-related motor and premotor cortex, it has been interpreted as confirmation for motor system activation in speech perception. Further converging evidence came from studies using a range of methods, including fMRI and MEG/electroencephalography (EEG) with source localization (e.g., Watkins et al., 2003; Watkins and Paus, 2004; Wilson et al., 2004), and it could be demonstrated that activation spreads rapidly from superior temporal to inferior frontal areas (Pulvermüller et al., 2003, 2005; see Tomasello et al., 2016, for converging evidence from computational modeling). Sound-evoked activity in the motor or sensorimotor system is not specific to speech sounds as compared with other acoustic stimuli, because similar patterns of motor activation have also been seen for nonlinguistic sounds, in particular for the sounds of mouth-produced or manual actions (Hauk et al., 2006; Scott et al., 2006; Etzel et al., 2008). However, apart from showing motor involvement in speech perception, Fadiga et al's (2002) work and related studies suggested specificity of activation at a more fine-grained level. In particular, the tongue-related articulatory-phonological nature of the /r/ sound may have contributed to localization specificity.[7] As we discuss below, this was investigated in detail in further studies.

## Does sensorimotor cortex contain phonological information relevant for speech perception?

Some fMRI studies investigated whether, during speech perception, activity in frontoparietal and articulatory motor areas reflect phonological information, in particular about the phonemic features 'place of articulation' (Pulvermüller et al., 2006; Raizada and Poldrack, 2007) and 'voicing' (Myers et al., 2009). Pulvermüller et al. (2006) had subjects

---

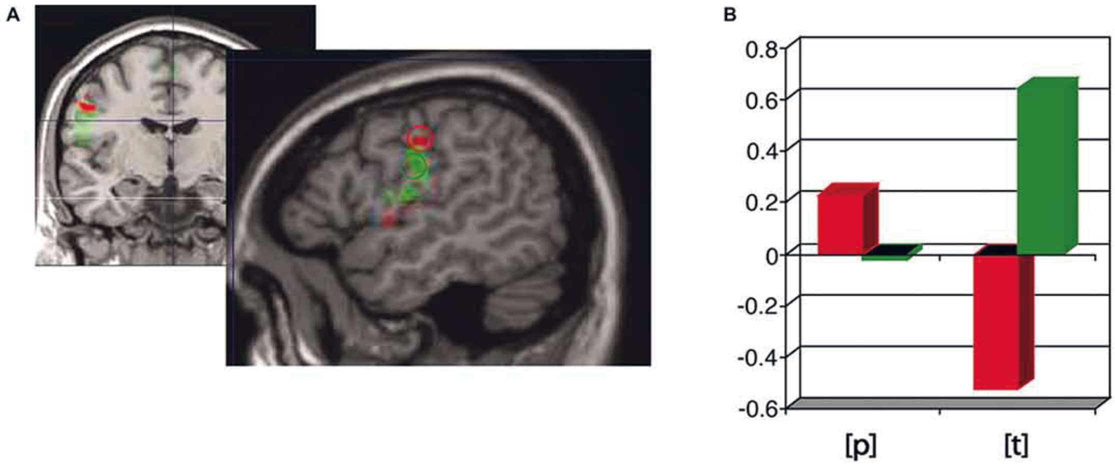
[7] Apart from phonological features, the high arousal and general motor activity associated with the rolling /r/ of Italian may be relevant for the observed specificity.

attentively listen to syllables starting with a lip-related bilabial /p/ or a tongue-related alveolar phoneme /t/. In the absence of any overt motor task, stimuli were passively presented during silent breaks where the MRI scanner was switched off, using a technique known as ‘sparse imaging’ (Hall et al., 1999; Peelle et al., 2010), so as to allow speech perception without scanner noise overlay. After the linguistic perception part of the experiment, participants produced non-linguistic minimal lip and tongue movements and these movement localizer tasks were used to define lip and tongue regions of interest (ROIs) in sensorimotor cortex. When using these ROIs, and also when examining a range of subsections of the precentral cortex, the authors found that during perception of syllables starting with lip-related and tongue-related sounds, the corresponding relatively more dorsal vs. ventral sectors of sensorimotor cortex controlling those articulators were differentially activated. In other words, the motor cortex activation as a whole contained information about the place of articulation of the perceived phonemes (see Figure 3.2 top).

In recent years, the univariate fMRI studies of the brain correlates of speech perception were complemented by experiments using the novel analysis method called multivariate pattern analysis, or MVPA (Haxby et al., 2001; Norman et al., 2006; Haynes, 2015). This method offers a way of testing whether fine-grained voxel-by-voxel activation patterns within specific brain areas contain information about stimulus types, for example about phonetic and phonemic features of speech. Initially, the application of MVPA to fMRI activity in studies on phonological processing focused on temporal cortex, where successful decoding of vowel identity could be demonstrated (Formisano et al., 2008). Recently, this approach has been extended to activity not only in temporal, but also in fronto-parietal areas (Arsenault and Buchsbaum, 2015; Correia et al., 2015; Evans and Davis, 2015). Arsenault and Buchsbaum (2015) found a reliable place of articulation classification throughout superior and middle temporal cortex and in the left subcentral gyrus, an area at the intersection of precentral and postcentral cortices also active during articulation (Huang et al., 2002; Bouchard et al., 2013;

Bouchard and Chang, 2014). However, these authors did not report reliable phonetic feature classification in the precentral motor cortex or inferior frontal cortex. Correia et al. (2015) trained classifiers on one phonetic feature using specific phonemes (e.g., place of articulation with stop consonants) and tested if performance generalized to the same feature exhibited by different phoneme types (e.g., fricatives). Such cross-phoneme-type generalization was successful in a large sensorimotor region, including precentral motor regions, IFG, and the postcentral somatosensory cortex.

An innovative study by Evans and Davis (2015) used MVPA of phonological processing and employed representational similarity analysis, or RSA (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013). This approach allows for testing models predicting the degree of similarity of neuronal patterns elicited by multiple pairs of stimuli. Using a 'searchlight' approach, one can then calculate the 'representational (dis)similarity' of neuronal patterns associated with different stimuli and see which of the several models of predicted similarity most closely resembles the actually observed similarities of neuronal activity patterns. Evans and Davis (2015) tested models which predicted similarity according to acoustic features (e.g., noise or speaker identity) or phonemic properties (e.g., phoneme identity and place of articulation). By using these advanced image analysis methods (MVPA and RSA) the authors were able to disentangle brain activity patterns related to acoustic vs. phonemic similarity, an important issue previously not addressed by most previous imaging studies. Their results revealed that local neuronal activations reflect a graded hierarchy: in primary auditory cortex, neural patterns code for the acoustic form of speech only, irrespective of phonemic features. In bilateral superior temporal cortex, both acoustic and phonemic information is coded. Finally, in left precentral gyrus, the highest degree of



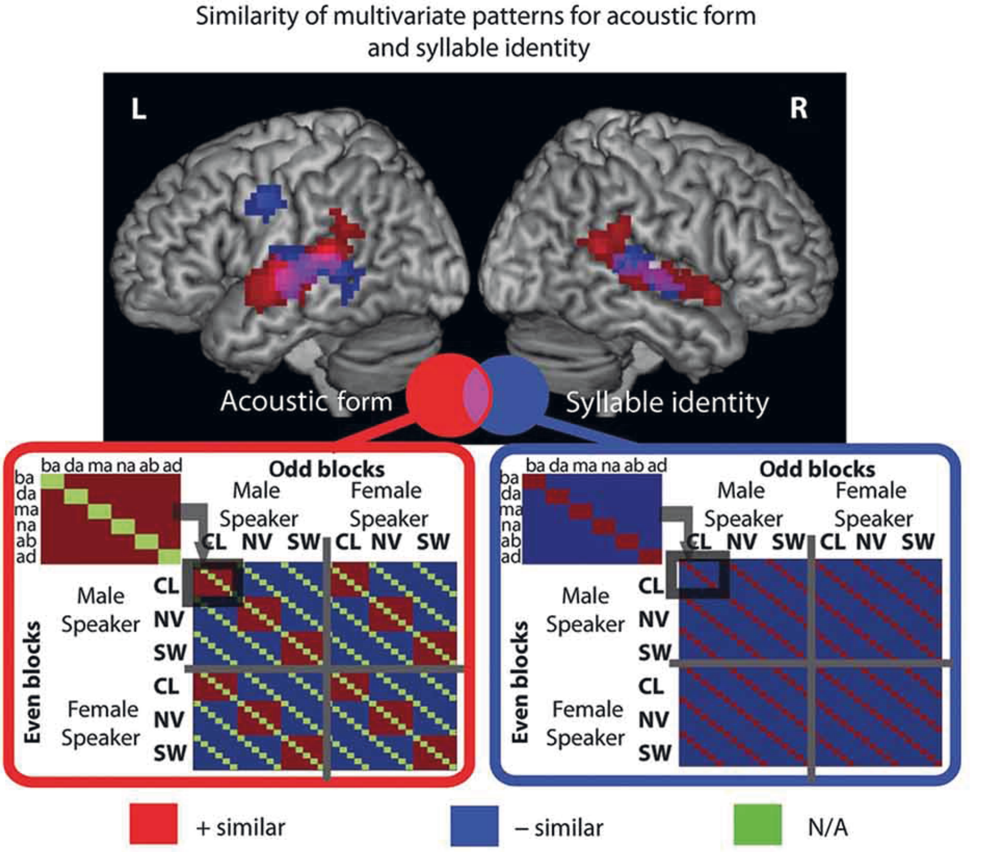


**Figure 3.2. Functional magnetic resonance imaging (fMRI) studies showing presence of phoneme-related information in motor systems during passive syllable perception.**

**(Top) (A)** Regions of interest (ROIs) were derived from non-linguistic minimal lip and tongue movements. Lip ROI shown in red, tongue ROI in green. **(B)** Differential activation (arbitrary units) in those same ROIs during passive perception of lip- and tongue-related phonemes /p/ and /t/, indicating an interaction between ROI and place of articulation (PoA) of the perceived phoneme. Adapted from Pulvermüller et al. (2006; Figure 2), Copyright (2006) National Academy of Sciences, Washington, DC, USA.

**(Bottom)** Representational similarity analysis (RSA) revealed that in pre- and postcentral motor regions the similarity of multivariate patterns reflects syllable identity, but not acoustic form; in contrast, in temporal regions, the similarity of patterns reflects both acoustic form and syllable identity. Patterns in precentral gyrus additionally reflect phoneme identity and CV structure (not shown in figure). Adapted from Evans and Davis (2015; Figure 3).

abstraction is found, with patterns reflecting phonemic aspects exclusively (phoneme and syllable identity and consonant-vowel structure; see Figure 3.2 bottom). In sum, the majority of studies report phoneme mapping across a fronto-parieto-temporal perisylvian region and some innovative recent findings from RSA indicate that the motor system of the human brain is unique in mapping phonemic properties of speech relatively independent of acoustic features.

## Some discrepancies between recent findings

In a recent study, Arsenault and Buchsbaum (2016) tried to replicate Pulvermüller et al.'s (2006) univariate results on double dissociations between activation for tongue- and lip-related (alveolar/bilabial) speech sounds and conducted additional multivariate analyses. They report a failure to find such a double dissociation in the precentral gyrus, both when defining the ROIs based on coordinates taken from the original study and when using their own motor localizer ROIs. As true replication failures may be of significance, a second look at both studies is warranted. Closer inspection in fact shows major differences between the original and the attempted replication. Whereas Pulvermüller et al. (2006) had chosen a localizer task with only minimal articulator movements (to foster focal brain activation), such a task was not included in the new attempt. Rather, Arsenault and Buchsbaum (2016) based their own ROIs on a task requiring silent, but overt articulation of lip- and tongue-related phonemes (/p/ and /t/), a strategy which had not led to significant results in the earlier work. Secondly, Pulvermüller et al. (2006) used 5 and 8 mm ROIs, whereas Arsenault and Buchsbaum (2016) used 10 mm ROIs. Thirdly, whereas Pulvermüller et al. (2006) refrained from using an overt motor task in the speech perception condition – to avoid general task-related activation in the motor system – Arsenault and Buchsbaum's (2016) subjects had to perform a button press task on some trials. And finally, Pulvermüller et al. (2006) had spent effort to reduce scanner noise by applying sparse imaging techniques, and additionally used matched noise stimuli as a baseline for the speech perception condition, so as to allow for good signal-to-noise ratios in the speech-evoked hemodynamic response. In contrast, Arsenault and Buchsbaum presented their sounds during scanning so that all phoneme stimuli were overlaid by scanner noise. Considering these substantial differences between studies, the more recent work appears as a replication failure in two senses, with regard to the results and methods of the pre-existing work. Below, we present an analysis of the recent literature to find out which of the methodological aspects of Arsenault and Buchsbaum's (2016) work might be responsible for the failure to replicate phoneme-related motor system activity (see the following section on the role of scanner noise).

Apart from their purported replication attempt using univariate methods, Arsenault and Buchsbaum (2016) analyzed their data using MVPA. They trained a classifier on a subset of the perception trials and tested it on a different subset. Instead of a searchlight approach, they tested the classifier in three anatomically defined ROIs, in the precentral and central sulcus as well as in the postcentral somatosensory cortex. Although classification was unsuccessful in the precentral and central ROIs, results revealed significant decoding of place of articulation in left postcentral somatosensory cortex, in line with the findings by Correia et al. (2015). We also note that explaining the presence of articulator-related information in somatosensory cortex requires the invocation of motor mechanisms because the motor movements of the different articulators are causal for any specific somatosensory sensations related to speech sounds – hence the need for including somatosensory cortex in integrative action perception models of language (see introduction of this chapter; Pulvermüller, 1992, 2013; Pulvermüller and Fadiga, 2010). In finding no MVPA mapping of phonological information in the motor system, Arsenault and Buchsbaum's (2016) results are in apparent contrast with the work by Evans and Davis (2015) and Correia et al. (2015) discussed above. We now turn to possible explanations of the observed discrepancies.

## The role of (scanner) noise and overt motor tasks

In order to explain the discrepancies in results about the motor system's role as an indicator of phoneme processing, it is necessary to pay special attention to subtle but possibly crucial differences between studies. In Table 3.1 (attached to the end of this chapter), we compiled a list of fMRI studies that found phonology-related information in specific cortical areas during (mostly passive) speech perception. The table lists studies that investigated the cortical loci of *general phoneme-related activity* during speech perception (studies 1-5) as well activity carrying *specific phonological information* (studies 6-15), for example, activation differences between phonemes, phonological features and/or feature values (such as [+bilabial] or [+front]). Comparing studies against each other shows that the crucial methodological factors which predict acoustically induced phonological activation of, and information in, fronto-parietal areas are (i) the use of 'silent gap', or 'sparse' imaging (Hall et al., 1999; Peelle et al., 2010) and (ii) the absence of a requirement to perform button presses during the experiments. Both of these features are amongst those that distinguished Arsenault and Buchsbaum (2016) from Pulvermüller et al. (2006).

### *The role of scanner noise*

Why would avoiding scanner noise be so important for finding brain activation related to speech perception in frontal areas? Arsenault and Buchsbaum (2016) argue that "according to previous literature, the background scanner noise […] should actually have *increased* the role of the PMC in speech perception" (p.1238; their italics). However, a closer look at the literature shows that the reverse likely applies; Table 3.1 shows that those studies which avoided scanner noise, button presses, or both (No. 1-3, 5-10, 13-14) all found activation (or MVPA decoding) in left motor cortex or IFG during speech perception; in contrast, those studies where both scanner noise and button presses were present (No. 4, 11, 12, 15; marked bold) found no involvement of left frontal or motor regions. The only exception to this rule is study 11 (Du et al., 2014), which reports precentral phonemic information in spite of noise and button presses on every trial. Crucially, however, and in contrast to Arsenault and Buchsbaum's (2016) statement, Du et al. (2014) found phoneme-related information in the ventral PMC (vPMC) only at the lowest noise level (headphone-attenuated scanner noise with no additional noise; Figure 3.3 D); at higher noise levels, successful phoneme classification could not be shown in vPMC anymore (Figure 3.3 A-C), but still in dorsal PMC (dPMC). They conclude that "adding noise weakened the power of phoneme discrimination in almost all of the above mentioned areas [see Figure 3.3 D] except the left dorsal M1/PMC which may index noise-irrelevant classification of button presses via the right four fingers" (Du et al., 2014; p. 7128). The caveat is likely given that there was a one-to-one-mapping between response buttons and phoneme category and this wasn't counterbalanced in Du et al's study. Decoding in inferior frontal areas (insula/Broca's region) was somewhat more robust to noise. However, in contrast to all other studies in Table 3.1, Du et al. (2014) used an active syllable identification task on every trial; it is therefore unclear whether decoding in these areas reflects phonological information or, alternatively, decision-related processes or response selection/preparation (see Binder et al., 2004). In contrast, of particular interest for articulatory information are precentral motor areas—those which were the focus of Arsenault and Buchsbaum's (2016) investigation; crucially, in these areas (as well as in superior temporal and inferior parietal regions), Du et al. (2014) found decoding to be most fragile, appearing only at the lowest noise levels.

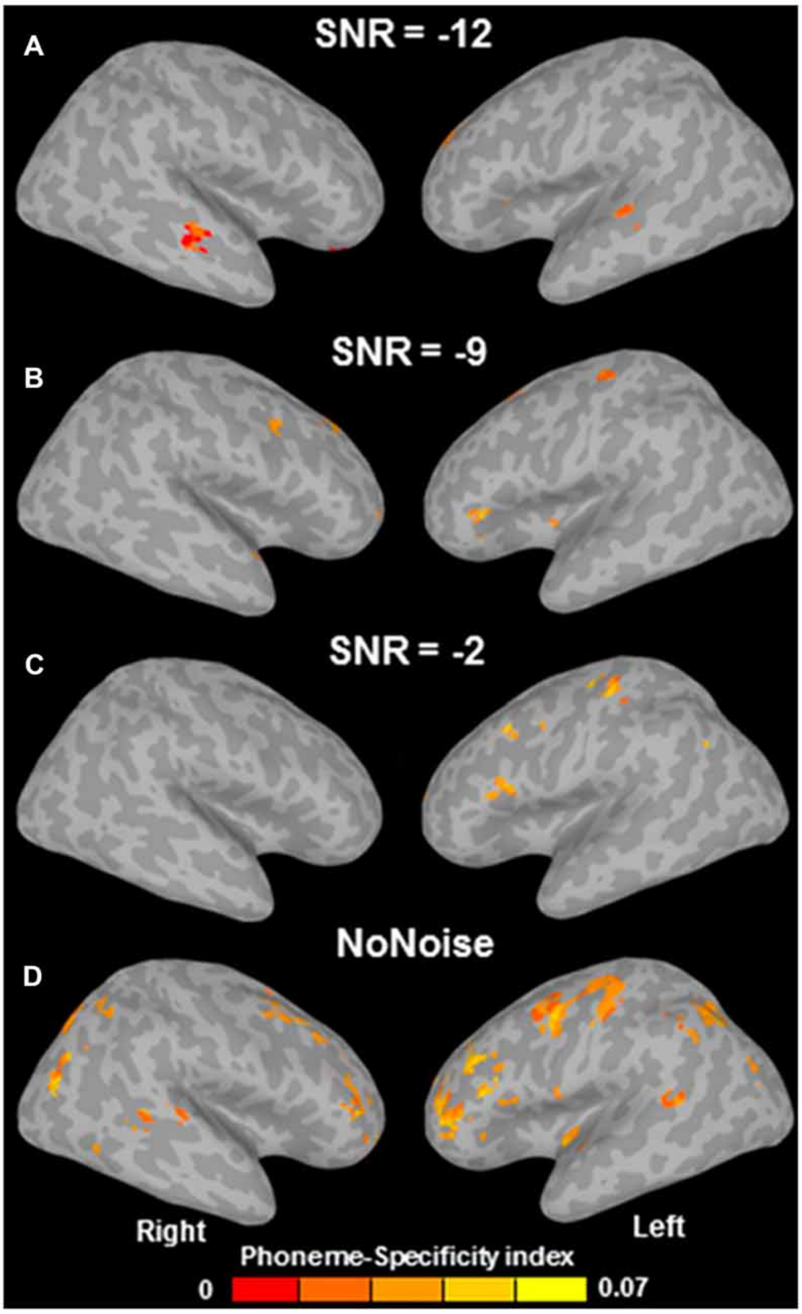


**Figure 3.3. Multivariate pattern analysis (MVPA) phoneme-specificity maps as a function of signal-to- noise ratio (SNR; in dB).** A more negative SNR indicates more additional noise on top of scanner noise attenuated by 25 dB (which was always present, even in the “no noise” condition). Successful MVPA decoding of phoneme identity in ventral premotor cortex (PMC) can only be seen in the “no noise” condition **(D)**, whereas with increasing noise **(A–C)**, decoding is unsuccessful in ventral PMC, but still successful in dorsal PMC and inferior frontal regions (see main text for detailed discussion). Adapted from Du et al. (2014; Figure S4).

