SL

In Favor of the Phonemic Principle: a Review of Neurophysiological and Neuroimaging Explorations

Alessandra Rampinini - Emiliano Ricciardi

Abstract

In the last thirty years, *in vivo* brain structural and functional exploration has sparked vivid light on the neural correlates of language. Along these lines, the study of phonological competence has offered a 'neural view' into the organization of basic speech-sensitive areas, improving the sensitivity of pre-surgical mapping and brain-computer interface-based communication. Nevertheless, only rarely the significance of these results has been recognized in the context of a century-long discussion around the theoretical, physical and cognitive consistency of the phoneme itself. Here we review recent investigations into speech perception, imagery and production at the segmental level through neuroimaging and neurophysiological techniques, showing that phonemes are processed as discrete entities, which are categorized in cognition as unique products of their acoustic and articulatory features, despite the seamless flow of the speech signal. These results seem to expand the scope of the motor theory of speech perception.

Keywords: neurolinguistics, phonology, motor theory of speech perception.

1. *Introduction*

Selecting the phonological form of a word engages processes well beyond the sensory and motor peripheral nervous system, echoing all the way up to the cerebral cortex for language perception and production (Hickok and Poeppel, 2000; Hickok, 2009). As a complement to this idea that phonological information is represented centrally and cognitively thanks to its salience, the existence of a discrete unit of speech, such as the phoneme, was already driving intense debate among linguists long before the cognitive neurosciences approached the psychophysics of speech. Worth mentioning, within the Italian phonological community, is the exchange between Marotta (2010) and Albano Leoni (2011), reprised as well in Paoloni (2012), where the argument of an abstract symbolic structure encapsulating phonemes as invariant cognitive entities conflicts with the 'phonic shape' proposal for whole-word perception.

Received: December 2016
Accepted: February 2017

To shed some further light on this debate, one must recollect how the notion of phoneme came about in the development of phonology as a scientific field: the French term *phonème* became popular at a time when linguistics was facing an existential shift from its purely historical and comparative instantiation to a synchronic perspective, which then led the way to the general, empirically-driven observations about the structure of natural languages characterizing its later developments (Mugdan, 2011; 2014). Within a few years, the abstract, psychologically-rooted nature of phonemes was first hypothesized by Baudouin de Courtenay (Stankiewicz, 1972) at the very dawn of the cognitive sciences, while the structural and functional neuroanatomy of language were being discussed by neurologists observing the first scientifically documented cases of production and perception aphasia (Broca, 1861; 1865; Wernicke, 1874). By then, the phoneme was already being presented as a mental representation, linking perceptual and motor features through associational-transformational processes engaging memory of perceived speech to shape it into a cognitive entity. In the first half of the last century, the School of Prague further elaborated this elusive concept of a minimal discrete unit pertaining to a language-specific set or *inventory* (Trubetzkoy, 1969), built upon a notion of *pars minima* dating all the way back to the classical ages and set the tone for any future debates, as thoroughly reviewed in Albano Leoni (2009). Meanwhile, discussions around the conceptual, 'psychological' consistency of phonemes (Sapir, 1925), as well as their physical outcome as «bundled distinctive [*ed.* mechanical-sensorimotor] features» (Bloomfield, 1933), tipped the scale in favor of a *phonemic principle*, and progressively towards a phonological, rather than phonetic, interpretation of the nature of language sounds (Hymes and Fought, 1981). The phonemic principle was later re-considered following Bloomsfield's observations, as data derived from speech analysis showed how features of the acoustic or articulatory kind overlap in time during speech production and are perceived as such, being determined by the physics of sound waves travelling across the oral cavity, interacting with its shape and the obstacles therein (Jakobson *et al.*, 1952; Jakobson and Halle, 1956 [2010^2]; Chomsky and Halle, 1968).

In such a climate of particle dissection, the 'cognitive status' of the phoneme kept being scrutinized and debated from a mostly theoretical standpoint for years to come, and, to a certain extent, it still is, particularly when it comes to establish its ancient but still influential role as *pars minima* (Albano Leoni, 2009). Nonetheless, the possibility of phonological categorization being an organizing principle arising and rooted in the human brain has lacked sufficient support by neuroscientific data up until very recently, when new methodologies for brain functional exploration became available. Thus, phonological competence has been studied extensively, although seldom for its own sake: the ultimately practical scope around which neuroscientists organized the study of phonology in the brain has led to a detachment of neural data from linguistic theory, so that the debate around the nature of the phoneme is now partly disengaged from the voluminous, recent experimental evidence.