Therefore, taking into consideration the caveats about Du et al.'s (2014) design, the following tentative conclusions can be offered: speech motor systems, but equally inferior parietal areas and superior temporal cortex – the latter being a site widely agreed to contribute to phonological processing – index phonological information processing only if the speech is presented without noise or with only moderate noise overlay.

Still, some studies reported that a contribution of frontal or motor systems *further increases* when stimuli become moderately more difficult to understand, for example because of noise overlay (Murakami et al., 2011; Osnes et al., 2011; Adank et al., 2012; Hervais-Adelman et al., 2012), motor-perturbed speech distortions (Nuttall et al., 2016), increased subjective dissimilarity between the perceived and the participant's own voice (Bartoli et al., 2015), or age-related hearing impairment (Du et al., 2016). However, that an increasing contribution of motor systems with increasingly challenging listening conditions does not logically entail that this contribution is generally absent in non-noisy listening conditions[8]. This is seemingly at odds with some TMS studies that found no evidence for motor involvement in speech perception without noise (D'Ausilio et al., 2012). However, null effects in the absence of noise are equally open to an explanation in terms of ceiling effects (see Sato et al., 2009, for a similar ceiling effect due to low task complexity). Note that normal speech is characterized by lots of redundancies due to co-articulation, requiring much information to be omitted before a measurable change in perception or comprehension performance can be found. Hence, if TMS to motor systems produces no effect in a task with high overall accuracy, this is likely a result of a ceiling effect or methodological factors (as, in general, TMS tends to produce weak effects) rather than indicating that motor systems' contributions are indeed absent—apart from the obvious fact that absence of evidence in favor of an effect is no evidence of its absence. Taking a broader perspective, there is abundant

---

[8] Furthermore, it is well-known that completely noise-free and 'perfect' listening conditions rarely occur outside the context of laboratory experiments (D'Ausilio et al., 2012)

evidence for motor systems activation during language processing in sparse-imaging fMRI experiments (see Table 1), as well as in other noise-free (and passive) tasks (Fadiga et al., 2002; Möttönen et al., 2013, 2014; Shtyrov et al., 2014; Grisoni et al., 2016).

In summary, motor systems' contributions to speech processing tend to show up already with no noise and might further increase with moderate noise overlay. However, with too strong noise overlay (which non-attenuated scanner noise might constitute), this contribution disappears again. This observation is problematic for models viewing perceptually-induced motor system activation as correlate of a prediction process *only* effective under noisy or otherwise challenging perceptual conditions (Hickok, 2015b).[9]

### *The role of overt motor tasks*

We now turn to the second important methodological point, the role of overt motor responses (e.g., occasional or constant button presses). Arsenault and Buchsbaum's (2016) study is subject to such a motor response confound. Subjects had to press a button occasionally on catch trials (11% of trials), to ensure they were paying attention. Therefore, subjects had to be prepared throughout the experiment to respond, thus leading to constant preparatory activity in the motor system. Such preparatory activity does not only involve the primary and pre- motor cortex, but, in addition, adjacent supplementary motor and prefrontal cortices as well, as is evident from studies investigating the so-called readiness potential and related preparatory brain indicators (Kornhuber and Deecke, 1965; Babiloni et al., 2001). Note that hand representations in somatosensory and motor areas lie side by side with articulator representations, especially of the lips. Presumably, preparatory neuronal activity in motor regions induced by a button press task causes a ceiling effect, which leads to a reduced chance of seeing small speech-sound induced articulator-related activity in motor cortex.

---

[9] Note that there are also similar proposals for motor-induced forward predictions in speech perception without constraints (Skipper et al., 2005, 2007), which sit well with the available data.

Indeed, previous studies using lexical decision tasks requiring a button press also often found no evidence of semantically related activation in motor cortex, whereas most studies using passive paradigms found such ‘semantic somatotopy’ (Carota et al., 2012; Kemmerer, 2015a). This discrepancy is best explained by preparatory hand-motor activity (for discussion, see Pulvermüller et al., 2001). A similar effect could be at work both in Arsenault and Buchsbaum (2016) and in their earlier study (2015) which even required a button press on every trial (plus a gender identification task). This pattern of results is consistent with the statement that preparatory motor activity and hence overt button press tasks work against the detection of phonological information processing in the articulatory motor cortex. This position would also offer an explanation why Arsenault and Buchsbaum (2015, 2016), despite finding no MVPA decoding in precentral cortex, nonetheless reported successful discrimination in the postcentral somatosensory regions, where any preparatory motor activity is minimal or absent, hence not leading to a ceiling effect as in precentral motor regions.

In sum, a review of a range of neuroimaging experiments on speech processing shows that the factors noise overlay and overt motor tasks explain why some previous univariate and multivariate fMRI studies found evidence for phoneme-specific activation in frontal cortex, including Broca’s and precentral areas, and why others did not.[10] The mechanisms underlying these effects need further clarification but a tentative mechanistic explanation can be offered in terms of acoustic phonemic signal-to-noise ratios reflected in the fronto-central cortex, which must decrease both with overlay of acoustic noise and ‘motor noise’ which may result

---

[10] A recent ECoG study (Cheung et al., 2016) reported superior-temporal along with pre- and post-central activation to single syllables presented in a button press task. Differences in brain responses were found reflecting the massive acoustic differences between stop and fricative sounds, but not for the fine acoustic differences between stop consonants with different place of articulation. As a button press task was used und data analysis focused on one specific neurophysiological measure, high frequency responses, these results do not motivate strong conclusions on motor recruitment in speech perception per se or the absence of phonemic discrimination in fronto-parietal cortex.

from preparatory motor movements. These two factors, especially in combination (see studies 4, 11, 12, and 15 in Table 3.1), seem to cause a loss of phoneme-related activation in frontal areas, which also explains the unsuccessful replication attempt of Arsenault and Buchsbaum (2016) and the discrepancies of their work with other recent studies (Evans & Davis, 2015; Correia et al., 2015)[11]. Therefore, a clear take home message from this review can be phrased as follows: in order to map the full cortical signature, including motor activity, of speech recognition and processing, it is advantageous to avoid (i) acoustic noise and (ii) overt motor responses. A further suggestion is to avoid tasks focusing attention on stimulus aspects which are not in focus (e.g., speaker identity when investigating phonological features), as this also has an impact on MVPA decoding (Bonte et al., 2014). An analogous suggestion may apply to other perception-related brain activity patterns as well.

### Excursus: Cross-decoding from miming to perception as the critical test?

A methodologically innovative aspect of Arsenault and Buchsbaum's (2016) study, compared to previous MVPA studies on this topic, was that they also used multivariate cross-classification, or cross-decoding (see Kaplan et al., 2015, for a review of the method). In this approach, a machine learning classifier is trained to distinguish a difference between types of stimuli in one condition or brain area and its performance is then tested on a different condition or brain area. Arsenault and Buchsbaum (2016) applied this logic to the difference between silent syllable articulation ('miming') and speech perception conditions. Classifiers were trained on the distinction between bilabial and alveolar place of articulation (PoA) on the miming data; they then investigated whether that same classifier could decode PoA from

---

[11] Our literature review does not rule out other factors as additional explanatory variables, such as ROI-based vs. searchlight analyses, or searchlight size; for example, Correia et al. (2015) found successful classifier performance in left IFG only with a searchlight radius of 20 mm, but not 10 mm (see also Lee et al., 2012, for discussion of the influence of searchlight size on MVPA analyses).

the fMRI patterns in the speech perception condition as well. Crucially, this cross-modality decoding from miming to perception did not succeed, which, according to Arsenault and Buchsbaum (2016) would be “the critical test of motor theories of speech perception”.

This latter statement is problematic, however; no explanation is given as to why this cross-decoding should constitute “the critical test”. This view seems to imply that substantial similarities should exist between the cortical activity patterns seen during speech production and perception. In contrast, the crucial prediction of action-perception integration models of speech which was vindicated by Pulvermüller et al. (2006) was that phoneme perception involves access to multimodal phoneme representations which, due to their multimodal character, *include* neurons with articulatory function in the speech motor system (cf. Galantucci et al., 2006). The key finding (see Figure 3.2 top) was that lip and tongue regions of motor cortex were differentially activated during speech perception, indicating that “information about articulatory features of speech sounds is accessed in speech perception” (Pulvermüller et al., 2006, p.7868). The link between perception and articulator movement conditions in Pulvermüller et al. (2006) consists in the fact that subregions of motor cortex (lip vs. tongue) were defined as ROIs based on the articulator movement localizer and in the perception condition, these same ROIs exhibited similar differential activity depending on the perceived phoneme. Thus, what the conditions had in common was that both of them produced articulator-specific activation of subregions of motor cortex. But this does not suggest that there should be more general and wide-ranging similarities in neural activation patterns between these conditions. In fact, empirical evidence clearly shows large differences between speech production and perception. For example, the strong motor activity controlling overt articulator muscle movements during speech production is different from the slightly enhanced excitability of articulatory motor regions in speech perception (Fadiga et al., 2002) and clear dissociations at the level of neural activity have also been demonstrated using fMRI (Figure 1 in Pulvermüller et al. 2006; see also Markiewicz and Bohland, 2016). Apart from differences in *degree* of activation (e.g., motor activity being strong in production but weak/sub-threshold in perception), further important differences between production and perception are obvious. For example, trivially, subjects are overtly moving their articulators in production thus generating somatosensory self-stimulation, whereas both of these processes are absent in passive speech perception. Likewise, acoustic stimulation with speech sounds leads to acoustic processes not present during speech motor programming or silent articulation. Already due to these obvious cognitive-behavioral and related neurophysiological differences alone, significantly different neuronal activation patterns are to be expected between production and perception. However, such necessary differences cannot argue against shared auditory and sensory *mechanisms*, i.e. production and perception mechanisms may both involve the activation of shared action-perception circuits as one of their components.

In summary, it appears unreasonable to expect *identical* neural activation for motor action and concordant perception (in this case silent articulation or ‘miming’ of speech sounds and their perception). Rather, the aspects of neural activity shared between perception and production can only be a subset of the total activity patterns present during both. Hence, when testing a classifier in a condition which shares only some of the relevant processes with the condition it was trained on, it is no surprise that cross-decoding is difficult. Such a result fits well with general observations from other MVPA studies, which found, firstly, that in general cross-decoding performance is reduced when performed across different modalities (auditory vs. written word presentation; Akama et al., 2012), but, critically, that cross-modal classification accuracies are often asymmetrical depending on cognitive features. For example, Cichy et al. (2012) found that cross-decoding from imagery to perception was less successful than vice versa, supposedly because the neural patterns of imagery are only a subset of those of perception (see also de Borst and de Gelder, 2016). Similarly, Oosterhof et al. (2012) found that cross-decoding was more successful when training on imagery and testing on action execution than vice versa. Hence, it appears as generally difficult to succeed with cross-decoding of perceptual/cognitive patterns from motor tasks; in the case of Arsenault and Buchsbaum (2016) additional complications were introduced because a motor response task was present in the perception condition, but not the miming condition; conversely, head motion induced artifacts might have been present in the miming but not the perception condition. Hence, further differences between the two conditions were introduced, which could contribute to the classifier learning features which are discriminative only in miming but not in perception and vice versa. Therefore, both the motor response task, while being problematic in itself (as discussed above), and the fact that overt articulation rather than minimal articulator movements were used, likely also contributed to difficulties in multivariate cross-decoding by adding further differences between conditions. In conclusion, Arsenault and Buchsbaum's (2016) lack of success in decoding speech perception information based on miming data does not come as a surprise and cannot be interpreted as evidence for or against specific neurocognitive models.

## The functional relevance of (phonological information in) sensorimotor cortex for speech perception and understanding

The neurophysiological experiments reviewed above show that phonological information about perceived speech, including abstract phonemic distinctive features such as place of articulation, is reflected in differential patterns of activation in motor cortex. These results are of great theoretical interest, as they help to decide between competing theories that view speech perception either as a fractionated sensory process or as an interactive mechanism involving both action and perception information and mechanisms.

However, the mere activation of sensorimotor cortex in perception could be due to intentional articulatory activity, which adds to the perception mechanism from which it is otherwise functionally divorced. Such motor activity may be sub-threshold and may thus appear while no corresponding movement or muscle activity occurs. Motor activity during, but entirely independent of perception, may be linked to motor preparation or to predicting future perceptual input. To judge this possibility, it is critical to find out whether perceptually-induced motor activation indeed carries a more general function in speech processing. Already some brain activation studies suggest a functional role of motor cortex activation in speech processing. One study found that the magnitude of speech-evoked motor activity reflects working memory capacities of experiment participants (Szenkovits et al., 2012). Other work showed that perceptually-induced motor activation reflected the type of language learning by which novel 'pseudo-words' had been acquired. Fronto-central cortical responses to novel sequences of spoken syllables increased when subjects familiarized themselves with these items by repeated articulation, whereas passive perceptual learning of the same speech items did not lead to comparable sensorimotor activation (Pulvermüller et al., 2012; Adank et al., 2013). Further indication of functional contributions of motor systems to speech perception and comprehension comes from the observation that practice in producing unfamiliar sounds or accents significantly improves their discrimination/comprehension (Catford and Pisoni, 1970; Adank et al., 2010; Kartushina et al., 2015). Similarly, learning-induced plasticity in the motor system has been shown to alter speech percepts (Lametti et al., 2014). Therefore, perceptually induced motor activity may signify articulatory learning, working memory and long-term memory for speech sounds and spoken word forms.[12]

The strongest statement of an integrative active perception account, however, addresses a putative causal role of motor systems in perceptual processing. Is the motor system causal

---

[12] Park et al. (2016) suggest a functional role of the articulatory motor cortex even during audiovisual speech perception, since visually perceived lip movements were found to entrain oscillations in lip motor cortex and the degree of coherence correlated with comprehension accuracy.

for speech perception and understanding? To decide this crucial issue, a neuropsychological research strategy is required, which investigates whether functional changes in the sensorimotor cortex impact on speech perception. Indeed, TMS studies have demonstrated that the motor system has a causal influence on the discrimination and classification of speech sounds (Meister et al., 2007; D'Ausilio et al., 2009; Möttönen and Watkins, 2009; Krieger-Redwood et al., 2013; Rogers et al., 2014; Murakami et al., 2015). Similar TMS modulation in phonological tasks has also been demonstrated for the inferior frontal and supramarginal gyrus (SMG; Hartwigsen et al., 2010a, 2010b, 2016). Over and above any general causal influence on speech discrimination performance, a phoneme specific effect of local sensorimotor stimulation has been demonstrated by a number of TMS studies comparing speech sounds with different place of articulation (usually bilabials vs. alveolars, see D'Ausilio et al., 2009; Möttönen and Watkins, 2009). These studies showed a facilitation of phonological discrimination of 'body-part congruent' sensorimotor stimulation on the processing of phonemes. For example, tongue area TMS specifically accelerated (and improved) the perceptual classification of 'tongue sounds' such as /d/ and /t/. These results converge with the earlier fMRI study on the topographical specificity of the place of articulation of speech sounds in sensorimotor cortex. In addition to showing phoneme-specific topographic activation, they also indicate a causal role of motor cortex in perception.

As mentioned before, research addressing the causality question requires a neuropsychological research strategy whereby the manipulated independent variable is the change of brain states (e.g., by TMS) and the measured dependent variable is a behavioral response, for example the accuracy and/or latency of a button press. Therefore, all neuropsychological studies require an overt motor task and any task administered in an experimental laboratory is to a degree 'unnatural', such studies are open to criticisms. Researchers holding a critical attitude towards action perception theory, for example Hickok (2014), choose to criticize the use of phoneme identification and discrimination tasks as ‘unnatural’ and possibly engaging processes not required in everyday language use and understanding. This position does not come without any reason, as pressing a button labeled with the letter ‘p’ or ‘d’ is certainly not an activity normal listeners would frequently engage in when hearing and processing speech. In this context, it has been argued that TMS might not modulate perception but rather decision-related processes instead. Different mappings on motor system areas might therefore reflect aspects of decisions, not phonological information. However, an explicit investigation of this issue using signal detection theory found that after TMS to lip motor cortex, changes in speech perception tasks are driven by changes in perceptual sensitivity but not by decision-related processes such as response bias (Smalle et al., 2015). Furthermore, even in the absence of any task, Möttönen et al. (2013) found that an attention-independent neurophysiological index of speech sound processing known as the mismatch negativity or MMN (Näätänen et al., 1997), was reduced following TMS to lip motor cortex. This result shows that sensorimotor cortex stimulation modulates a major physiological marker of speech perception even in the absence of a task, although a follow-up MEG study found that this modulation appeared relatively late and was not specific to place of articulation (Möttönen et al., 2014). In sum, MMN studies indicate that articulatory motor cortex reflects speech sound processing, rather than decision related processes such as response bias, and that functional changes in this part of the motor system reduces neurophysiological correlates of speech sound processing.One may still ask, however, how this TMS functional change relates to language comprehension under normal conditions, as speech sound discrimination tasks do not provide conclusive evidence about any causal role in language comprehension. The standard task with which psycholinguists investigate single word comprehension uses pictures and has subjects select a picture related to a spoken word. This *word-to-picture-matching task* (WPMT) was applied recently in two TMS experiments. In one experiment (Schomers et al., 2015), pictures were shown whose typical verbal