To date, linguistic practice, mostly in teaching, has built upon the validity of the phoneme as an abstract notion, and phonological change is still explained by interactions between phonemes, through their distinctive features. Thus, on one hand, a debate around the nature of phonemes exists, though without sufficient, up-to-date empirical evidence; whereas, on the other hand, phonology is still mostly taught as if the issue had never surfaced.

In this review, evidence will be presented from neuroscientific data investigating the phoneme and its representation in cognition, converging on a shared result which speaks in favor of a cognitive validity for the phonemic principle: as it will be shown, the brain indeed processes phonemes as discrete bundles of articulatory-acoustic features, and organizes them into categories shaped by such characteristics.

2. *Neuroscientific investigations into the nature of the phoneme*

In this section, we will review recent data emerging from the vast neuroscientific literature investigating phonological categorization in the cerebral cortex of humans through various *in vivo* techniques, such as Electrocorticography (ECoG), Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI), offering neurophysiological as well as metabolic measurements of neuronal activity, which neuroscientists seek to correlate with a task being performed at the time of data acquisition – such as perception, production or imagery of speech sounds (Penfield and Jasper, 1954; Ogawa *et al.*, 1990; Huettel *et al.*, 2004; Niedermeyer and da Silva, 2005; Miller *et al.*, 2007; Marini, 2008).

2.1. *Multivariate classifiers and machine learning in brain exploration*

Most of the studies reviewed here exploit the potentialities offered by the latest developments for second-level statistical analysis of neuroimaging data, namely, *multivariate classifiers* describing the content of signals coming from electrical activity across the cerebral cortex (ECoG, EEG) or from the consumption of oxygen carried by hemoglobin throughout the cerebral vasculature (fMRI), itself tied to neuronal activity (Villringer and Dirnagl, 1994). Classifiers can reveal the organization of information content within and across cortical regions without any *a priori* assumptions, in what is called a 'data driven' perspective (Naselaris *et al.*, 2011), under the premise that the cortex is a complex, network-based structure whose data points are analyzed as part of a *pattern* of activated neural populations, tied to the processing of a certain stimulus (Kriegeskorte *et al.*, 2006; Lemm *et al.*, 2011). The recent innovation brought by these statistical methods to the treatment of brain data was momentous, in that it crucially allowed neuroscientists to abandon the *massive univariate approach*, where they would assume scarcely grounded statistical relationships between data points.

By 'letting data speak', machine learning algorithms instead *train* and *test* themselves: in fact, they learn the statistical behavior of the data we feed them, thus developing the power to predict any further feedings, i.e., to classify any new stimulus by placing it into its pertaining category. Consequently, multivariate approaches abandon the robustness, as well as the inner limits, of univariate analyses to break into a deeper layer of understanding than activation studies (Davis *et al.*, 2014). In fact, multivariate classifiers and machine learning offer the unique possibility to study many cognitive subtleties from a data-driven perspective, to observe the actual, pattern-based *organizing principle* driving categorization of certain stimuli, and crucially change the question from 'which area significantly engages in a specific task?' to 'how is information represented in a specific area during a specific task?' (Kriegeskorte and Bandettini, 2007), without assuming dependencies between voxels analyzed separately and brought together later on, instead considering them as a statistically interconnected system. For the neurolinguists seeking a correlation between several, often colliding theoretical models proposed to describe different aspects of language (phonology included), it is now possible to witness the brain processing not just speech perception, production or imagery in one or the other cortical area, but also to unwrap the code subtending activation, i.e., to see which regions 'recognize' phonemes on the basis of a model derived from their distinctive features, and how accurately categorization is performed.

Therefore, most of the experiments presented in this review take advantage of multivariate classifiers and machine learning to investigate the cognitive processing of phonemes (both vowels and consonants) in the cerebral cortex, with different aims, yet sharing the common result of showing that phonemes are indeed a cognitive entity, entailing the features described, as far back as eighty years ago, through the qualitative description of spoken and heard speech.

2.2. *Electrocorticography*

In this section we will review the vast ECoG literature on the topic of phonological competence[1]. For the sake of clarity, given the amount of work around this topic (mostly due to the significance of this very spatially and temporally fine-grained methodology for BCI and pre-surgical mapping) speech production, perception and imagery are treated in separated subsections.

2.2.1. *Production*

As regards the numerous ECoG-based inquiries into the speech network existing in literature, intraoperative speech *production* studies have been considered particularly often due to the necessity of assessing the eloquent cortex prior to neurosurgery, so as to avoid the unwilling impairment of verbal communication. With this aim, Pei *et al.* (2011) were able to classify four vowels and nine pairs of consonants embedded in whole words in the premotor and primary motor cortices (PM, M1), Broca's area and the posterior superior temporal gyrus (pSTG) using multivariate techniques, similarly to what Leuthardt *et al.* (2011) did in different patients with sparse

---

[1] Initially developed in the context of epilepsy treatment, ECoG measures the postsynaptic activity of local field potentials (small groups of close-by neurons), springing from pyramidal cells and mediated vertically throughout the cerebral cortex, up to the *pia mater*, across the cerebrospinal fluid (CSF), the *arachnoid* and the *dura mater*. The ECoG signal travels through a microarray of epidural or subdural electrodes placed onto – or just below – the exposed surface of the outermost membrane, the *dura mater*, through a craniotomy, and, despite its invasiveness, it features fast temporal and fine spatial resolution, making it an optimal technique for the mapping of brain activity; moreover, while electroencephalography is negatively affected by electrical signal travelling through the aforementioned tissue/fluid layers and the skull (notoriously poorly conductive), ECoG is not concerned by this problem. With this methodology, it is thus possible to measure brain activity in the temporal and spatial domain with optimal precision.

vowel-vowel or vowel-rest contrasts; interestingly, Bouchard *et al.* (2013) succeeded in identifying the consonant-to-vowel transition during spoken syllables by correlating cortical activity and articulator features for the cardinal vowels [a][i][u] and several consonants defined by four places of articulation. Along those lines, Rescher *et al.* (2013) demonstrated on a more general basis the correlation between the flow of natural speech and significant power increase in the ECoG high gamma band in Broca's area, PM and M1 in patients undergoing eloquent cortex pre-surgical assessment.

Similarly, Chakrabarti *et al.* (2013) successfully decoded 12 psychophysically-based Mel-cepstral frequency coefficients calculated from continuous speech production from high gamma band ECoG activity, similarly to a later single-patient study where the authors decoded twelve consonant-vowel (CV) syllables during, again, natural fluent reading (Kanas *et al.*, 2014). In Bouchard and Chang (2014), the ventral somatosensory cortex (vSMC) was found to predict acoustic parameters across vowel categories and different renditions of the same vowel on a single-trial basis, once again using the cardinal vowels; that same year, a study by Mugler *et al.* (2014) provided a correlation between the International Phonetic Alphabet (IPA) chart for American English production and ECoG signal measured when subjects read words aloud, isolating single phonemes in such a way that those considered close in the IPA chart often ended up being adjacent in the ECoG multivariate measures; a result similar to a previous study by Tankus *et al.* (2012), where the confusability matrix of vowels alone had been predicted by the activity pattern of vowel-specific electrodes in the rostral anterior cingulate and medial orbitofrontal cortices (rAC, MOF), representing a spatial-anatomical axis identical to that shown in the IPA, and reaching optimal performance when two different renditions of the same vowel were incorporated in the calculations by averaging their coefficients. Interestingly, in this study STG did not show vowel-specific activities, but rather its electrodes engaged in the processing of all vowels. Finally, in yet another investigation into the relationship that neural activity entertains with the IPA chart, during continuous speech production groups of consonants and vowels defined by, once again, *place* (labial, dorsal coronal) and *manner* (obstruent, sonorant) of articulation, *voicing*, as well as the vowel-consonant broad distinction, showed differential activity patterns (Lotte *et al.*, 2015): place-related activity patterns span all along the precentral gyrus (PrCG), while manner- and voicing-related ones are isolated to a smaller sensorimotor region proposed in a previous study by Brown *et al.* (2008) as the seat of laryngeal control for phoneme production (*i.e.* controlling the vocal cords to manage airflow, and thus the emission of voice). Strangely, Lotte and colleagues did not manage to observe cortical activity that was predictive of tongue position in vowels.

2.2.2. *Perception*

In speech perception, Chang *et al.* (2010) were able to classify three consonants in pSTG, where Pasley *et al.* (2012) later reconstructed the coarse spectrogram of whole words. Quite interestingly, the primary acoustic cortices (A1) of the ferret in Mesgarani *et al.* (2008) and the rat in Engineer *et al.* (2008) also encoded phonological information so as to allow robust phoneme categorization; Mesgarani *et al.* (2014) later on demonstrated how single human STG electrode sites are tuned to consonant groups defined by *manner* and *place* of articulation, and to three vowel groups (high- and low-front, low-back) defined by the integrated weight of their first and second formants (F_1 and F_2); crucially, this study demonstrated that harmonic structure is encoded at a higher-order level than simple frequential tuning, thus differentiating human from animal results, since phonological categorization in human perception extended well beyond the actual borders of the primary acoustic cortex to leak into areas engaged in associative, higher-order processing unique to humans. New results were recently obtained even with sung vowels gated at very short durations (2 to 128ms) in the guinea pig A1 with multi-unit recordings, showing robust vowel identification through timbre structure at varying pitches despite the degradation occurred from gating (Occelli *et al.*, 2016); in this study, the human behavioral counterpart showed identification rates as low as chance when the duration of the presented vowels was progressively reduced, suggesting that vowel recognition might depend on duration as a cue that the presented stimulus is in fact a language sound.