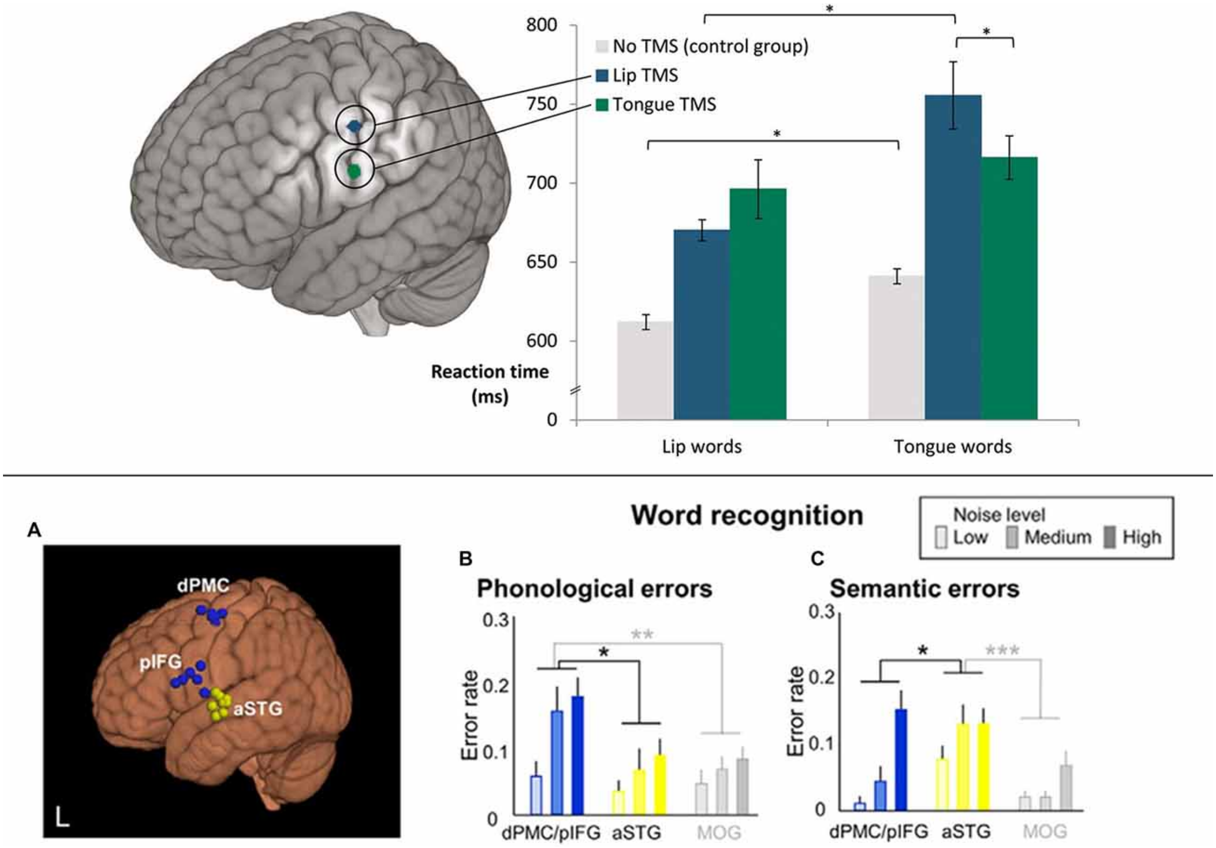


**Figure 3.4.** Transcranial magnetic stimulation (TMS) studies showing causal effects of frontal cortex stimulation on speech comprehension (word-to-picture matching).
**(Top)** Double TMS pulses to different articulator representations in motor cortex (lip vs. tongue) led to relative facilitation in word comprehension responses for words starting with a phoneme related to the congruent articulator, as revealed by a significant interaction of stimulation locus and word type ("lip words" vs. "tongue words"). $*p < 0.05$. Adapted from Schomers et al. (2015; Figure 1) by permission of Oxford Univ. Press, material published under a Creative Commons Attribution (CC-BY-NC) license.
**(Bottom)** A simultaneous virtual lesion in both dPMC and pIFG (using "double-knockout" thetaburst TMS) led to significantly increased semantic and phonological errors in word recognition (word-to-picture matching). $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Adapted from Murakami et al. (2015; Figure 6).

labels were phonological 'minimal pairs' only differing in their word-initial phoneme, which was either a [+bilabial] lip-related or [+alveolar] tongue-related speech sound (for example, pictures of a deer and a beer were shown while the spoken word 'deer' was presented). TMS to lip- and tongue-controlling precentral sulcus differentially influenced reaction times in the comprehension of spoken words starting with [+bilabial] and [+alveolar] phonemes, respectively (see Figure 3.4 top), thus demonstrating a causal role of sensorimotor cortex on speech comprehension. As in previous studies using sub-threshold single or double TMS, a relative facilitation effect was revealed by response times. In another recent experiment, Murakami et al. (2015) used a ‘double-knockout’ thetaburst TMS protocol, a novel technique where two different brain areas are stimulated with bursts of theta frequency TMS pulses (Huang et al., 2005), causing long-lasting (up to 60 min) functional degradation simultaneously in both areas. After such ‘double-knockout’ of both pIFG and dPMC an increase in phonological errors in a WPMT was observed (see Figure 3.4 bottom). Interestingly, this effect did not significantly interact with noise level, indicating that noise overlay was not a crucial factor in observing involvement of frontal areas in speech comprehension (see section on “The Role of Scanner Noise”).

In conclusion, sensorimotor articulatory cortex does not only reveal phoneme-specific activation signatures during speech perception, it also takes a differential phoneme-specific causal role in speech perception and word comprehension. Importantly, as both facilitation and error-induction could be observed in speech comprehension tasks, the causal role of sensorimotor cortex in perceptual tasks receives strong support.

## Conclusion

So, is the sensorimotor system relevant for speech perception and comprehension? Considering the evidence available across methods, studies and laboratories, this question receives a clear ‘Yes’. Still, noise overlay and motor tasks during speech perception may cancel any measurable phonologically related activation in the motor system, including multivoxel pattern information reflecting phonological specificity.

Evidence from univariate analyses of fMRI data has long shown that various parts of the speech motor system are activated during passive speech perception. Some of these studies even found specific phonological information, e.g. about place of articulation or voicing, present in these areas. Recently, several fMRI studies using MVPA replicated and extended the earlier findings. An open question that remains is what the precise role of the different regions in the sensorimotor system is, in particular the IFG, the premotor, primary motor and somatosensory cortices (see Hertrich et al., 2016, for a recent review on the role of the supplementary motor area). Mechanistic neurobiological models suggest that the roles of neurons in primary, secondary and higher multimodal areas in both frontal and temporal lobes can be understood in terms of distributed functional circuits within which distributionally different patterns of activation are the basis of the perception, recognition and working-memory storage of phonemes and meaningful units (Pulvermüller and Garagnani, 2014; Grisoni et al., 2016).

Still, there is substantial divergence between some of the reported findings regarding the precise locations where phonological information can be detected in the neurometabolic response (see Table 3.1). We argue here that at least a significant portion of this variance can be explained by differences in methods, in particular by the features of scanner noise and preparatory motor activity. Activity in motor cortex, especially precentral gyrus, seems to be vulnerable to both (whereas activity close to auditory areas and in somatosensory cortex is not as much influenced by preparatory motor activity). Hence, in order to observe motor system activity in perception experiments, it is of the essence to reduce acoustic noise and 'motor noise' as much as possible, i.e., to use sparse imaging and avoid having subjects engage in (even only occasional) button presses throughout the experiment. Finally and most importantly, any discrepancies in fMRI results are secondary in light of clear evidence from TMS that modulation of sensorimotor and frontoparietal areas causes functional changes in speech perception and comprehension, both measured neurophysiologically (Möttönen et al., 2013, 2014) and behaviorally (D'Ausilio et al., 2009; Möttönen and Watkins, 2009; Hartwigsen et al., 2010a, 2010b, 2016; Rogers et al., 2014; Bartoli et al., 2015; Murakami et al., 2015; Schomers et al., 2015; Smalle et al., 2015).

**TABLE 3.1 | Overview of functional magnetic resonance imaging (fMRI) studies investigating involvement of inferior frontal, sensorimotor and inferior parietal systems in syllable perception.**

| No. | Study | Stimuli | Phonetic features investigated | Task | Button presses | Sparse imaging | Analyses | Baseline | Activation/Decoding found in . . . | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | **Prefrontal areas** | **Motor areas** | **Somatosensory and inferior parietal areas** |
| 1 | Benson et al. (2001) | 15 C/VC/CVC syllables | n/a | none | never | yes | univariate | non-speech tones | left BA 9, 10 | left BA 6 | left SMG (BA 40) |
| 2 | Wilson et al. (2004) | /pa/, /gi/ | n/a | none | never | no | univariate | rest/silence | not reported | ventral (v) BA 4, 6 | right SMG (BA 40) |
| 3 | Wilson and Iacoboni (2006) | 50 consonants embedded between two /ɑ/ vowels | n/a | none | never | no | univariate | rest/silence | not reported | v BA 4, 6 | not reported |
| **4** | Szenkovits et al. (2012) | monosyllabic pseudowords | n/a | one-back repetition detection task | **10% of trials** | **no** | univariate | non-speech buzzes | not reported | not reported[d] | not reported |
| 5 | Grabski et al. (2013) | 9 vowels | n/a | none | never | yes | univariate | rest/silence | BA 44, left BA 45 | right BA 6 | not reported |
| 6 | Pulvermüller et al. (2006) | /pæ/, /tæ/, /pl/, /tl/ | Place | none | never | yes | univariate, ROI-based[a] | matched noise stimuli | not reported | left v BA 4, 6 (differential activation of lip vs. tongue regions) | not reported |
| 7 | Raizada and Poldrack (2007) | /ba/, /da/ | Place | detect occasional quieter stimulus | 6% of trials | yes | repetition adaptation | n/a | left middle frontal cortex (amplification of response to stimulus pairs differing in place) | not reported | left SMG (amplification of response to stimulus pairs differing in place) |
| 8 | Myers et al. (2009) | /da/, /ta/ | Voicing | detect occasional high-pitched stimulus | 37.5% of trials | yes | repetition adaptation | n/a | left inferior frontal sulcus (release from adaptation only for stimuli differing in voicing) | not reported | not reported |
| 9 | Lee et al. (2012) | 10 CV syllables on /ba/-/da/ continuum | Place | detect occasional quieter stimulus | 11% of trials | yes | MVPA (searchlight) | n/a | left BA 44 (decoding of place) | left pre-SMA (decoding of place) | not reported |

*(Continued)*

**TABLE 3.1 | (Continued).**

| No. | Study | Stimuli | Phonetic features investigated | Task | Button presses | Sparse imaging | Analyses | Baseline | Activation/Decoding found in . . . | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | **Prefrontal areas** | **Motor areas** | **Somatosensory and inferior parietal areas** |
| 10 | Chevillet et al. (2013)[b] | /da/–/ga/ continuum | Place | dichotic listening (detect in which ear the sound persisted longer) | always | yes | repetition adaptation | n/a | not reported | left BA 6 (release from adaptation for stimulus pairs differing in place of articulation) | not reported |
| **11** | Du et al. (2014) | /ba/, /ma/, /da/, /ta/ | Place | active syllable identification (4-AFC) | **always** | **no** (but scanner noise attenuation by 25 dB) | MVPA (searchlight) | n/a | Insula/Broca's area (decoding of place)—at low/moderate noise levels only[e] | left v BA 6 (decoding of place)—at low noise levels only[f] | left inferior parietal lobule (decoding of place)—at low noise levels only |
| **12** | Arsenault and Buchsbaum (2015) | 16 CV syllables | Place, manner, voicing | gender identification task | **always** | **no** | MVPA (ROI-based) | n/a | not reported | not reported | left subcentral gyrus (decoding of place) |
| 13 | Evans and Davis (2015) | /ba/, /da/, /ma/, /na/, /ab/, /ad/ | Place, manner, phoneme identity, CV structure (CV vs. VC) | one-back repetition detection task | 8% of trials | yes | MVPA-based RSA (searchlight)[c] | rest/silence | not reported | left precentral gyrus (decoding of syllable and phoneme identity and CV structure) | left postcentral gyrus (decoding of syllable identity) |
| 14 | Correia et al. (2015) | 24 CV syllables | Place, manner, voicing | none | never | yes | MVPA (searchlight) | n/a | IFG (decoding of place/manner) | right inferior precentral gyrus (decoding of place) | postcentral gyrus, SMG (decoding of place/manner) |
| **15** | Arsenault and Buchsbaum (2016) | 8 CV syllables: /ba/, /pa/, /va/, /fa/, /da/, /ta/, /za/, /sa/ | Place (manner/voicing not analyzed) | detect occasional blank trials | **11% of trials** | **no** | univariate and MVPA (both ROI-based) | n/a | not reported | not reported | left postcentral gyrus (decoding of place) |

[a] *Differential activation of subregions (lip vs. tongue) in left BA 4, 6 depending on place of articulation (lip vs. tongue; see* ***Figure 3.2*** *Top).* [b] *See also Alho et al. (2016) for similar results from MEG.* [c] *Representational similarity analysis (RSA) used to disentangle acoustic and phonological similarity.* [d] *Not reported in whole-brain analysis; but note that in a ROI-based analysis a correlation between working memory abilities and motor/premotor activation in speech perception was found.* [e] *Due to the active identification task on every trial, this might reflect decision-related processes or response selection/preparation (see main text for discussion).* [f] *In dorsal BA 6 (PMC) decoding was still observed at higher noise levels (see* ***Figure 3.3****), but this might reflect classification of responses rather than speech perception (see main text for discussion).*

# 4. Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence

---

This chapter is based on:

Schomers, M.R., Kirilina, E., Weigand, A., Bajbouj, M., Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. ***Cerebral Cortex*** 25(10):3894, https://doi.org/10.1093/cercor/bhu274

## Abstract

Classic wisdom had been that motor and premotor cortex contribute to motor execution but not to higher cognition and language comprehension. In contrast, mounting evidence from neuroimaging, patient research, and transcranial magnetic stimulation (TMS) suggest sensorimotor interaction and, specifically, that the articulatory motor cortex is important for classifying meaningless speech sounds into phonemic categories. However, whether these findings speak to the comprehension issue is unclear, because language comprehension does not require explicit phonemic classification and previous results may therefore relate to factors alien to semantic understanding. We here used the standard psycholinguistic test of spoken word comprehension, the word-to-picture-matching task, and concordant TMS to articulatory motor cortex. TMS pulses were applied to primary motor cortex controlling either the lips or the tongue as subjects heard critical word stimuli starting with bilabial lip-related or alveolar tongue-related stop consonants (e.g., “pool” or “tool”). A significant cross-over interaction showed that articulatory motor cortex stimulation delayed comprehension responses for phonologically incongruent words relative to congruous ones (i.e., lip area TMS delayed “tool” relative to “pool” responses). As local TMS to articulatory motor areas differentially delays the comprehension of phonologically incongruous spoken words, we conclude that motor systems can take a causal role in semantic comprehension and, hence, higher cognition.

## Introduction

One of the most fundamental debates in current neuroscience addresses the role of the frontal lobe in perception and understanding. The mirror neuron literature has shown action-perception linkage and premotor activation in perceptual and comprehension processes (Rizzolatti et al., 1996, 2014), but to which degree this frontal activation is necessary for understanding remains controversial. The left frontal cortex' contribution to symbol understanding has long been in the focus of brain research on language. In neurolinguistics, the classic view had been that the left inferior frontal cortex, including Broca's area (Brodmann areas [BA] 44, 45), and adjacent articulatory motor areas, including inferior motor and premotor cortex (BA 4, 6), are involved in speech production only, whereas for speech perception and language comprehension, auditory and adjacent Wernicke's area in superior-temporal cortex (including BA 41, 42, 22) are necessary and sufficient (Wernicke, 1874; Lichtheim, 1885; Geschwind, 1970). In sharp contrast with this position, action-perception theory of language postulates that the neuronal machineries for speech production and understanding are intrinsically linked with each other, implying a double function of both superior-temporal and, critically, inferior-frontal cortex in both production and comprehension (Pulvermüller and Fadiga, 2010). A crucial prediction of the latter perspective, which goes against the classic approach, is that left inferior frontal and even articulatory motor areas are causally involved in language comprehension, including the understanding of single words.

Several arguments seem to support a role in comprehension of left inferior frontal cortex and adjacent articulatory motor systems. Liberman and his colleagues pointed out in the context of their "motor theory of speech perception" that there is great context-dependent variability of acoustic-phonetic features of speech sounds that fall in the same phonemic category and that a link between these disparate auditory schemas might be possible based on the similar articulatory gestures performed to elicit the sounds (Liberman and Mattingly, 1985; Galantucci et al., 2006). In view of the neurobiological mechanisms of speech perception, it was argued that the correlated articulatory motor and auditory information available to the human brain during babbling and word production fosters the development of action-perception links between frontal articulatory and temporal auditory areas. Formation of such frontotemporal links for phonemes is implied by the fundamental principle of Hebbian correlation learning and the neuroanatomical connectivity profile between frontal and temporal cortex (Rilling et al., 2008), thus yielding functional frontotemporal connections between neurons controlling articulatory motor movements performed to produce speech sounds and neurons involved in acoustic perception of the corresponding speech sounds (Braitenberg and Pulvermüller, 1992; Pulvermüller et al., 2006; Pulvermüller and Fadiga, 2010). In this view, neuronal circuits distributed across superior-temporal and inferior-frontal areas that reach into articulatory motor cortex carry both language production and comprehension. Further arguments against a functional separation of receptive and productive language areas come from studies in patients with post-stroke aphasia. Broca's aphasics, in addition to their speech production deficits, appear to be impaired in speech perception tasks, which typically involve the explicit discrimination between, or identification of, phonemes or syllables (Basso et al., 1977; Blumstein et al., 1977, 1994; Utman et al., 2001). However, attempts at replicating these findings were not always successful and, crucially, as critical lesions are sometimes extensive, lesion studies do not provide unambiguous evidence for a general involvement of left inferior frontal cortex in speech perception. For example, only three Broca's aphasics in a group study documenting a perception deficit indeed had lesions restricted to frontoparietal cortex (Utman et al., 2001); whereas one study (Rogalsky et al., 2011) failed to detect profound speech perception deficits in two aphasics with selective inferior frontal lesions, another study (Caplan et al., 1995) documented them in a different set of two such patients.

In this context, it is important to emphasize the distinction between the *perception* of phonemes and the *comprehension* of meaningful language. Although all linguistic signs serve the role of communicating meaning, phonemes, the smallest units that *distinguish between* meaningful signs, do themselves not *carry* meaning. If they are presented outside the context of words, they lack their normal function as meaning-discriminating units. Importantly, speech perception performance of aphasia patients dissociates from their ability to understand the meaning of single words (Miceli et al., 1980). Therefore, a deficit in a speech perception task does not imply a comprehension failure. Although some authors have argued that word comprehension is still relatively intact in some patients with aphasia, because semantic evidence can in part be used to reconstruct missing phonetic information (Basso et al., 1977; Blumstein et al., 1994), it is possible that word comprehension experiments reveal additional information about left inferior frontal cortex function. Spoken word comprehension deficits in patients with Broca's aphasia have been documented when stimuli were degraded and embedded in noise but not when they were spoken clearly (Moineau et al., 2005). However, a significant delay in comprehending clearly spoken single words was reported in this and related studies (Utman et al., 2001; Yee et al., 2008). Still, these delays may be attributable to general cognitive impairments or strategic aspects of language processing. Clear evidence from small well-documented lesions in left inferior frontal cortex documents comprehension deficits for specific semantic types of words, especially action words (Neininger and Pulvermüller, 2001, 2003; Kemmerer et al., 2012), and a similar category-specific deficit is present in patients with degenerative brain diseases primarily affecting cortical regions in the frontal lobe, including Motor Neuron Disease and Parkinson's Disease (Bak et al., 2001; Boulenger et al., 2008). In summary, patient studies could so far not finally clarify whether the left inferior frontal cortex or the articulatory motor system serve a general causal role in speech comprehension. Any firm conclusions are hampered by uncertainties about precise lesion sites, known differences between speech perception and comprehension, uncertainties about the interpretation of delays in speech comprehension and the category-specific nature of some well-documented comprehension deficits.