2.2.3. *Imagery*

A bridge towards restoring communication in physically impaired individuals has been suggested by few studies assessing speech imagery for BCI control, although only a few of them have investigated fine-grained, segmental-level information content. Among these, Pei *et al.* (2011) classified vowel-specific information in the premotor cortex, and consonant pairs in STG, while Leuthardt *et al.* (2011) classified couples of vowels and vowels against rest during imagery of speech, similarly to what they did with speech production. Moreover, Ikeda *et al.* (2014) classified the cardinal vowels in a sparse set of cortical locations, with varying success at different ECoG frequency bandwidths and in different subjects.

2.3. *Electroencephalography*

The temporally fast variations characterizing basic speech have made electrophysiology the most direct way to obtain signal that easily correlates to phonological processing; along these lines, in addition to ECoG, EEG as well has offered clues into basic speech[2]. Callan *et al.* (2000) showed Independent Component Analysis (Hyvärinen and Oja, 2000) to distinguish auditory evoked potentials from speech perception (listening phase of a vowel listening-plus-articulation task) and speech imagery, as well as speech 'planning' responses in the same tasks. Crucially, data as early as these showed involvement of the bilateral temporoparietal cortex, Broca's area, left M1 and PMC. Soon after, Alku *et al.* (2001) demonstrated with magnetoencephalography (MEG)[3] and EEG that the aperiodic signal characterizing vowel sounds through the laryngeal vibration is fundamental to identify a vowel even in the presence of a signature formant structure, advising against the oversimplification of the two-formant, merely acoustic model of vowel distinction, thus arguing in favor of laryngeal vibration as a motor-based component for sustained speech decoding – even though Manca and Grimaldi (2016) show that formant distance directly correlates with cortical distance and MEG activity magnitude in the auditory representation of vowel sets.

Relatively to the perception of single phonemes, a comprehensive study by Wang *et al.* (2012) showed the IPA feature chart being predicted by just the phase pattern of the EEG signal in the 2-9 Hz frequency bandwidth for eight initial consonants embedded in CV syllables, and four isolated vowels. Similarly, Kim *et al.* (2014) revealed how F_2 of the cardinal vowels could be easily classified on a single trial basis after extracting speech perception-

---

[2] By providing an electrode cap that is placed directly upon the head, EEG measures signal from the local field potentials of pyramidal cells traveling through the cortex, the CSF, the meninges and finally the poorly conductive skull: its sub-optimal spatial resolution and signal degradation notwithstanding, this simple, non-invasive and temporally optimal setup for the measuring of neural activity has been exploited in the study of speech since the earliest days of neurolinguistics (Weiss and Mueller, 2003).

[3] Similarly to EEG, MEG provides direct, noninvasive measurements of brain activity through the detection of magnetic fields generated by the synapse, providing finer spatial resolution through an increased number of measurement points than EEG, and equally good temporal resolution (Hämäläinen *et al.*, 1993).

related features from the alpha EEG band, with the aim of testing a good EEG-based pipeline for BCI-oriented feature extraction. Again in the context of BCI, Yoshimura *et al.* (2016) classified EEG-based imagery of the vowels [a] and [i] from current dipoles calculated on the fMRI-reconstructed cortical surface, showing an improvement in cortical current-based classification accuracy. Silent reading of single-vowel Chinese ideograms was tested by Wang *et al.* (2013) but good classification rates were achieved only for the ideogram (either one of two) *vs* rest condition, with lower accuracies in the distinction between the two ideograms, where DaSalla *et al.* (2009) had similarly succeeded in classifying two imagined vowels [a] and [u], though with better results in the vowel *vs.* rest than the vowel *vs.* vowel condition.

2.4. *Functional magnetic resonance imaging*

The study of the language faculty in physiology and pathology and the fMRI technique have long entertained a particularly fruitful relationship (Price, 2000; 2010; 2012), thanks to the noninvasiveness of this methodology, which exploits the magnetic properties of hemoglobin through the readout of a scanner equipped with magnetic and radiofrequency coils, to map brain function with good-to-optimal spatial resolution and acceptable temporal resolution. fMRI crucially differs from ECoG and EEG, providing an *indirect* measurement of neuronal activity, tied to the consumption of oxygen, thus describing brain function from a complementary and equally significant *metabolic perspective*.