In addition to patient studies, neuroimaging studies employing a wide variety of methods (univariate and multivariate fMRI, EEG/MEG, and connectivity analysis) have found activation of left inferior frontal and articulatory motor cortices in speech perception (Pulvermüller et al., 2003, 2006; Wilson et al., 2004; Osnes et al., 2011; Lee et al., 2012; Chevillet et al., 2013; Liebenthal et al., 2013; Alho et al., 2014; Du et al., 2014). One study by Pulvermüller et al. (2006) even revealed that information about the specific place of articulation of a passively perceived speech sound is manifest in focal activation of the articulatory representations in the motor system. In addition, transcranial magnetic stimulation (TMS)-evoked activation of articulatory muscles (lip or tongue) was shown to be increased while listening to meaningful speech (Fadiga et al., 2002; Watkins et al., 2003) and even when viewing hand gestures associated with spoken words (Komeilipoor et al., 2014). However, although these results show that speech perception and comprehension elicit left inferior frontal and articulatory motor cortex activation, the possibility exists that such activation is consequent to but not critical to perception and comprehension processes (Mahon and Caramazza, 2008).

TMS can also be used to reveal a possible causal role of motor areas in language processing, thereby overcoming some of the limitations mentioned earlier. TMS induces functional changes in cortical loci, which can be localized with an accuracy of 5–10 mm (Walsh and Cowey, 2000). This neuropsychological research strategy has been applied using tasks requiring explicit classification of speech sounds embedded in meaningless syllables (Meister et al., 2007; D'Ausilio et al., 2009; Möttönen and Watkins, 2009) (for reviews, see Möttönen and Watkins, 2012; Murakami et al., 2013). The results showed a causal effect on the perceptual classification of noise-embedded speech sounds.

In this context, it is important to highlight once again that, in previous TMS studies, phonemes were presented in isolation; thus, the tasks did not entail the comprehension of the meaning of speech and, therefore, it remains unclear whether the influence of articulatory motor areas extends to normal speech comprehension and semantic understanding.

In summary, on the basis of existing data from patient, neuroimaging, and TMS studies, the role of inferior frontal and motor cortex in language comprehension remains controversial. Some authors deny any role of these areas in comprehension completely or acknowledge an influence only in artificial tasks and/or degraded listening conditions (Hickok, 2009, 2014; Rogalsky et al., 2011). Others implicate only anterior areas, in particular anterior parts of Broca's area in semantic processing, whereas more posterior parts of frontal cortex (posterior Broca's area and premotor/motor cortex) are considered at most to play a role in phonological processing (Bookheimer, 2002; Gough et al., 2005). However, the crucial hypothesis about the role of left inferior frontal, and most importantly, left articulatory motor cortex in meaningful spoken word comprehension, still awaits systematic experimental testing.

To clarify the left articulatory motor cortex's role in single word comprehension, we here used the standard psycholinguistic test of spoken language comprehension, the word-to-picture-matching task. Naturally spoken words, which were minimal pairs differing only in their initial phoneme (e.g., “pool” and “tool”), were used as critical word stimuli to ascertain that, in the context of the experiment, the initial phonemes served their normal role as meaning-discriminating units. All critical phonemes were either bilabial (lip-related) or alveolar (tongue-related) stop consonants. Immediately before onset of the spoken word, TMS was applied to left articulatory motor cortex, either to the lip or tongue representation. For word-to-picture matching, two images appeared immediately after the word and subjects had to indicate as fast as possible, which of them corresponded to the meaning of the word. Classic brain-language models predict no influence of motor cortex TMS on single word comprehension, whereas action-perception theory of language suggests a modulation of language comprehension performance specific to phonological word type and specific to TMS locus.

## Materials and Methods

### ***Participants***

Thirteen monolingual native speakers of German (6 females) with a mean age of 22 years (range: 18–28 years) participated in the experiment for financial compensation or course credit. They had no history of neurological or psychiatric illness, normal or corrected-to-normal vision, and normal hearing (as assessed by a questionnaire). All participants were right-handed (Oldfield, 1971) (laterality quotient $M$ = 88.8, SD = 19.4). Participants provided written informed consent prior to participating in the study, and procedures were approved by the Ethics Committee of the Charité University Hospital, Berlin, Germany.

### ***Stimuli***

Stimuli were a total of 140 German words. Forty-four of those were critical experimental stimuli, and 96 were filler stimuli. The critical word stimuli were 22 minimal pairs differing only in their initial phoneme (see Table 4.1). In each pair, one word started with a bilabial (and therefore “lip-related”) stop consonant ([b] or [p]), and the other with an alveolar (“tongue-related”) stop consonant ([d] or [t]). We refer to these words as “lip words” and “tongue words”, respectively. Nineteen of the 22 word pairs were also matched in the feature (voicing), yielding similar voice onset times. Tongue and lip words were matched for the following psycholinguistic variables (mean ± SD) obtained from the dlexDB database (Heister et al., 2011): normalized word type frequency (lip words: $M$ = 18 ± 28.9; tongue words: $M$ = 81 ± 192), normalized average character bigram corpus frequency (lip words: $M$ = 48 942 ± 26 130; tongue words: $M$ = 57 234 ± 26 907) and normalized average character trigram corpus frequency (lip words: $M$ = 39 609 ± 26 864; tongue words: $M$ = 49 216 ± 27 421). All comparisons between lip words and tongue words were non-significant ($P > 0.05$).

We took efforts to distract subjects from the phonological nature of the present experiment and its focus on tongue- vs. lip-related phonemes. To this end, 96 irrelevant "filler" stimuli were added to the 44 experimental ones, including 24 semantically, but not phonologically related word pairs (e.g., "apple" vs. "cherry"), 12 minimal pairs differing in their initial phonemes, which were not bilabial vs. alveolar stops (e.g., "key" vs. "fee"), and 12 word pairs differing in the final phonemes (e.g., "gun" vs. "gum"). The full list of filler stimuli is provided in Table 4.2.

All stimulus words were naturally spoken by a female native speaker of German during a single session in a soundproof chamber. They were recorded using an ATR1200 microphone (Audio-Technica Corporation), digitized (44.1 kHz sampling rate), and stored on disk. After recording, noise was removed and loudness of the each recording normalized to -20 dB spl using "Audacity" (http://audacity.sourceforge.net). Average duration of stimuli was 432 ms (SD ± 50 ms) for lip words and 445 ms (SD ± 41 ms) for tongue words; a paired *t*-test showed no significant difference in duration.

For each stimulus word, a corresponding picture was chosen. The corresponding picture showed an object the word is typically used to speak about. Two of the experimental words were too abstract to find an obvious pictorial correlate of their meaning, and a circle was used as abstractness symbol so that the decision between pictorial alternatives had to be based on the distinction between one semantic match and one "abstractness" indicator. For the remaining 42 critical stimulus words, semantically related pictures could be produced easily.

### TMS Methods

Structural MRI images were obtained for all participants (3T, Tim Trio Siemens, T1-weighted images, isotropic resolution 1 × 1 × 1 $mm^3$) and used for frameless stereotactic neuronavigation (eXimia Navigated Brain Stimulation, Nexstim). TMS pulses were generated by a focal biphasic figure-of-eight coil (eXimia 201383P). Coil position was maintained at roughly 90° to the central sulcus in the direction of the precentral gyrus. Using single TMS pulses, the lip representation in the left motor cortex was localized in each subject by measuring electromyographic activity in two surface electrodes attached to the right orbicularis oris muscle (as described in Möttönen and Watkins, 2009), or, if this was not possible, the first dorsal interosseus (FDI) muscle was localized according to standard procedures (Rossini, 1994). Target articulator loci not directly localized in an individual were calculated using the method described in a previous TMS study by D'Ausilio et al. (2009): Montreal Neurological Institute (MNI) coordinates for the left lip and tongue representation from an earlier fMRI study (Pulvermüller et al., 2006) were converted to individual subjects' head space. In subjects for whom only the individual FDI location was available, the target coordinates were adjusted according to the difference between the actual FDI location and standard FDI coordinates (Niyazov et al., 2005) ($x$, $y$, $z$ = −37, −25, 58). In those subjects where the lip representation could be localized, only the tongue coordinate was adjusted according the difference between actual and projected lip location. The average MNI coordinates of the actual stimulation sites were as follows: lip MNI $x$, $y$, $z$ = −55.4, −9.2, 43.9 and tongue MNI $x$, $y$, $z$ = −59.4, −7.4, 22.8 (see Fig. X). Thus, the distance between average actual stimulation sites and standard MNI peaks found in the fMRI study by Pulvermüller et al. (2006) and used by D'Ausilio et al. (2009) was 2.5 mm for the lip site (MNI peak: $x$, $y$, $z$ = −56, −8, 46) and 3.5 mm for the tongue site (MNI peak: $x$, $y$, $z$ = −60, −10, 25). The average distance between the two stimulation sites was 22.3 mm.

In those subjects in whom lip motor threshold could be determined ($n = 3$), stimulation intensity was 90% of the lip motor threshold. In subjects for whom only FDI motor threshold was available ($n = 10$), on average 100% of the FDI threshold (Rossini, 1994) was used because higher intensities are normally required to magnetically stimulate the articulators.

Average intensity was 34.6% of total stimulator output. TMS pulses (10 Hz, i.e., separated by 100 ms) were delivered before the onset of the spoken target word. Previous studies (Moliadze et al., 2003; Mottaghy et al., 2006) showed that facilitatory TMS effects emerge when the delay between TMS and critical stimulus onset is in the range from >100 to 500 ms. We chose a delay of 200 ms between the last pulse and spoken word onset to minimize acoustic interference of the clicking sounds accompanying TMS pulses with the spoken stimuli. We initially used 3 pulses to potentially further increase TMS efficacy (Kammer and Baumann, 2010); however, after running 3 subjects, we noticed that this caused a risk of coil overheating and thus used 2 pulses for the remaining subjects.

### *Experimental Procedures*

Each subject attended two TMS sessions, separated by at least 2 weeks. There were two blocks per session. In each block, 70 trials were administered, each with 22 critical trials and 48 filler trials. Trial order was pseudo-randomized with no more than three stimuli of one type allowed in direct succession. A trial consisted of one spoken word immediately followed by a picture pair. Subjects had to indicate as fast and as accurately as possible which of the pictures matched the spoken words. For each word pair, the two complementary spoken targets were presented in separate blocks of each session. Alternative pseudorandomization orders were produced using the software Mix (van Casteren and Davis, 2006) so that all stimuli were identical in the two sessions but in a different order. In each block, TMS was delivered to one articulator locus, the order being counterbalanced over subjects. Stimulus word delivery was through in-ear headphones (Koss Corporation), which also provided attenuation of noise created by the TMS machine and stimulation.

To reduce the degree of redundancy immanent to the speech signal and avoid ceiling effects (see Discussion for further explanation), we individually adjusted the sound pressure level to a range where subjects were still able to repeat most of a set of test words correctly

(*M* = 69%, SEM = 3.7%) in a pilot screening. This screening was followed by a training session consisting of 11 trials of word-to-picture matching but with different items than those used in the experiment. During the training session, TMS pulses were also delivered so that subjects could familiarize themselves with the task, in particular with carrying it out under concurrent TMS stimulation.

Each trial started with a fixation cross in the middle of the screen displayed for an interval randomly varying between 1 to 2 s, after which the target sound was played via the headphones. The TMS pulses were delivered just prior to the presentation of the auditory stimulus such that the last pulse occurred 200 ms before the word onset. 700 ms after spoken word onset, the two target images appeared simultaneously for 500 ms on screen to the left and right of fixation. Subjects were instructed to make their response as quickly and accurately as possible; responses were allowed within a period of 1.5 s after image onset. The following trial started after an inter-trial period ranging between 1.0 and 2.5 s. Subjects responded using their left hand, pressing a left arrow key for the image on the left and a right arrow key for the image on the right. Each response key was associated with lip words and tongue words equally often and in a randomized order. Stimuli were presented on a Windows PC running Matlab 2012 and the Cogent 2000 toolbox (http://www.vislab.ucl.ac.uk/cogent.php). The script triggered the onset of the TMS pulses through a direct BNC cable to the TMS device.

Twelve separate right-handed subjects participated in a behavioral control study. They completed four blocks of 70 trials each with the same items and randomization orders as in the TMS experiment but without TMS application and all in one single session.

### *Data Analysis*

Only data from experimental items were analyzed (see Stimuli). A total of 10 trials were excluded because no TMS pulse was given on these trials due to technical problems with the neuronavigation system. Reaction times below 350 ms were excluded as outliers (1.6% of data) as well as those exceeding ±2 SDs of each subject-and-session-specific mean (4.7% of data). Furthermore, there were five word pairs for which accuracy across all subjects was not significantly different from chance (tested using chi-square tests) and therefore those word pairs were excluded from analysis. The same procedures were applied to the data from the control experiment without TMS.

For the TMS data, we conducted a repeated-measures analysis of variance (ANOVA) with a 2 × 2 design with factors TMS Location (Lip/Tongue) and Word Type (Lip/Tongue). We analyzed both reaction time (correct responses only) and accuracy (trials where no response was given were discarded, 4.1% of experimental trials) in separate ANOVAs; both were *z*-score-transformed to each subjects' mean and SD. We also analyzed reaction time data with a linear mixed-effects model using R version 3.0.1 and the package lme4 (Bates et al., 2014). In this model, we employed the same 2 × 2 design as in the ANOVA but also added random intercepts for each word and each subject. To test significance of effects, we calculated degrees of freedom using the approximation described by Kenward and Roger (1997).

**Table 4.1. List of experimental stimuli**

| | | |
|---|---|---|
| Bauer – Dauer | Pier – Tier | Buch – Tuch |
| Bube – Tube | Beil – Teil | Paste – Taste |
| Bach – Dach | Becher – Dächer | Packer – Tacker |
| Pop – Top | Panne – Tanne | Bank – Dank |
| Bass – Dass | Punk – Tank | Passen – Tassen |
| Punkt – tunkt | Birne – Dirne | Bill – Dill |
| Bühne – Düne | Bellen – Dellen | Ben – Denn |
| Bier – Dir | | |

**Table 4.2. List of filler stimuli**

(a) semantically, but not phonologically related word pairs

| | | |
|---|---|---|
| Hocker – Sessel | Apfel – Kirsche | Tomate – Karotte |
| Tasche – Tüte | Schere – Zange | Schaufel – Besen |
| Schlange – Krokodil | Auge – Nase | Korb – Fass |
| Schrank – Regal | Maus – Ratte | Schiff – Boot |
| Baum – Strauch | Haus – Hütte | Feile – Säge |
| Messer – Löffel | Klavier – Orgel | Fahrrad – Roller |
| Rucksack – Koffer | Hut – Mütze | Brot – Croissant |
| Flasche – Kanne | Rollstuhl – Krücken | Flugzeug – Ballon |

(b) word pairs differing in the initial phonemes (but not bilabial vs. alveolar pairs)

| | | |
|---|---|---|
| Teller – Keller | Kessel – Sessel | Dose – Hose |
| Berg – Zwerg | Stift – Gift | Kind – Wind |
| Mauer – Sauer | Garten – Karten | Pappe – Klappe |
| Tonne – Sonne | Gasse – Kasse | Hahn – Zahn |

(c) word pairs differing in the final phonemes

| | | |
|---|---|---|
| Wurst – Wurm | Braten – Bratsche | Messer – Messen |
| Gleis – Gleich | Griff – Grill | Falle – Falke |
| Teich – Teig | Spatz – Spaten | Segel – Segen |
| Feier – Feile | Bauch – Baum | Stamm - Stall |

## Results

The ANOVA on reaction times revealed a significant main effect for Word Type (lip-word-initial, “lip word” vs. tongue-sound-initial, “tongue word”) ($F_{1,12} = 9.6$, $P = 0.009$) and a significant interaction between Word Type and TMS Location ($F_{1,12} = 7.1$, $P = 0.021$). Post-hoc Newman–Keuls tests showed a significant difference for tongue words between the two different stimulation locations ($P = 0.042$), and for the lip TMS Location, a significant difference between Word Types ($P = 0.001$) (Fig. 4.1). The linear mixed-effects model confirmed the significant main effect of Word Type ($F = 4.8$, $P = 0.036$) and significant interaction of Word Type by TMS Location ($F = 5.19$, $P = 0.023$).

The ANOVA on accuracy data revealed a significant main effect of word type ($F_{1,12} = 5.4$, $P = 0.038$). Accuracy overall was higher for lip words ($M = 80.6\%$ correct) than that for tongue words ($M = 73\%$ correct) (SE of the difference = 2%). The interaction between Word

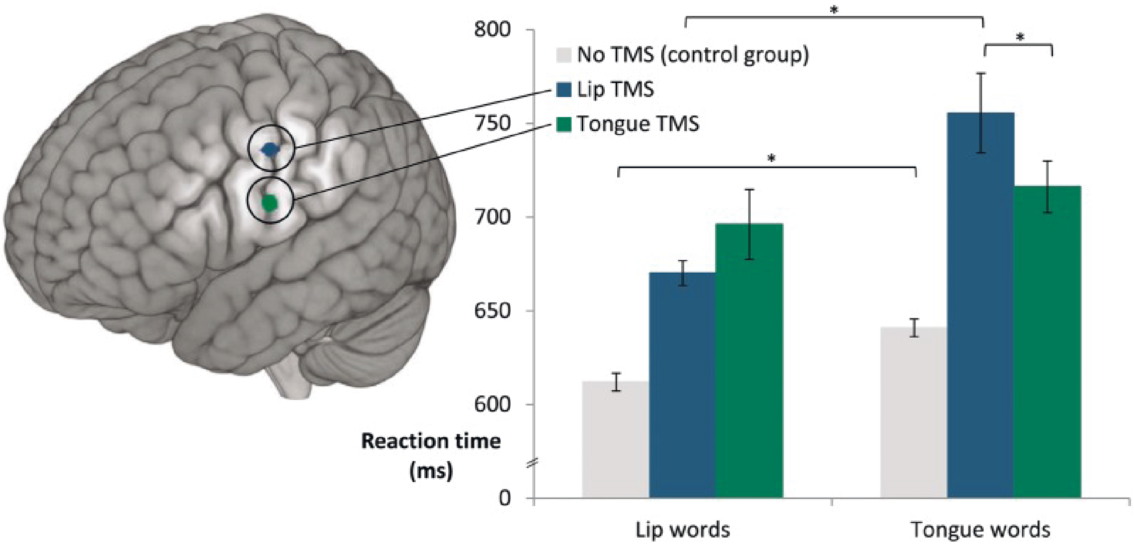


**Figure 4.1**
**(Left)** Average stimulation locations for lip and tongue representation shown on a standard MNI brain (lip ***x***, ***y***, ***z*** = −55.4, −9.2, 43.9; tongue ***x***, ***y***, ***z*** = −59.4, −7.4, 22.8).
**(Right)** Significant interaction of word type by TMS location (reaction time data). The label ‘Lip words’ denotes words starting with bilabial lip-related phonemes, whereas ‘Tongue words’ denotes words starting with alveolar tongue-related phonemes. Error bars show ±1 SEM after removing between-subject variance (Morey, 2008) *$P < 0.05$.

Type and TMS Location was not significant in the accuracy data ($F_{1,12}$ = 1.2, $P$ = 0.29). Nonetheless, there was a similar trend toward lower accuracy with TMS delivered to articulatory areas incongruent with the word-initial phoneme both for lip words (mean difference = 2.2%, SE = 2.6%) and for tongue words (mean difference = 2.3%, SE = 3.5%).

For the behavioral control experiment performed in 12 separate subjects without TMS, a paired *t*-test revealed that lip words were responded to faster (*M* = 612 ms) than tongue words (*M* = 641 ms), (*P* = 0.002, SE of the difference = 5.1 ms). A paired *t*-test on the accuracy values revealed no significant difference between word types (lip words: *M* = 79% correct, tongue words: *M* = 81% correct, *P* = 0.55).

## Discussion

We examined the effect of TMS to motor cortex on the comprehension of single spoken words by stimulating the articulatory motor representations of the tongue and lips and measuring word-to-picture-matching responses to words starting with tongue- and lip-produced stop consonants (“lip words” and “tongue words”). The analysis of reaction times showed a significant interaction between word type (lip vs. tongue word) and the location of TMS to motor cortex (lip vs. tongue locus) (Fig. 4.1). To our knowledge, this result demonstrates for the first time that articulatory motor cortex can exert a causal effect on the comprehension of single meaningful spoken words.

Further analysis of the interaction through post-hoc comparisons showed that reaction times to tongue words were significantly prolonged, by 39 ms in the average, with incongruent TMS stimulation (i.e., to the lip representation) compared with congruent TMS stimulation (of the tongue representation). Correspondingly, reaction times for lip words were delayed by an average 26 ms with incongruent TMS stimulation relative to congruous stimulation; although this latter post-hoc comparison did not reach significance, both incongruency effects together are manifest in the significant cross-over interaction.

One may ask why TMS produced a clear and significant effect on tongue-sound-initial words but not on lip-sound-initial word processing. Possible explanations include a ceiling effect for lip words, whose RTs were lower than those to tongue words (main effect $P = 0.009$), which was also confirmed in the control experiment without TMS ($P = 0.002$). This indicates that, in spite of all measures taken to control between stimulus materials, tongue-sound-initial items were already more difficult to process, thus being more sensitive to minimal TMS-elicited interference. A further possible explanation can be built on a recent suggestion by Bartoli et al. (2015): reconstructing tongue configurations from an acoustic signal might be computationally more demanding due to higher biomechanic complexity and higher degrees of freedom for tongue than for lip movements. This observation might explain why functional TMS interference was easier for tongue- than for lip-sound-initial words. A related point is that lip movements typically engage a smaller part of the motor system than even minimal tongue movements (see Fig. 1 in Pulvermüller et al., 2006). Thus, motor circuits for bilabial phonemes may be more focal than those for alveolar phonemes, making the former easier to functionally influence with TMS. Hence, TMS to focal lip areas might have been able to cause stronger interference effects on the processing of incongruent words than TMS to the relatively distributed tongue representations.

In essence, several reasons may explain why causal effects of motor cortex stimulation were pronounced for tongue-sound-initial words, but marginal for lip words considered separately; importantly, however, the significant interaction of stimulation site and phonemic word type proves that motor cortex stimulation altered word comprehension processes in a phoneme-specific manner. The behavioral pattern revealed by response time results was corroborated by accuracy data (see Results), which showed a relative processing disadvantage for word comprehension when incongruous motor cortex was stimulated. Note that both increase of response times and reduction of accuracies are consistent with a reduction in processing efficacy. Although the error data in themselves did not reach significance, this consistency is of the essence because it rules out the possibility of a speed-accuracy tradeoff.

The causal role of motor cortex in speech comprehension shown by our results is in line with neurobiological models of language based on Hebbian correlation learning (Braitenberg and Pulvermüller, 1992; Pulvermüller et al., 2006; Pulvermüller and Fadiga, 2010). When speaking, a speaker usually perceives the self-produced acoustic speech signals, thus implying correlated neuronal activity in inferior-frontal and articulatory motor cortex and in superior-temporal auditory areas. Correlation learning commands that frontotemporal circuits important for both language production and comprehension emerge. That temporal areas, apart from fronto-central ones, play a role in speech production appears uncontroversial (Lichtheim, 1885; Paus et al., 1996). The present study now presents unambiguous evidence that, apart from superior-temporal areas, also fronto-central sites and especially motor cortices take a role in speech comprehension, thus supporting the action-perception model.

We hypothesized that stimulation of a congruent sector of articulatory motor cortex (lip or tongue) would lead to faster word comprehension than the stimulation of incongruous motor areas, where competing motor programs may be activated thus causing a degree of interference. This hypothesis was confirmed by the significant interaction of stimulation site by word type. Our baseline condition without TMS was performed with a different population of subjects so that the faster responses without TMS may be due to general or specific effects of TMS or rather to a difference between subject groups. We did not include a sham TMS condition for the subjects receiving TMS because this would have made stimulus repetition unavoidable. Therefore, we cannot determine with certainty, based on the present data, whether the effects are indeed a result of interference and whether facilitation may have played a role. The previous TMS study showing motor cortex influence on phoneme classification indeed found evidence for both interference and facilitations (D’Ausilio et al., 2009). This issue should be further investigated for word comprehension. Furthermore, the observed interaction effect of word type and stimulation site in single word comprehension could have been caused directly by the stimulation of primary motor cortex neurons or, as D'Ausilio et al. (2009) suggested, by an indirect effect of primary on premotor circuits, which, in this view, would influence comprehension. Still, our results show clearly that superior-temporal cortex is not the only area causally involved in speech perception and comprehension.

Earlier studies had already shown that TMS to articulatory motor or premotor cortex has an effect on speech classification performance (Meister et al., 2007, D'Ausilio et al., 2009; Möttönen and Watkins, 2009; Sato et al., 2009). However, this earlier work has been criticized because explicit phoneme classification is never necessary in linguistic communication, and effects could have emerged at the level of phoneme classification, rather than perception per se (Hickok, 2009; McGettigan et al., 2010). This interpretation can now be ruled out on the basis of recent studies combining TMS and EEG/MEG by Möttönen et al. (2013, 2014), who found that after repetitive TMS (rTMS) of the lip representation in motor cortex, the mismatch negativity (MMN) to speech sounds was reduced even when subjects passively perceived these sounds without explicit classification tasks. Other studies could show that speech listening automatically induces changes in articulatory movements, thus providing further arguments for causal functional links between the perception and articulation of speech sounds in the absence of classification tasks (Yuen et al., 2010; D'Ausilio et al., 2014). Nonetheless, these striking results still do not speak to the comprehension issue, that is, to the question of whether earlier results on the relevance of motor systems generalize to the crucial level of semantic comprehension of meaningful speech. In order to monitor comprehension, a task needs to involve at least one of the critical aspects of semantics, for example the relationship between words and objects, as is the case in the word-to-picture-matching task used most frequently in psycho- and neurolinguistics to investigate semantic understanding.

Placing the phonemes in the context of whole words allowed us to investigate the role of motor cortex sectors related to these phonemes when these serve their normal function as meaning-discriminating units. A recent study by Krieger-Redwood et al. (2013) used a similar approach and tried to disentangle speech perception and comprehension processes. However, these authors used classification tasks, which in our view makes their results difficult to interpret. Subjects had to explicitly categorize either the final speech sound of words into phonological categories (e.g., ‘k’ vs. ‘t’) or their meaning into semantic categories (e.g., large vs. small). The authors found that rTMS to premotor cortex only delayed reaction times for phonological but not semantic judgments and conclude that the role of premotor cortex does not extend to accessing meaning in speech comprehension. Furthermore, that study used to a degree artificial ‘cross-spliced’ stimuli and TMS was to relatively dorsal premotor cortex (MNI $x$, $y$, $z$ = −52.67, −6.67, 43), which is consistent with our present lip site but distant from our tongue site, so that only a fraction of tongue-related phonemes might be affected. Finally, as mentioned earlier, both the phoneme classification and semantic classification tasks require cognitive processes of comparison and classification going beyond speech perception and comprehension. Therefore, it is unclear whether the delays in phonological categorization responses originate at the level of phoneme perception or rather at those of comparison and categorization. The delays in phonological and their absence in semantic categorization reported by these authors may therefore speak more to the categorization issue than to that of speech perception and understanding. In contrast, our results provide decisive information to resolve the speech comprehension debate as they show a causal role of articulatory motor cortex on speech comprehension when words serve their normal function as carriers of meaning.

Earlier TMS studies showing motor cortex influences on speech perception have also been criticized because of noise overlay of the auditory stimuli, which, apart from the use of 'artificial' phonological tasks, contributed to unnatural listening conditions. Though it can be argued that speech perception without noise is actually the exception rather than the norm in everyday life (D'Ausilio et al., 2012), it has been claimed that motor influences on speech perception might disappear when stimuli are not overlaid with noise (Rogalsky et al., 2011). Our results argue against this view by showing motor cortex influences on speech comprehension using non-synthesized stimuli without noise overlay. Although the reduction of sound pressure level (SPL) constitutes a decrease in signal-to-noise ratio (SNR), we see a major difference between our present approach and earlier studies which used noise to mask speech stimuli. By reducing SPL, the critical variable of change was the *degree of redundancy of the speech signal* (for discussion, see Wilson, 2009). For example, it is enough to hear part of the vowel [u] to identify the entire lexical item "pool", because co-articulation effects provide the listener with information about the preceding and subsequent phonemes (see, e.g., Warren and Marslen-Wilson, 1987). By minimizing redundancy through SNR reduction we observed a motor influence on the comprehension of spoken words not masked by noise, in line with the observation that earlier TMS studies (Fadiga et al., 2002; Watkins et al., 2003) and fMRI studies (Pulvermüller et al., 2006; Grabski et al., 2013) also found motor cortex activation in speech perception using stimuli without noise overlay. Furthermore, a recent study (Bartoli et al., 2015) showed that TMS to motor cortex causes interference in a speech discrimination task where task difficulty is induced by inter-speaker variability of naturally spoken syllables without noise.

Finally, our results also address two further objections that had been raised against earlier TMS studies investigating phoneme discrimination. First, it has been argued that motor cortex involvement in speech perception only occurs in tasks requiring phonemic segmentation (Sato et al., 2009), although the null effects in other tasks reported by that study could be due to differences in task complexity. Secondly, it is claimed that TMS to motor systems does not modulate perception of speech sounds but rather secondary decision processes such as response bias (Venezia et al., 2012). However, the fact that our task did not require segmenting or classifying phonemes argues against these views. Recognizing the word-initial sounds as either a bilabial or alveolar stop consonant was not the task per se but rather was implicit to understanding the whole spoken word and mapping it onto its meaning, as it is typical for natural language processing. Our results thus support that TMS to motor cortex indeed affects speech perception and comprehension (rather than any possible response bias or segmentation process). These conclusions are also in line with two recent studies which found that thetaburst TMS to motor cortex impairs syllable discrimination in a task unaffected by response bias (Rogers et al., 2014) and that rTMS to motor cortex affects sensitivity of speech discrimination, but not response bias (Smalle et al., 2015).

In summary, we here show a causal influence of articulatory motor cortex on the comprehension of meaningful words in the standard psycholinguistic task of word-to-picture matching. Furthermore, our results show that such effects can be obtained with naturally spoken stimuli without artificial noise overlay and are not due to response bias or other features epiphenomenal to the comprehension process. In the wider neuroscience debate about the frontal cortex' role in perceptual processing, the observed causal effect of motor cortex activation on language comprehension now demonstrates that the human motor system is not just activated in perception and comprehension, as previous research amply demonstrated, but that it also serves a critical role in the comprehension process itself. These results support an action-perception model of language (Pulvermüller and Fadiga, 2010) and are inconsistent with classic modular accounts attributing speech comprehension exclusively to temporal areas and denying a general causal contribution of motor systems to language comprehension (e.g., Wernicke, 1874; Lichtheim, 1885; Geschwind, 1970; Hickok, 2009, 2014).

# 5. Discussion

## Summary of results

### *Chapter 2*

Chapter 2 investigates which neurobiological mechanisms might be involved in representing word forms. Action-perception theory of language proposes that word forms are based on action-perception circuits (APCs) distributed across temporal and frontal cortices. Therefore, one specific question to test is whether the connectivity strength between these areas plays a role in establishing APCs for word forms. A neuroanatomically grounded computational model was used to investigate this question by comparing two model architectures with and without long-distance links between perisylvian language areas (reflecting the structural connectivity provided by the AF in humans) and comparing their emerging APCs in terms of activation dynamics. Results showed that auditory-articulatory APCs emerged in both model architectures, but that crucially, the APCs in the human model architecture with AF connectivity were much more robust. Specifically, the functional advantage provided by AF connectivity in the model was the emergence of reverberating activity in the APCs signifying verbal working memory (Fuster and Bressler, 2012). These findings underline a wealth of research indicating the importance of verbal working memory (VWM) for language learning and extend this existing research by showing that (i) VWM consists in reverberating activity between perisylvian language regions (consistent with evidence from behavioral and neuroimaging experiments), and (ii) that this reverberating activity in turn requires strong fronto-temporal connectivity. More strongly connected APCs might also imply stronger causality of frontal on temporal circuits parts in speech perception (which is the topic of Chapters 3 and 4).

### ***Chapter 3***

Chapter 3 addresses how speech perception and comprehension can be explained neurobiologically. Chapter 2 had already confirmed that frontal and motor systems as well as their connectivity is of key importance for *establishing* neuronal representations for word forms. Thus, Chapter 3 focuses on whether the resulting APCs distributed across perisylvian language regions play a functional role in speech recognition. While some brain models consider areas in temporal cortex as the locus of speech recognition (mapping sound to meaning), action-perception theory predicts that not only temporal areas, but equally speech motor areas in frontal cortex as well as inferior parietal areas should be activated by and play a functional role in speech processing. Chapter 3 reviews evidence from fMRI studies showing that speech perception causes activation across all perisylvian areas. A particular focus is on studies that were able to demonstrate that the observed activation carries specific phonological information, for example about place of articulation as well as other distinctive features. In recent years, the novel analysis of multivariate pattern analysis (MVPA) has become widespread, and several of the reviewed studies used this novel method to demonstrate phoneme-specific information in fronto-parietal areas during speech perception. The second part of Chapter 3 reviews studies which used TMS to demonstrate that frontal and inferior parietal areas also exert causal influences on speech processing. In sum, the findings reviewed in Chapter 3 confirm that phoneme-specific information is activated across all perisylvian areas in speech processing and that this activity can functionally contribute to speech processing.

### ***Chapter 4***

Chapter 4 addresses one specific issue implicated by the language mechanisms emerging from the neurocomputational modeling work in Chapter 2 and suggested by the literature review in Chapter 3, namely the functional causal role of frontal and motor mechanisms in language understanding, even at the basic level of comprehending single spoken words. The vast majority of studies investigating the role of frontoparietal areas in speech processing (reviewed in Chapter 3) used purely phonological tasks with meaningless syllables as stimuli and requiring phonological decision or identification tasks. However, there is an important distinction to be made between perception of meaningless phonemes and comprehension of meaningful words (see Introduction for details). Therefore, it was necessary to test whether TMS to articulatory motor regions also plays a causal role in speech comprehension of meaningful words, where phonemes serve their normal role as meaning-discriminating units. Chapter 4 now presents a TMS experiment which demonstrates for the first time that stimulation of articulatory motor cortex can functionally influence comprehension responses in the standard psycholinguistic test of speech comprehension, word-to-picture matching. Chapter 4 has been published as Schomers et al. (2015), and similar converging results have meanwhile also been reported by Murakami et al. (2015) (Figure 3.4; see Chapter 3 for further discussion).

## General discussion

The overall goal of the present dissertation was to investigate the neurobiological representation of word forms, which serve as the basis of meaning-carrying linguistic symbols. Specific questions were: How are spoken word forms represented neurobiologically? What neural mechanisms allow learning of novel word forms as well as comprehension of known words? Why do non-human primates lack these abilities? Which neuroanatomical features might underlie specifically human language abilities?

Action-perception theory provides an explanation in terms of distributed action-perception circuits (APCs) resulting from Hebbian learning. It was predicted that (i) strong fronto-temporal connectivity provided by the AF is a crucial part of the human brain mechanisms for verbal memory, and therefore important for word learning, and (ii) that speech comprehension is subserved by temporal, with inferior parietal areas and speech motor areas in inferior frontal cortex playing an additional causal role. These critical predictions, both of which could be confirmed, can be explained in terms of the same mechanism: strongly connected APCs interlinking phonological information about articulations and acoustic patterns of speech.

Returning to the discussion of the various neurobiological brain models of language presented in the Introduction, this dissertation demonstrates that the notion of a single, modular speech center in the brain must be wrong. Therefore, both the classic neurological models as well as more recent ones which posit, for example, a “gateway to understanding” in the superior temporal sulcus (Hickok, 2014) or a “lexical interface” (Hickok and Poeppel, 2004, 2007) in the posterior middle temporal gyrus (pMTG) (see Introduction for details) are clearly inconsistent with empirical evidence. To be fair, it is important to emphasize that Hickok and Poeppel do not view this purported interface as the ‘seat’ of meaning – they explicitly acknowledge that semantic representations are “widely distributed” (without going into detail about why and how exactly they might be widely distributed). But, importantly, they state that “middle posterior temporal regions are involved in the mapping between phonological representations in the STS and widely distributed semantic representations” (Hickok and Poeppel, 2007, p. 398). The findings of the present dissertation clearly speak against the idea that only one area in temporal cortex acts as a semantic binding site, by showing that a second locus, precentral cortex (along with inferior frontal cortex, as other work shows) contribute to semantic understanding. Furthermore, the neurocomputational and TMS studies show that not only semantic representations are distributed, but also the neuronal representations for word forms – that is, the action-perception circuits binding articulatory and acoustic-phonological information – are widely distributed across perisylvian regions.

In summary, the notion of a single binding site for mapping sound to meaning is not only theoretically unfounded (i.e., there is no compelling reason why there should be a single form-meaning interface, and why it should be located in the pMTG), but is also inconsistent with empirical evidence. Instead, the semantic link between sound and meaning appears to be mapped to a range of different brain areas. Word form representations are distributed across all perisylvian areas, whereas their corresponding semantic representations exhibit even wider distributions across the entire brain, including extrasylvian areas. These two sub-parts of an APC are bound together due to Hebbian learning, resulting in a strongly connected higher-order APC interlinking and representing both the word form and its meaning.

Following this summary of findings and general implications, I will now discuss how the results relate to some further specific questions of interest in the neuroscience of language.

### *The importance of working memory for building action-perception circuits for language*

In general, working memory is considered to play an important role in long-term memory formation (Ranganath et al., 2005; Blumenfeld and Ranganath, 2006). In line with this, a wealth of evidence has also demonstrated that long-term memory for word forms depends on verbal working memory abilities (Baddeley et al., 1988; Gathercole and Baddeley, 1989; Gathercole et al., 1997; Gathercole, 2006). Therefore, verbal working memory appears to be essential for establishing word form circuits, and Chapter 2 demonstrated that VWM ability is, in turn, subserved by strong fronto-temporal connectivity.

It is therefore the indirect link via VWM that provides a putative explanation for why the AF is an important prerequisite for learning word forms already at the purely phonological level (which was the focus of the work in Chapter 2). However, it is worth considering whether working memory might additionally be important for learning the meaning of words.

Words provide the basis for semantics, by serving as the symbols which can be associated with meaning. Therefore, the AF might not only be important for language because it enables working memory for word forms, but, additionally, strong verbal working memory might be of great importance for learning the meaning of new words which are used and experienced in immediate temporal congruency with the intended referent. When a new word is learnt there is activity both in perisylvian areas representing the word form and in extrasylvian areas representing aspects of the meaning of the word (Pulvermüller, 1999, 2013). In many cases this activity has a close temporal contiguity, for example when the meaning of a word is taught explicitly by establishing joint attention (Baldwin, 1995) and pointing to word referents simultaneously (e.g. "This is a dog"). However, there is evidence that infants effortlessly learn novel word meanings even when there is no temporal contiguity between word and referent (Akhtar and Tomasello, 1996), in particular in some cultures, where explicit teaching of word meanings through child-directed speech is rare (Schieffelin, 1985). In such instances, keeping in mind a novel phonological word form to bridge such a temporal delay is crucial for learning word meanings. Thus, the ability to keep the word form representation active for a longer time period might not only be important for learning the phonological word form; it might also promote learning the meaning of words, by providing a crucial 'bridge' for binding neuronal cell assemblies in the perisylvian and extrasylvian cortices, even without the input overlapping in time.

### *The relationship between working memory and speech comprehension*

The previous section presented arguments for why VWM is important for learning phonological word forms and possibly also for learning the meaning of words. However, a second focus of this dissertation relates to the neurobiological mechanisms involved in speech comprehension; since verbal working memory is so central to language learning, one may ask whether VWM is just an additional, 'optional' cognitive function, or whether it might also serve a more fundamental function in language, in particular during ‘normal’ speech comprehension.

This is an important question because the brain language models mentioned in the Introduction make competing claims about the function of VWM in language comprehension. In particular, Hickok and Poeppel (2007) argue that “speech recognition tasks involve lexical access processes, whereas speech perception tasks […] *instead* require processes that allow the listener to maintain sublexical representations in an active state” (p. 394; my italics). Thus, the implication is that VWM is involved *only* in speech perception tasks, whereas speech comprehension proceeds only along the ventral stream using “lexical access processes”. Furthermore, Hickok views VWM as an active and conscious process only: “tasks that require participants to consciously attend to phonemic information [...] may draw on motor resources (e.g., phonological working memory) that are not normally engaged in speech perception during auditory comprehension” (Hickok, 2012, p. 399).

Hickok does not deny a role of the dorsal stream in working memory, in fact, he explicitly acknowledges such a role (Hickok et al., 2003); thus, the denial of dorsal stream involvement in speech perception and comprehension fundamentally rests on the notion that VWM is not part of ‘normal’ speech processing. Therefore it is worth scrutinizing this viewpoint more carefully. It seems very difficult to see how language comprehension could function without any VWM. Note that Baddeley’s original notion of VWM addressed a form of rather ‘high-performance’ mnemonic activity, which was typically studied using challenging tasks, where sequences of up to seven different phonemes (or syllables, words) had to be kept in memory and, after a delay, reproduced precisely in the serial order in which they had been presented. Obviously, such high-performance VWM – requiring *active* articulatory rehearsal of multiple units in a fixed order after a delay – is a most surprising human skill which requires powerful and specific mechanism worth studying in much detail. However, it should not be ignored, that in language use, memory of linguistic units is constantly required, even though the burden on the memory mechanisms may sometimes be much reduced compared with the typical Baddeley task. Already in a word-to-picture matching task – the classic psycholinguistic task for assessing single word comprehension – the perceived word must be kept in working memory for a short time, while the subject tries to find the matching image. In other cases of real life language comprehension, meaningful words are not used in isolation, but unfold over time in the context of a whole sentence. Therefore, at the level of sentence processing, VWM is most definitely required for language comprehension, since in order to correctly understand the syntactic structure of a sentence, it is necessary to keep several words of the entire sentence in working memory for some time (Just and Carpenter, 1992; Lewis et al., 2006).

The notion that sentence comprehension requires VWM lies at the heart of theories proposing that VWM is the reason for activation of Broca's area in sentence comprehension (Smith and Jonides, 1999). Interestingly, Hickok is also among the authors proposing such a theory (Rogalsky et al., 2008; Rogalsky and Hickok, 2011). Therefore, one can argue that Hickok's proposal is logically inconsistent: he agrees that VWM recruits inferior frontal and motor areas (Hickok et al., 2003), and argues that both phonological tasks as well as sentence comprehension require VWM, but denies that motor regions play a role in language comprehension.

Of course, in addition to raising the issue of logical inconsistency, this dissertation also shows that this claim is simply empirically wrong, by showing that TMS stimulation to motor regions does affect speech comprehension in a word-to-picture matching task (see Chapter 3, and in particular, Chapter 4). In order to resolve this contradiction, Hickok and Poeppel's model would need to be modified and acknowledge that VWM plays a role in speech comprehension, and that comprehension must thereby involve the auditory dorsal stream in addition to the ventral stream.

However, the attempt to strictly delineate working memory from speech perception and comprehension processes might be ill-founded in the first place. Action-perception theory of language can provide a better explanatory framework; what are traditionally viewed as separate cognitive processes, working memory vs. speech comprehension, might rely to a considerable extent on the same neural substrates, namely action-perception circuits in perisylvian regions. From this perspective it makes little sense to strictly classify a given cognitive process or task as *either* language understanding *or* verbal working memory (similarly, one can argue that speech production and comprehension cannot clearly be delineated and need not be mutually exclusive).

The results of this dissertation support the idea that the processes of working memory and speech comprehension both rely on action-perception circuits as their neural basis. Note that in the computational modeling results from Chapter 2, passive presentation of the auditory component of a previously learnt pattern was sufficient for reverberating activity across auditory and motor areas. Crucially, the motor areas in the model, including premotor (PM) but also primary motor (M1) cortex – the areas most distant from the auditory area which receives input in the testing phase – were 'reactivated' (Figure 2.3), but only in the human architecture including the long-distance connectivity. Both VWM and language comprehension might thus be subserved by ignition and reverberation of phonological APCs, and the reverberating activity of such an APC necessarily includes some degree of 'reactivation' of speech motor regions.

Interestingly, correlations between individual differences in verbal working memory performance and speech motor system activations in passive speech perception have also been demonstrated experimentally (Szenkovits et al., 2012; Murakami et al., 2015). One possible explanation for this observed correlation would be that individuals with stronger AF connectivity (which also differs across individuals; López-Barroso et al., 2013) have stronger activation of motor systems due to greater AF strength, and that this leads to both stronger motor systems activation in passive perception/comprehension and stronger VWM abilities. A future project could directly test this hypothesis by investigating whether there is a correlation between AF strength and motor systems activation during speech processing.

### *The role of noisy/challenging listening conditions for motor involvement in speech comprehension*

The results from chapters 3 and 4 also provide a novel perspective on the role of noisy (or otherwise challenging) listening conditions for the involvement of the speech motor system in speech processing. As mentioned several times before, the idea that the motor system could play a causal and functional role in speech perception and comprehension has been met with considerable controversy. In 2009, two TMS studies provided strong support for this view, at least as far as speech perception goes (D'Ausilio et al., 2009; Möttönen and Watkins, 2009), which was immediately followed by intense debate in the field (e.g., Lotto et al., 2009; Scott et al., 2009; Wilson, 2009; McGettigan et al., 2010). Publication of the results of Chapter 4 (Schomers et al., 2015) has revived this debate to a certain extent (Glenberg, 2015; Hickok, 2015b; Kemmerer, 2015b; Rizzolatti and Sinigaglia, 2015; see Hickok, 2015a for responses). A large part of Hickok's contribution to these debates was an acknowledgment of motor influences in the face of empirical evidence on the one hand, but a 'downplaying' of their importance on the other hand (see also previous section). One reason for this is that TMS stimulation typically only leads to a very weak effect, often manifest only in slight changes in reaction time or accuracy (larger effects can be obtained using novel methods for stimulating several brain areas simultaneously with TMS; see section on *Limitations and outlook* for further discussion). As a result of this, some measures of increasing task difficulty need to be taken to avoid ceiling effects. The fact that TMS effects were weak and usually restricted to modulation of reaction times (as is typical for TMS studies), whereas neuroimaging experiments reported that motor systems contributions increase when the stimuli become *moderately* less intelligible (see Chapter 3 for detailed review), led Hickok to propose that motor systems *only* contribute to speech processing under particularly challenging listening conditions (Hickok et al., 2011; Rogalsky et al., 2011; Hickok, 2015b), presumably because challenging listening conditions lead to the recruitment of working memory mechanisms not normally involved in speech processing (cf. discussion of this topic in the previous section). For example, Hickok et al. (2011) proposed that the speech motor system can generate forward predictions in the context of a feedback control system where such predictions are necessary for guiding speech production (Guenther et al., 2006). Hickok therefore argues that, under such 'special' circumstances, this mechanism – evolved primarily for speech production – could be exapted to support speech perception, albeit emphasizing that this is only an 'exception' under certain special circumstances. As reviewed in Chapter 3, there is mounting evidence against this view. For one thing, the list of purported 'special circumstances' or 'exceptions' to a purely auditory view of speech processing has become so long that one can question whether they really constitute 'exceptions'. In addition, there is abundant evidence for motor systems activation during language processing in sparse-imaging fMRI experiments as well as in other noise-free tasks. Chapter 3 therefore proposes that the prevalent view about the role of noise might require slight revisions. Rather than being observed *only* in challenging listening conditions, it is proposed that motor contributions to speech processing are present even in noise-free listening conditions, that they increase *to some degree* with increasingly challenging listening conditions, but, with too strong noise overlay, actually disappear again (probably due to methodological reasons, as discussed in Chapter 3).

### *Implications for the evolution of language*

One of the core questions in the neuroscience of language concerns identifying the specific aspects of human brain anatomy and connectivity which enable human language and whether the absence of language abilities (at least ones comparable to those seen in humans) in non-human primate ancestors, such as chimpanzees, is due to anatomical or functional differences in the brain or due to other reasons. Some theories argue that chimpanzees could not develop language abilities because they simply lack the ability to produce these sounds, most importantly because they do not have a descended larynx. Obviously, if this is true, then non-human primates could not develop language (at least not based on the vast repertoire of speech sounds and their combinations as in humans) simply due to them lacking the anatomical prerequisites for producing such sounds. However, whether the theory of the uniquely human descended larynx is true is controversial (Fitch, 2000; Fitch and Reby, 2001). Independently of that, to what extent differences in brain anatomy and connectivity also play a role is an intriguing question. That is, even if non-human primates were in principle able to produce speech sounds like humans (as Fitch suggests), and assuming they would similarly engage in babbling as human infants do, would they be able to learn words and build auditory-motor action-perception circuits? The results of Chapter 2 suggest otherwise; rather, the results suggest that the human language faculty has neuroanatomical prerequisites in addition to requiring a suitable vocal tract apparatus. Interestingly, recent evidence suggests that speech production in humans is not only crucially reliant on the vocal tract apparatus, but that vocal motor control of the larynx also requires cortical connections between the laryngeal motor cortex and temporoparietal regions implicated in language (Simonyan, 2014; Hickok, 2016; Kumar et al., 2016b).

### *To what extent are the neural substrates of speech processing specific to language?*

As outlined in the introduction, there are certain similarities between action-perception theory of language and Liberman's motor theory of speech perception (MTSP). One important claim made by Liberman that does not seem to hold up in the face of neuroscience evidence, however, is the existence of a speech-specific phonetic module purported to constitute the innate link between perception and production and to be in charge of perceiving “intended phonetic gestures”.

The fMRI study by Pulvermüller et al. (2006) already provided evidence against this modular view, by demonstrating that the same areas of motor cortex controlling non-linguistic movements of articulators such as lips and tongue are also activated during linguistic perception (i.e. perceiving syllables). Moreover, TMS evidence (D’Ausilio et al., 2009, and the TMS experiment presented in Chapter 4) succeeded in demonstrating that stimulation of those same sections of motor cortex involved in non-linguistic movements also exerts a causal influence on speech perception and comprehension. Thus, the available evidence supports Liberman’s claim that speech perception involves access to the speech motor system (cf. Galantucci et al., 2006), but it does not support the claim that “speech is special” and subserved by a dedicated phonetic module.

Similarly, regarding VWM, one can ask whether this is subserved by a dedicated module, as was argued in the original proposal by Baddeley and Hitch who posited a dedicated “phonological loop” specifically involved in phonological working memory. Again, the results of this dissertation argue against this idea, but support the theory that VWM is based on reverberating activity between auditory and motor regions. The fronto-temporal connectivity provided by the AF might prove to be particularly suitable for the reverberating activity underlying phonological working memory in particular, but evidence indicates it is not speech- or language-specific, as it has also been shown to be involved in auditory working memory for music (Koelsch et al., 2009). Some authors suggest that the crucial factor is repeatability, or producibility. For example, auditory recognition memory is much better for words that are producible than reversed words (Schulze et al., 2012). Kumar et al. (2016a) found that working memory of a non-speech tone which was in a frequency range that was producible by subjects also led to functional connectivity between frontal and auditory cortices. Similarly, when perceiving sounds which have a native-like phonology (and are hence producible), the activation of motor systems in speech perception is stronger and auditory-motor functional connectivity is stronger (Wilson and Iacoboni, 2006; Londei et al., 2010; Swaminathan et al., 2013). As such, the AF might benefit activation of motor systems (and VWM) specifically for any auditory item that is producible or repeatable by humans. Speech is one prime example for this, and the fact that speech stimuli can easily be repeated is one of the essential properties of language which make it so useful. Thus, the AF provides a function particularly suitable to speech, but probably not restricted to it.

In sum, the findings of this dissertation do not warrant claims for any modules dedicated specifically to phonological processing or phonological working memory. Rather, they are in line with the general framework of action-perception theory (Pulvermüller and Fadiga, 2010) as well as embodied cognition views (Barsalou, 1999, 2008; Glenberg and Kaschak, 2002) in that they show that cognitive functions such as speech processing and verbal working memory can rely on domain-general sensory and motor areas and do not require any specialized 'modules' or domain-specific mechanisms.

## Limitations and outlook

Before concluding, I would like to discuss limitations of the studies presented in this dissertation as well as open questions, with an outlook on possible future research projects.

Regarding Chapter 2, it is important to stress that this was intended as an initial investigation of the role of long-distance fronto-temporal connectivity for the emergence of APCs. One obvious limitation is that the model does not include any parietal lobe regions yet. However, as Chapter 3 shows, there is increasing evidence that these regions also play an important role in speech perception, even though they are less well studied than temporal, inferior frontal and motor regions. Future modelling work could additionally include inferior parietal areas; as a result of this, it would also be possible to model the influence of the long direct segment of the AF vs. the two indirect segments of the AF (Catani et al., 2005).

However, one should note that López-Barroso et al. (2013) had found a correlation between structural connectivity and word learning abilities only for the long direct segment and not for the indirect segments of the AF. Therefore, the long direct segment is likely to be the most important structural connectivity prerequisite for human language abilities, although this requires further investigation.

Furthermore, the established models of verbal working memory posit that verbal working memory functions through articulatory subvocal rehearsal (D'Esposito, 2007), which is in line with a wealth of evidence showing the involvement of speech production systems in VWM (see Chapter 2). Therefore, the fact that the modelling results presented here also evidenced reactivation of primary motor and premotor areas supports this view, as well as supporting the theory that VWM does not require a specific 'module' for VWM but relies instead on the same brain areas involved in speech production (see previous section for further discussion on implications for modularity). However, one could also test whether lesioning areas M1 and PM in the model leads to abolishment or significant impairment of the reverberating activity signifying VWM, in order to test this hypothesis more explicitly.

A further limitation of Chapter 2 that should be noted is that long-distance connectivity is probably one of the key features, possibly *the* key feature of human neuroanatomical connectivity enabling word learning; however, one should emphasize that there are probably many other cognitive factors which also mediate word learning, and as a result of this, other brain areas also play important additional roles in word learning, such as the hippocampus (Breitenstein et al., 2005; Sederberg et al., 2007) and the amygdala (Ripollés et al., 2014). Finally, learning novel phonological word forms also significantly benefits from existing semantic associations (Hawkins et al., 2015; Savill et al., 2015b, 2016), an aspect not modeled so far, since the focus was purely on the phonological aspects. However, the model architecture used here has been extended to model semantic associations by including extrasylvian areas as well (Garagnani and Pulvermüller, 2016; Tomasello et al., 2016), and further research on semantic aspects of word learning will be conducted using this model in the future.

A further important question not yet addressed by the computational model is the total number of distinct ‘words’ that can be learned by the different architectures. Given that there are a few reports in the literature of chimpanzees learning a small repertoire of words, albeit only with extremely long periods of training, one could surmise that AF connectivity affects not only the overall robustness of cell assemblies and reverberatory activity, but also the total number of words that can be learnt and memorized. This could also be investigated in a future computational modeling project.

Regarding the involvement of motor areas in frontal cortex in speech processing, most previous fMRI and TMS studies (as reviewed in Chapter 3) have focused on primary motor and premotor cortex as well as somatosensory and inferior parietal cortex. However, as pointed out at the end of Chapter 3, it has been suggested that other motor areas which have been known to contribute to speech production – such as the cerebellum and the supplementary motor area – also play roles in speech perception (Ackermann et al., 2007; Hertrich et al., 2016). A major open question concerns the precise role of the different motor areas in both speech production and perception. In particular, the exact interplay between inferior frontal areas, premotor areas, primary motor cortex and other motor areas remains to be clarified.

Regarding the experiment described in Chapter 4 (as well as Murakami et al., 2015), it should be noted that using TMS, these results demonstrated that articulatory motor cortex plays a causal role in speech comprehension; this had long been doubted due to the use of explicit phonological identification/discrimination tasks in previous work. The results of Chapter 4 allow to rule out the hypothesis that motor cortex plays absolutely no role in speech comprehension, as proposed by some models. However, since no effect on accuracy was observed (despite there being a non-significant trend), it is unclear whether motor cortex is necessary for speech comprehension. One reason why the experiment in Chapter 4 did not produce effects on accuracy is because a single-pulse TMS paradigm was used, which does not produce a sustained ‘virtual lesion’ and is therefore not expected to lead to a decline in accuracy. In contrast, repetitive TMS, in particular thetaburst stimulation (Huang et al., 2005) might produce stronger effects, affecting accuracy instead of (or in addition to) reaction times. For example, in Rogers et al. (2014), thetaburst TMS to the lip area of primary motor cortex led to a large and significant reduction of accuracy in a phoneme discrimination task. However, when it comes to speech comprehension, similarly large effects might be harder to obtain, likely due to the fact that normal speech is characterized by a high degree of redundancy (see Chapters 3 and 4). In particular, two recent TMS studies demonstrated that repetitive TMS of a single cortical site does not lead to a measurable effect on speech comprehension, whereas stimulating two areas does (Murakami et al., 2015; Hartwigsen et al., 2016). Therefore, an open question is whether further experiments using a double-knockout thetaburst approach similar to that of Murakami et al. (2015) can produce decreases in accuracy in a word comprehension task.

## Conclusion

In conclusion, the findings of this dissertation lend support to the action-perception theory of language. The neuronal representation of word forms can be explained through action-perception circuits (APCs) which emerge based on Hebbian learning principles and long-distance connectivity in perisylvian brain areas. These APCs provide a basis for both verbal working memory as well as meaning-carrying linguistic symbols (word forms). Therefore, they also take a functional role in speech recognition, i.e. mapping sound to meaning. Because of the distributed nature of APCs, action-perception theory predicts that there cannot be a single sound-to-meaning interface, but that, instead, speech comprehension is subserved by many cortical loci across the entire perisylvian language regions, as supported by a wealth of empirical evidence presented in this dissertation. Finally, this dissertation also underlines the central role which verbal working memory plays for language abilities, not only in establishing APCs neural representations of words, but also in language comprehension.

# References

Acheson, D. J., Hamidi, M., Binder, J. R., and Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *J. Cogn. Neurosci.* 23, 1358–1367.

Ackermann, H., Mathiak, K., and Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *The Cerebellum* 6, 202–213.

Adank, P., Davis, M. H., and Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia* 50, 77–84.

Adank, P., Hagoort, P., and Bekkering, H. (2010). Imitation improves language comprehension. *Psychol. Sci.* 21, 1903–1909. doi:10.1177/0956797610389192.

Adank, P. M., Rueschemeyer, S.-A., and Bekkering, H. (2013). The role of accent imitation in sensorimotor integration during processing of intelligible speech. *Front. Hum. Neurosci.* 7, 634.

Akama, H., Murphy, B., Na, L., Shimizu, Y., and Poesio, M. (2012). Decoding semantics across fMRI sessions with different stimulus modalities: a practical MVPA study. *Front Neuroinform* 6, 24.

Akhtar, N., and Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *Br. J. Dev. Psychol.* 14, 79–93. doi:10.1111/j.2044-835X.1996.tb00695.x.

Alho, J., Green, B. M., May, P. J., Sams, M., Tiitinen, H., Rauschecker, J. P., et al. (2016). Early-latency categorical speech sound representations in the left inferior frontal gyrus. *NeuroImage* 129, 214–223.

Alho, J., Lin, F.-H., Sato, M., Tiitinen, H., Sams, M., and Jääskeläinen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Front. Psychol.* 5, 394.

Amir, Y., Harel, M., and Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J. Comp. Neurol.* 334, 19–46.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266.

Anderson, M. L. (2015). Précis of after phrenology: neural reuse and the interactive brain. *Behav. Brain Sci.* 16, 1–22.

Andoh, J., and Martinot, J.-L. (2008). Interhemispheric compensation: A hypothesis of TMS-induced effects on language-related areas. *Eur. Psychiatry* 23, 281–288. doi:10.1016/j.eurpsy.2007.10.012.

Arbib, M. A. (2016). Towards a computational comparative neuroprimatology: framing the language-ready brain. *Phys. Life Rev.* 16, 1–54.

Arsenault, J. S., and Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *J. Neurosci.* 35, 634–642. doi:10.1523/JNEUROSCI.2454-14.2015.

Arsenault, J. S., and Buchsbaum, B. R. (2016). No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychon. Bull. Rev.* 23, 1231–1240. doi:10.3758/s13423-015-0988-z.

Artola, A., Bröcher, S., and Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* 347, 69–72.

Artola, A., and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* 16, 480–487.

Babiloni, F., Carducci, F., Cincotti, F., Del Gratta, C., Pizzella, V., Romani, G. L., et al. (2001). Linear inverse source estimate of combined EEG and MEG data related to voluntary movements. *Hum. Brain Mapp.* 14, 197–209.

Baddeley, A. (1993). Short-term phonological memory and long-term learning: A single case study. *Eur. J. Cogn. Psychol.* 5, 129–148.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839.

Baddeley, A., Papagno, C., and Vallar, G. (1988). When long-term learning depends on short-term storage. *J. Mem. Lang.* 27, 586–595.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., and Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease–dementia–aphasia syndrome. *Brain* 124, 103–120.

Baldwin, D. A. (1995). "Understanding the link between joint attention and language," in *C. Moore & P. J. Dunham (Eds.). Joint attention: Its origins and role in development* (Hillsdale, NJ: Lawrence Erlbaum), 131–158.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2008). Grounded Cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639.

Bartoli, E., D'Ausilio, A., Berry, J., Badino, L., Bever, T., and Fadiga, L. (2015). Listener–speaker perceived distance predicts the degree of motor contribution to speech perception. *Cereb. Cortex* 25, 281–288. doi:10.1093/cercor/bht257.

Bartoli, E., Maffongelli, L., Campus, C., and D'Ausilio, A. (2016). Beta rhythm modulation by speech sounds: somatotopic mapping in somatosensory cortex. *Sci. Rep.* 6, 31182. doi:10.1038/srep31182.

Basso, A., Casati, G., and Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex* 13, 85–95.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. http://CRAN.R-project.org/package=lme4.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion–symptom mapping. *Nat. Neurosci.* 6, 448–450.

Belleville, S., Peretz, I., and Arguin, M. (1992). Contribution of articulatory rehearsal to short-term memory: Evidence from a case of selective disruption. *Brain Lang.* 43, 713–746.

Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., and Geschwind, N. (1973). Conduction aphasia: a clinicopathological study. *Arch. Neurol.* 28, 339–346.

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang.* 78, 364–396. doi:10.1006/brln.2001.2484.

Bibbig, A., Wennekers, T., and Palm, G. (1995). A neural network model of the cortico-hippocampal interplay and the representation of contexts. *Behav. Brain Res.* 66, 169–175.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301.

Bishop, D. V. M., Brown, B. B., and Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *J. Speech Lang. Hear. Res.* 33, 210–219.

Blumenfeld, R. S., and Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *J. Neurosci.* 26, 916–925.

Blumstein, S. E., Baker, E., and Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia* 15, 19–30.

Blumstein, S. E., Burton, M., Baum, S., Waldstein, R., and Katz, D. (1994). The role of lexical status on the phonetic categorization of speech in aphasia. *Brain Lang.* 46, 181–197.

Bogen, J. E., and Bogen, G. M. (1976). Wernicke's Region–where is it? *Ann. N. Y. Acad. Sci.* 280, 834–843.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.

de Borst, A. W., and de Gelder, B. (2016). fMRI-based multivariate pattern analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and auditory cortices. *Cereb. Cortex.* doi:10.1093/cercor/bhw211.

Bouchard, K. E., and Chang, E. F. (2014). Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J. Neurosci.* 34, 12662–12677.

Bouchard, K. E., Mesgarani, N., Johnson, K., and Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327–332.

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., and Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia* 46, 743–756.

Braitenberg, V. (1978). "Cell assemblies in the cerebral cortex," in *Theoretical approaches to complex systems* (Springer), 171–188.

Braitenberg, V., and Pulvermüller, F. (1992). Entwurf einer neurologischen Theorie der Sprache. *Naturwissenschaften* 79, 103–117.

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity*. Berlin: Springer.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* 25, 958–968.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d’une observation d’aphémie (perte de la parole). *Bull. Société D’anatomie* 36, 330–57.

Brysbaert, M., Stevens, M., Mandera, P., and Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant’s age. *Front. Psychol.* 7, 1116. doi:10.3389/fpsyg.2016.01116.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D’Esposito, M., et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119, 119–128.

Buchsbaum, B. R., and D’Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778.

Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697.

Call, J. E., and Tomasello, M. E. (2007). *The gestural communication of apes and monkeys.* Taylor & Francis Group/Lawrence Erlbaum Associates.

Camalier, C. R., D’Angelo, W. R., Sterbing-D’Angelo, S. J., de la Mothe, L. A., and Hackett, T. A. (2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18168–18173. doi:10.1073/pnas.1206387109.

Caplan, D., Gow, D., and Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology* 45, 293–298.

Carota, F., Moseley, R., and Pulvermüller, F. (2012). Body-part-specific representations of semantic noun categories. *J. Cogn. Neurosci.* 24, 1492–1509.

van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589.

Catani, M., Allin, M. P., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., et al. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci.* 104, 17163–17168.

Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

Catford, J. C., and Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *Mod. Lang. J.* 54, 477–481.

Cheung, C., Hamilton, L. S., Johnson, K., and Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife* 5, e12577.

Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* 33, 5208–5215. doi:10.1523/JNEUROSCI.1870-12.2013.

Cichy, R. M., Heinzle, J., and Haynes, J.-D. (2012). Imagery and perception share cortical representations of content and location. *Cereb. Cortex* 22, 372–380.

Clark, D., and Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia* 41, 304–317. doi:10.1016/S0028-3932(02)00163-X.

Correia, J. M., Jansma, B. M., and Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* 35, 15015–15025.

Damasio, H., and Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain* 103, 337–350.

D’Ausilio, A., Bufalari, I., Salmas, P., and Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex* 48, 882–887.

D’Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., et al. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130418. doi:10.1098/rstb.2013.0418.

D’Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.

Davachi, L., Maril, A., and Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. *J. Cogn. Neurosci.* 13, 1059–1070.

D’Esposito, M. (2007). From cognitive to neural models of working memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 761–772.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. USA* 109, E505–E514.

Dick, A. S., Bernal, B., and Tremblay, P. (2014). The language connectome new pathways, new concepts. *The Neuroscientist* 20, 453–467.

Dick, A. S., and Tremblay, P. (2012). Beyond the arcuate fasciculus: Consensus and controversy in the connectional anatomy of language. *Brain* 135, 3529–3550.

Diesch, E., Eulitz, C., Hampson, S., and Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain Lang.* 53, 143–168.

Douglas, R. J., and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27, 419–451.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M.-T., and Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain* 130, 1432–1441.

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci.* 111, 7126–7131.

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* 7, 12241.

Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Hüppi, P. S., and Hertz-Pannier, L. (2014). The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. *Neuroscience* 276, 48–71. doi:10.1016/j.neuroscience.2013.12.044.

Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J. F., Bihan, D. L., and Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensori-motor networks. *Cereb. Cortex* 19, 414–423. doi:10.1093/cercor/bhn097.

Duncan, J. (2006). EPS Mid-Career Award 2004: brain mechanisms of attention. *Q. J. Exp. Psychol.* 59, 2–27.

Eggert, J., and Van Hemmen, J. L. (2000). Unifying framework for neuronal assembly dynamics. *Phys. Rev. E* 61, 1855.

Epstein, C. M., Meador, K. J., Loring, D. W., Wright, R. J., Weissman, J. D., Sheppard, S., et al. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clin. Neurophysiol.* 110, 1073–1079.

Etzel, J. A., Gazzola, V., and Keysers, C. (2008). Testing simulation theory with cross-modal multivariate classification of fMRI data. *PLoS ONE* 3, e3690. doi:10.1371/journal.pone.0003690.

Evans, S., and Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cereb. Cortex* 25, 4772–4788. doi:10.1093/cercor/bhv136.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267.

Fitch, W. T., and Reby, D. (2001). The descended larynx is not uniquely human. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1669–1675.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., et al. (2015). Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci.* 112, 2871–2875.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). "Who" is saying "What"? Brain-based decoding of human voice and speech. *Science* 322, 970–973.

Fridriksson, J., Bonilha, L., and Rorden, C. (2007). Severe Broca's aphasia without Broca's area damage. *Behav. Neurol.* 18, 237–238.

Fridriksson, J., Fillmore, P., Guo, D., and Rorden, C. (2015). Chronic Broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cereb. Cortex* 25, 4689–4696.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78.

Fritz, J., Mishkin, M., and Saunders, R. C. (2005). In search of an auditory engram. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9359–9364. doi:10.1073/pnas.0503998102.

Fry, D. B. (1966). "The development of the phonological system in the normal and deaf child," in *Smith, F., and Miller, G.A. (Eds). The genesis of language.* (MIT Press), 187–206.

Fuster, J. M. (2003). *Cortex and mind: Unifying cognition.* Oxford University Press.

Fuster, J. M. (2006). The cognit: a network model of cortical representation. *Int. J. Psychophysiol.* 60, 125–132.

Fuster, J. M., and Bressler, S. L. (2012). Cognit activation: a mechanism enabling temporal integration in working memory. *Trends Cogn. Sci.* 16, 207–218.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.

Garagnani, M., and Pulvermüller, F. (2013). Neuronal correlates of decisions to speak and act: Spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas. *Brain Lang.* 127, 75–85. doi:10.1016/j.bandl.2013.02.001.

Garagnani, M., and Pulvermüller, F. (2016). Conceptual grounding of language in action and perception: a neurocomputational model of the emergence of category specificity and semantic hubs. *Eur. J. Neurosci.* 43, 721–737. doi:10.1111/ejn.13145.

Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention–language interactions in the human brain. *Eur. J. Neurosci.* 27, 492–513.

Garagnani, M., Wennekers, T., and Pulvermüller, F. (2009). Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network. *Cogn. Comput.* 1, 160–176.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Appl. Psycholinguist.* 27, 513–543. doi:10.1017/S0142716406060383.

Gathercole, S. E., and Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *J. Mem. Lang.* 28, 200–213.

Gathercole, S. E., Hitch, G. J., Martin, A. J., and others (1997). Phonological short-term memory and new word learning in children. *Dev. Psychol.* 33, 966.

Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.

Gierhan, S. M. (2013). Connections for auditory language in the human brain. *Brain Lang.* 127, 205–221.

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3, 1116–1133.

Glenberg, A. M. (2015). Big Myth or Major Miss? Gregory Hickok's "The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition." *Am. J. Psychol.* 128, 533–539. doi:10.5406/amerjpsyc.128.4.0533.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016.

Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., et al. (2013). Shared and distinct neural correlates of vowel perception and production. *J. Neurolinguistics* 26, 384–408. doi:10.1016/j.jneuroling.2012.11.003.

Graziano, M. S. (2016). Ethological action maps: a paradigm shift for the motor cortex. *Trends Cogn. Sci.* 20, 121–132.

Graziano, M. S., Taylor, C. S., and Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34, 841–851.

Grisoni, L., Dreyer, F. R., and Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cereb. Cortex* 26, 2353–2366.

Guellaï, B., Streri, A., and Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: some critical questions. *Front. Psychol.* 5, 812.

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

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). Sparse temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010a). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci.* 107, 16494–16499.

Hartwigsen, G., Price, C. J., Baumgaertner, A., Geiss, G., Koehnke, M., Ulmer, S., et al. (2010b). The right posterior inferior frontal gyrus contributes to phonological word decisions in the healthy brain: Evidence from dual-site TMS. *Neuropsychologia* 48, 3155–3163. doi:10.1016/j.neuropsychologia.2010.06.032.

Hartwigsen, G., Saur, D., Price, C. J., Ulmer, S., Baumgaertner, A., and Siebner, H. R. (2013). Perturbation of the left inferior frontal gyrus triggers adaptive plasticity in the right homologous area during speech production. *Proc. Natl. Acad. Sci.* 110, 16402–16407. doi:10.1073/pnas.1310190110.

Hartwigsen, G., Weigel, A., Schuschan, P., Siebner, H. R., Weise, D., Classen, J., et al. (2016). Dissociating parieto-frontal networks for phonological and semantic word decisions: a condition-and-perturb TMS study. *Cereb. Cortex* 26, 2590–2601. doi:10.1093/cercor/bhv092.

Hauk, O., Shtyrov, Y., and Pulvermüller, F. (2006). The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory–motor networks by sounds associated with finger and tongue movements. *Eur. J. Neurosci.* 23, 811–821.

Hawkins, E., Astle, D. E., and Rastle, K. (2015). Semantic Advantage for Learning New Phonological Form Representations. *J. Cogn. Neurosci.* 27, 775–786. doi:10.1162/jocn_a_00730.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.

Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* 87, 257–270.

Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., et al. (2011). dlexDB–eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychol. Rundsch.* 62, 10–20.

Hertrich, I., Dietrich, S., and Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neurosci. Biobehav. Rev.* 68, 602–610.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., and Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang. Cogn. Process.* 27, 1145–1166.

van den Heuvel, M. P., Bullmore, E. T., and Sporns, O. (2016). Comparative connectomics. *Trends Cogn. Sci.* 20, 345–361.

van den Heuvel, M. P., and Sporns, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243.

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *J. Commun. Disord.* 45, 393–402. doi:10.1016/j.jcomdis.2012.06.004.

Hickok, G. (2014). *The myth of mirror neurons: the real neuroscience of communication and cognition*. WW Norton & Company.

Hickok, G. (2015a). Response to Rizzolatti and Sinigaglia and to Glenberg. *Am. J. Psychol.* 128, 539–549.

Hickok, G. (2015b). The motor system’s contribution to perception and understanding actions: clarifying mirror neuron myths and misunderstandings. *Lang. Cogn.*, 1–9.

Hickok, G. (2016). A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychon. Bull. Rev.*, 1–8.

Hickok, G., Buchsbaum, B. R., Humphries, C., and Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Cogn. Neurosci. J. Of* 15, 673–682.

Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Huang, J., Carr, T. H., and Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Hum. Brain Mapp.* 15, 39–53. doi:10.1002/hbm.1060.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron* 45, 201–206.

Huyck, C. R., and Passmore, P. J. (2013). A review of cell assemblies. *Biol. Cybern.* 107, 263–288.

Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci.* 106, 1245–1248.

Jacquemot, C., and Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10, 480–486.

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112.

Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci.* 97, 11793–11799.

Kaiser, M., and Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput Biol* 2, e95.

Kakei, S., Hoffman, D. S., and Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science* 285, 2136–2139.

Kammer, T., and Baumann, L. W. (2010). Phosphene thresholds evoked with single and double TMS pulses. *Clin. Neurophysiol.* 121, 376–379.

Kandel, E. R., Schwartz, J. H., and Jessel, T. M. (2000). *Principles of Neural Sciences*. New York: McGraw-Hill.

Kaplan, J. T., Man, K., and Greening, S. G. (2015). Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Front. Hum. Neurosci.* 9, 151. doi:10.3389/fnhum.2015.00151.

Karabanov, A. N., Paine, R., Chao, C. C., Schulze, K., Scott, B., Hallett, M., et al. (2015). Participation of the classical speech areas in auditory long-term memory. *PloS One* 10, e0119472.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J. Acoust. Soc. Am.* 138, 817–832.

Kemmerer, D. (2015a). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychon. Bull. Rev.* 22, 1068–1075.

Kemmerer, D. (2015b). Does the motor system contribute to the perception and understanding of actions? Reflections on Gregory Hickok's The myth of mirror neurons: The real neuroscience of communication and cognition. *Lang. Cogn.* 7, 450–475.

Kemmerer, D., Rudrauf, D., Manzel, K., and Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* 48, 826–848.

Kenward, M. G., and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53, 983–997.

Kimppa, L., Kujala, T., Leminen, A., Vainio, M., and Shtyrov, Y. (2015). Rapid and automatic speech-specific learning mechanism in human neocortex. *NeuroImage* 118, 282–291. doi:10.1016/j.neuroimage.2015.05.098.

Kimppa, L., Kujala, T., and Shtyrov, Y. (2016). Individual language experience modulates rapid formation of cortical memory circuits for novel words. *Sci. Rep.* 6, 30227. doi:10.1038/srep30227.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Hum. Brain Mapp.* 30, 859–873.

Komeilipoor, N., Vicario, C. M., Daffertshofer, A., and Cesari, P. (2014). Talking hands: tongue motor excitability during observation of hand gestures associated with words. *Front. Hum. Neurosci.* 8, 767. doi:10.3389/fnhum.2014.00767.

Kornhuber, H. H., and Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüg. Arch. Für Gesamte Physiol. Menschen Tiere* 284, 1–17.

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., and Jefferies, B. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *J. Cogn. Neurosci.* 25, 2179–2188.

Kriegeskorte, N., and Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412.

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4.

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., and Griffiths, T. D. (2016a). A brain system for auditory working memory. *J. Neurosci.* 36, 4492–4505. doi:10.1523/JNEUROSCI.4341-14.2016.

Kumar, V., Croxson, P. L., and Simonyan, K. (2016b). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *J. Neurosci.* 36, 4170–4181.

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., and Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *J. Neurosci.* 34, 10339–10346.

Lee, Y. S., Turkeltaub, P., Granger, R., and Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32, 3942–3948.

Leonard, M. K., Cai, R., Babiak, M. C., Ren, A., and Chang, E. F. (2016). The peri-Sylvian cortical network underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. *Brain Lang.* doi:10.1016/j.bandl.2016.06.001.

Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454.

Liao, D. A., Kronemer, S. I., Yau, J. M., Desmond, J. E., and Marvel, C. L. (2014). Motor system contributions to verbal and non-verbal working memory. *Front. Hum. Neurosci.* 8, 753. doi:10.3389/fnhum.2014.00753.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431.

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.

Liberman, A. M., and Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn. Sci.* 4, 187–196.

Lichtheim, L. (1885). On aphasia. *Brain* 7, 433–484.

Liebenthal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., and Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *J. Neurosci.* 33, 15414–15424. doi:10.1523/JNEUROSCI.1511-13.2013.

Locke, J. L. (1995). *The child's path to spoken language*. Harvard University Press.

Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. D., Romani, G.-L., et al. (2010). Sensory-motor brain network connectivity for speech comprehension. *Hum. Brain Mapp.* 31, 567–580. doi:10.1002/hbm.20888.

Longcamp, M., Anton, J.-L., Roth, M., and Velay, J.-L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage* 19, 1492–1500.

López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A., and de Diego-Balaguer, R. (2013). Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci.* 110, 13168–13173.

López-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T. F., and Rodríguez-Fornells, A. (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cereb. Cortex* 21, 2742–2750.

Lotto, A. J., Hickok, G. S., and Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* 13, 110–114. doi:10.1016/j.tics.2008.11.008.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Malenka, R. C., and Bear, M. F. (2004). LTP and LTD: an embarrassment of riches. *Neuron* 44, 5–21.

Markiewicz, C. J., and Bohland, J. W. (2016). Mapping the cortical representation of speech sounds in a syllable repetition task. *NeuroImage* 141, 174–190.

Matsumoto, R., Nair, D. R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., et al. (2004). Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain* 127, 2316–2330.

McGettigan, C., Agnew, Z. K., and Scott, S. K. (2010). Are articulatory commands automatically and involuntarily activated during speech perception? *Proc. Natl. Acad. Sci. USA* 107, E42–E42.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi:10.1038/264746a0.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696. doi:10.1016/j.cub.2007.08.064.

Mesulam, M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* 28, 597–613.

Miceli, G., Gainotti, G., Caltagirone, C., and Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain Lang.* 11, 159–169.

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., and Davis, K. R. (1978). Broca aphasia Pathologic and clinical. *Neurology* 28, 311–311.

Moineau, S., Dronkers, N. F., and Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *J. Speech Lang. Hear. Res.* 48, 884.

Moliadze, V., Zhao, Y., Eysel, U., and Funke, K. (2003). Effect of transcranial magnetic stimulation on single-unit activity in the cat primary visual cortex. *J. Physiol.* 553, 665–679.

Mottaghy, F. M., Sparing, R., and Töpper, R. (2006). Enhancing picture naming with transcranial magnetic stimulation. *Behav. Neurol.* 17, 177–186.

Möttönen, R., Dutton, R., and Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cereb. Cortex* 23, 1190–1197. doi:10.1093/cercor/bhs110.

Möttönen, R., Järveläinen, J., Sams, M., and Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *NeuroImage* 24, 731–737. doi:10.1016/j.neuroimage.2004.10.011.

Möttönen, R., van de Ven, G. M., and Watkins, K. E. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *J. Neurosci.* 34, 4064–4069.

Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825.

Möttönen, R., and Watkins, K. E. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology* 26, 1103–1118.

Murakami, T., Kell, C. A., Restle, J., Ugawa, Y., and Ziemann, U. (2015). Left dorsal speech stream components and their contribution to phonological processing. *J. Neurosci.* 35, 1411–1422.

Murakami, T., Restle, J., and Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia* 49, 2045–2054. doi:10.1016/j.neuropsychologia.2011.03.034.

Murakami, T., Ugawa, Y., and Ziemann, U. (2013). Utility of TMS to understand the neurobiology of speech. *Front. Psychol.* 4, 446.

Myers, E. B., Blumstein, S. E., Walsh, E., and Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.* 20, 895–903.

Näätänen, R., Lehtokoski, A., Lennest, M., Luuki, A., Alliki, J., Sinkkonen, J., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.

Nasir, S. M., and Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci.* 106, 20470–20475.

Neininger, B., and Pulvermüller, F. (2001). The Right Hemisphere's Role in Action Word Processing: a Double Case Study. *Neurocase* 7, 303–317.

Neininger, B., and Pulvermüller, F. (2003). Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia* 41, 53–70.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., and Passingham, R. (2004). The inferior frontal gyrus and phonological processing: an investigation using rTMS. *J. Cogn. Neurosci.* 16, 289–300.

Niyazov, D. M., Butler, A. J., Kadah, Y. M., Epstein, C. M., and Hu, X. P. (2005). Functional magnetic resonance imaging and transcranial magnetic stimulation: effects of motor imagery, movement and coil orientation. *Clin. Neurophysiol.* 116, 1601–1610.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430.

Nourski, K. V., Steinschneider, M., McMurray, B., Kovach, C. K., Oya, H., Kawasaki, H., et al. (2014). Functional organization of human auditory cortex: Investigation of response latencies through direct recordings. *Neuroimage* 101, 598–609.

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., and Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage* 128, 218–226. doi:10.1016/j.neuroimage.2015.12.038.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Hum. Brain Mapp.* 27, 562–571.

Obleser, J., and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19.

Obleser, J., Lahiri, A., and Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *Neuroimage* 20, 1839–1847.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Oosterhof, N. N., Tipper, S. P., and Downing, P. E. (2012). Visuo-motor imagery of specific manual actions: A multi-variate pattern analysis fMRI study. *NeuroImage* 63, 262–271.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends Cogn. Sci.* 2, 455–462.

Osnes, B., Hugdahl, K., and Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage* 54, 2437–2445.

Palm, G. (1982). *Neural assemblies*. Berlin: Springer.

Pandya, D. N. (1994). Anatomy of the auditory cortex. *Rev. Neurol. (Paris)* 151, 486–494.

Pandya, D. N., and Yeterian, E. H. (1985). "Architecture and connections of cortical association areas," in *Peters, A. & Jones, E.G. (Eds). Association and auditory cortices* (Springer), 3–61.

Papagno, C., and Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Q. J. Exp. Psychol.* 48, 98–107.

Park, H., Kayser, C., Thut, G., and Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife* 5, e14521. doi:10.7554/eLife.14521.

Pascual-Leone, A., Gates, J. R., and Dhuna, A. (1991). Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. *Neurology* 41, 697–702.

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

Paulesu, E., Vallar, G., Berlingeri, M., Signorini, M., Vitali, P., Burani, C., et al. (2009). Supercalifragilisticexpialidocious: how the brain learns words never heard before. *Neuroimage* 45, 1368–1377.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., and Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.* 8, 2236–2246.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., and Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage* 52, 1410–1419.

Penfield, W., and Rasmussen, T. (1950). *The cerebral cortex of man*. New York: MacMillan.

Penfield, W., and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., et al. (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci.* 108, 16056–16061.

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci.* 108, 5419–5424.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170.

Pinker, S. (1994). *The language instinct: The new science of language and mind*. Penguin UK.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10, 15–35.

Price, C. J., and Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends Cogn. Sci.* 6, 416–421.

Pulvermüller, F. (1992). Constituents of a neurological theory of language. *Concepts Neurosci.* 3, 157–200.

Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci.* 22, 253–336.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470.

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360.

Pulvermüller, F., and Fadiga, L. (2016). "Brain language mechanisms built on action and perception," in *Hickok, G., & Small, S. L. (Eds.) Neurobiology of Language* (Amsterdam: Elsevier).

Pulvermüller, F., and Garagnani, M. (2014). From sensorimotor learning to memory cells in prefrontal and temporal association cortex: a neurocomputational study of disembodiment. *Cortex* 57, 1–21.

Pulvermüller, F., Garagnani, M., and Wennekers, T. (2014a). Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. *Biol. Cybern.* 108, 573–593.

Pulvermüller, F., Härle, M., and Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain Lang.* 78, 143–168.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso Del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. USA* 103, 7865–7870. doi:10.1073/pnas.0509989103.

Pulvermüller, F., Kiff, J., and Shtyrov, Y. (2012). Can language-action links explain language laterality?: an ERP study of perceptual and articulatory learning of novel pseudowords. *Cortex* 48, 871–881.

Pulvermüller, F., Moseley, R. L., Egorova, N., Shebani, Z., and Boulenger, V. (2014b). Motor cognition–motor semantics: Action perception theory of cognition and communication. *Neuropsychologia* 55, 71–84. doi:10.1016/j.neuropsychologia.2013.12.002.

Pulvermüller, F., and Preissl, H. (1991). A cell assembly model of language. *Netw. Comput. Neural Syst.* 2, 455–468.

Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2003). Spatiotemporal dynamics of neural language processing: an MEG study using minimum-norm current estimates. *NeuroImage* 20, 1020–1025.

Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892.

Raizada, R. D., and Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56, 726–740.

Ranganath, C., Cohen, M. X., and Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *J. Cogn. Neurosci.* 17, 994–1010.

Rauschecker, A. M., Deutsch, G. K., Ben-Shachar, M., Schwartzman, A., Perry, L. M., and Dougherty, R. F. (2009). Reading impairment in a patient with missing arcuate fasciculus. *Neuropsychologia* 47, 180–194. doi:10.1016/j.neuropsychologia.2008.08.011.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi:10.1038/nn.2331.

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci.* 97, 11800–11806.

Rilling, J. K. (2014). Comparative primate neuroimaging: insights into human brain evolution. *Trends Cogn. Sci.* 18, 46–55.

Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., and Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Front. Evol. Neurosci.* 3, 11.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Rioult-Pedotti, M.-S., Friedman, D., and Donoghue, J. P. (2000). Learning-induced LTP in neocortex. *Science* 290, 533–536.

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H.-J., et al. (2014). The role of reward in word learning and its implications for language acquisition. *Curr. Biol.* 24, 2606–2611.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., and Rozzi, S. (2014). Cortical Mechanisms Underlying the Organization of Goal-Directed Actions and Mirror Neuron-Based Action Understanding. *Physiol. Rev.* 94, 655–706.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., and Sinigaglia, C. (2015). Curious Book on Mirror Neurons and Their Myth. *Am. J. Psychol.* 128, 527–533.

Rogalsky, C., and Hickok, G. (2011). The role of Broca's area in sentence comprehension. *J. Cogn. Neurosci.* 23, 1664–1680.

Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., and Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase* 17, 178–187.

Rogalsky, C., Matchin, W., and Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. *Front. Hum. Neurosci.* 2, 14.

Rogers, J. C., Möttönen, R., Boyles, R., and Watkins, K. E. (2014). Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. *Front. Psychol.* 5, 754.

Rolls, E. T., and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford: Oxford University Press.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Romero, L., Walsh, V., and Papagno, C. (2006). The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 18, 1147–1155.

Rossini, P. M. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroenceph Clin Neurophysiol* 91, 79–92.

Roy, A. C., Craighero, L., Fabbri-Destro, M., and Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *J. Physiol.-Paris* 102, 101–105.

Sato, M., Tremblay, P., and Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111, 1–7.

de Saussure, F. (1964). *Course in General Linguistics*. London: Peter Owen.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., et al. (1993). Language comprehension in ape and child. *Monogr. Soc. Res. Child Dev.*, i–252.

Savill, N., Ashton, J., Gugliuzza, J., Poole, C., Sim, Z., Ellis, A. W., et al. (2015a). tDCS to temporoparietal cortex during familiarisation enhances the subsequent phonological coherence of nonwords in immediate serial recall. *Cortex* 63, 132–144. doi:10.1016/j.cortex.2014.08.018.

Savill, N., Ellis, A. W., and Jefferies, E. (2016). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2016.03.006.

Savill, N., Metcalfe, T., Ellis, A. W., and Jefferies, E. (2015b). Semantic categorisation of a word supports its phonological integrity in verbal short-term memory. *J. Mem. Lang.* 84, 128–138.

Saygin, Z. M., Norton, E. S., Osher, D. E., Beach, S. D., Cyr, A. B., Ozernov-Palchik, O., et al. (2013). Tracking the roots of reading ability: White matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *J. Neurosci.* 33, 13251–13258. doi:10.1523/JNEUROSCI.4383-12.2013.

Schieffelin, B. B. (1985). "The acquisition of Kaluli," in *Slobin, D.I. (Ed.) The crosslinguistic study of language acquisition: Theoretical issues* (Erlbaum), 525–593.

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., and Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cereb. Cortex* 25, 3894–3902.

Schomers, M. R., and Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Front. Hum. Neurosci.* 10, 435. doi:10.3389/fnhum.2016.00435.

Schulze, K., Vargha-Khadem, F., and Mishkin, M. (2012). Test of a motor theory of long-term auditory memory. *Proc. Natl. Acad. Sci.* 109, 7121–7125.

Schüz, A., and Braitenberg, V. (2002). The Human Cortical White Matter: Quantitative Aspects of Cortico-Cortical Long-Range Connectivity.

Schwartz, J.-L., Basirat, A., Ménard, L., and Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *J. Neurolinguistics* 25, 336–354. doi:10.1016/j.jneuroling.2009.12.004.

Schwartz, J.-L., Berthommier, F., and Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, B69–B78.

Scott, B. H., Leccese, P. A., Saleem, K. S., Kikuchi, Y., Mullarkey, M. P., Fukushima, M., et al. (2015). Intrinsic connections of the core auditory cortical regions and rostral supratemporal plane in the macaque monkey. *Cereb. Cortex*, bhv277.

Scott, B. H., and Mishkin, M. (2016). Auditory short-term memory in the primate auditory cortex. *Brain Res.* 1640, Part B, 264–277. doi:10.1016/j.brainres.2015.10.048.

Scott, B. H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. *Proc. Natl. Acad. Sci.* 109, 12237–12241.

Scott, B. H., Mishkin, M., and Yin, P. (2014). Neural correlates of auditory short-term memory in rostral superior temporal cortex. *Curr. Biol.* 24, 2767–2775.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.

Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107.

Scott, S. K., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302.

Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083.

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., et al. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cereb. Cortex* 17, 1190–1196.

Shatz, C. J. (1992). The developing brain. *Sci. Am.* 267, 60–67.

Shewan, C. M. (1980). Phonological processing in Broca's aphasics. *Brain Lang.* 10, 71–88.

Shtyrov, Y. (2011). Fast mapping of novel word forms traced neurophysiologically. *Cognition* 2, 340. doi:10.3389/fpsyg.2011.00340.

Shtyrov, Y., Butorina, A., Nikolaeva, A., and Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc. Natl. Acad. Sci.* 111, E1918–E1923. doi:10.1073/pnas.1323158111.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., and Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage* 12, 657–663.

Shtyrov, Y., Nikulin, V. V., and Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *J. Neurosci.* 30, 16864–16867.

Shtyrov, Y., Pihko, E., and Pulvermüller, F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *NeuroImage* 27, 37–47.

Simonyan, K. (2014). The laryngeal motor cortex: its organization and connectivity. *Curr. Opin. Neurobiol.* 28, 15–21.

Skeide, M. A., and Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* 17, 323–332.

Skipper, J. I., Devlin, J. T., and Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* 164, 77–105.

Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25, 76–89.

Skipper, J. I., Wassenhove, V. van, Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399. doi:10.1093/cercor/bhl147.

Smalle, E. H., Rogers, J., and Möttönen, R. (2015). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cereb. Cortex* 25, 3690–3698.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., et al. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* 502, 894–923.

Smith, E. E., and Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science* 283, 1657–1661.

Stasenko, A., Bonn, C., Teghipco, A., Garcea, F. E., Sweet, C., Dombovy, M., et al. (2015). A causal test of the motor theory of speech perception: a case of impaired speech production and spared speech perception. *Cogn. Neuropsychol.* 32, 38–57.

Stewart, L., Walsh, V., Frith, U., Rothwell, J., and others (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Ann. N. Y. Acad. Sci.*, 433–435.

Strand, F., Forssberg, H., Klingberg, T., and Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words—an event related fMRI Study. *Brain Res.* 1212, 48–54.

Strijkers, K., and Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Lang. Cogn. Neurosci.*, 1–20.

Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., and Möttönen, R. (2013). Motor excitability during visual perception of known and unknown spoken languages. *Brain Lang.* 126, 1–7. doi:10.1016/j.bandl.2013.03.002.

Szenkovits, G., Peelle, J. E., Norris, D., and Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia* 50, 1380–1392.

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., and Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex* 48, 82–96. doi:10.1016/j.cortex.2011.10.001.

Tomasello, R., Garagnani, M., Wennekers, T., and Pulvermüller, F. (2016). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2016.07.004.

Tremblay, P., and Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* 162, 60–71. doi:10.1016/j.bandl.2016.08.004.

Tremblay, S., Shiller, D. M., and Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature* 423, 866–869.

Tsumoto, T. (1992). Long-term potentiation and long-term depression in the neocortex. *Prog. Neurobiol.* 39, 209–228.

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., and Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *NeuroImage* 31, 1284–1296.

Utman, J. A., Blumstein, S. E., and Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. *Brain Lang.* 79, 444–472.

Venezia, J. H., Saberi, K., Chubb, C., and Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Front. Psychol.* 3, 157.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281, 1188–1191.

Walsh, V., and Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nat. Rev. Neurosci.* 1, 73–80. doi:10.1038/35036239.

Warren, P., and Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Percept. Psychophys.* 41, 262–275. doi:10.3758/BF03208224.

Watkins, K. E., Strafella, A. P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994.

Watkins, K., and Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *J. Cogn. Neurosci.* 16, 978–987.

Wennekers, T., Garagnani, M., and Pulvermüller, F. (2006). Language models based on Hebbian cell assemblies. *J. Physiol.-Paris* 100, 16–30.

Wernicke, C. (1874). *Der aphasische Symptomenkomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau, Germany: Kohn und Weigert.

Westermann, G., and Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain Lang.* 89, 393–400.

Wilson, H. R., and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychon. Bull. Rev.* 8, 44–57.

Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends Cogn. Sci.* 13, 329–330.

Wilson, S. M., and Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage* 33, 316–325. doi:10.1016/j.neuroimage.2006.05.032.

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.

Yeatman, J. D., Dougherty, R. F., Rykhlevskaia, E., Sherbondy, A. J., Deutsch, G. K., Wandell, B. A., et al. (2011). Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. *J. Cogn. Neurosci.* 23, 3304–3317. doi:10.1162/jocn_a_00061.

Yee, E., Blumstein, S. E., and Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: evidence from eye movements. *J. Cogn. Neurosci.* 20, 592–612. doi:10.1162/jocn.2008.20056.

Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* 24, 603–612.

Young, M. P., Scannell, J. W., and Burns, G. (1995). *Analysis of cortical connectivity*. Heidelberg: Springer.

Young, M. P., Scannell, J. W., Burns, G., and Blakemore, C. (1994). Analysis of connectivity: neural systems in the cerebral cortex. *Rev. Neurosci.* 5, 227–250.

Yuen, I., Davis, M. H., Brysbaert, M., and Rastle, K. (2010). Activation of articulatory information in speech perception. *Proc. Natl. Acad. Sci. USA* 107, 592–597.

Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849.

Zipser, D., Kehoe, B., Littlewort, G., and Fuster, J. (1993). A spiking network model of short-term active memory. *J. Neurosci.* 13, 3406–3420.

# List of publications

This dissertation is based on the following original publications:

Schomers, M.R., Kirilina, E., Weigand, A., Bajbouj, M., Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. ***Cerebral Cortex*** 25(10):3894; https://doi.org/10.1093/cercor/bhu274

Schomers, M.R., Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. ***Frontiers in Human Neuroscience*** 10:435; https://doi.org/10.3389/fnhum.2016.00435

Schomers, M. R., Garagnani, M., Pulvermüller, F. (2017). Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. ***Journal of Neuroscience*** 37(11):3045*;* https://doi.org/10.1523/jneurosci.2693-16.2017

For reasons of data protection, this page is not published in the online version

For reasons of data protection, this page is not published in the online version

# Erklärung

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbständig verfasst habe und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Malte Schomers, Berlin, September 2016