Few fMRI studies have investigated the organization of phonological information processing and organization, and once again, mostly on the basis of single speech tasks, with a few early, univariate but relevant approaches. Among these, crucially to the idea that the fMRI signal could pick up phonologically-relevant variation, Jacquemot *et al.* (2003) contrasted free and word-selective variation in pseudowords presented to Japanese and French speakers, showing a significant «phonological grammar effect» (Jacquemot *et al.*, 2003: 9544) in the left STG and the left anterior supramarginal gyrus (SMG), whose activity increased when processing phonologically relevant, but not merely acoustic changes in both groups[4]. Later on, Rimol *et al.* (2005)

---

[4] Trials comprised pseudowords contrasting by the variables of acoustic change *vs* no-change, phonological change *vs.* no-change, and silence, coherently with the phonological system of a subject's mother tongue. Participants would hear a triplet of auditory stimuli, of which the first two were iden-

contrasted perception of plosive consonants, CV syllables and noise, revealing the sub-syllabic nature of activations in what would soon become the classical locus of phoneme multivariate classification in speech perception through both ECoG and fMRI, that is, the left STS (Liebenthal *et al.*, 2005), as well as in the middle temporal gyrus (MTG). Moreover, a recent study by Bonte *et al.* (2016) demonstrated how phonological skills directly correlate with the magnitude of activity in the left posterior STG, *planum temporale* (PT) and right STS/STG during a cardinal vowel perception with match-to-sample task, beyond age-related effects in phonological and speaker processing.

Multivariate pattern analysis on basic speech was initiated in fMRI by a seminal study demonstrating a highly organized 'cortical phonology' in the primary and secondary auditory cortices (Heschl's Gyrus, HG; STS), where a multivariate classifier distinguished patterns unique to each cardinal vowel during speech perception (Formisano *et al.*, 2008); similarly, Bonte *et al.* (2014) demonstrated how the perceived cardinal vowels could be decoded from activity within those same regions, which, furthermore, disengaged from classifying speaker identity during cross-decoding, thus suggesting that activity of early auditory regions in phonological processing may reflect task-dependent, top-down activity.

3. *Production in Perception and vice-versa*

Widespread interest towards a better understanding of how, in fact, production and perception of speech may share spatial (or information content) features may offer a neural underpinning to those models, such as the motor theor(ies) of speech perception (Lane, 1965; Studdert-Kennedy *et al.*, 1970; Liberman and Mattingly, 1985; Galantucci *et al.*, 2006), postulating the acquisition and use of language through covert-articulatory rehearsal mechanisms, which may be explained nowadays by *in vivo* brain mapping with the concept of regional crosstalk and multivariate, pattern-based code sharing. Moreover, not only a role for the premotor and motor cortices in speech perception may be argued to pin such theories to brain activity, but also the resonance of acoustic regions during speech imagery and speech production may expand this debated model into an 'acoustic theory of speech production', or,

---

tical, and give a similarity judgment for the third stimulus according to the experimental variables (change, no change, silence, either acoustic or phonological).

better, finally link the two ends of the speech network into a comprehensive model. In this direction, data as early as those in Zatorre and Evans (1992) indicated a role for the prefrontal cortex, a locus of speech production, in processing perceived speech, with Fadiga *et al.* (2002) showing Muscular Evoked Potentials (MEP) during speech listening. Further fMRI-based evidence from Buchsbaum *et al.* (2001) showed how STS, lateral and dorsal pSTG, the parietal operculum (PO) as well as primarily motor, prefrontal regions (Broca's area included) were engaged in both imagery and perception of multisyllabic pseudowords. A similar, shared network of frontotemporal regions discriminating phonologically relevant stimuli was shown by Opitz *et al.* (2003) in Broca's area and broadly the IFG, middle frontal gyrus (MFG) and PrCG, as well as pSTG, as usual in perception. Moreover, Wilson *et al.* (2004), as well as Watkins and Paus (2004), expanded the knowledge on Broca's area recruitment beyond visually-mediated speech (see also Skipper *et al.*, 2005). Furthermore, Wise *et al.* (1999) as well as Indefrey and Levelt (2004) demonstrated the involvement of perception-specific areas during speech *production*, with lesion studies also showing sensible production impairments as a consequence to damage in the posterior language cortex, particularly the PT (Buchsbaum *et al.*, 2011).

Recently, Correia *et al.* (2015) provided multivariate, fMRI-based evidence for phonemic classification of features typical to speech production in a sparse, bilateral network of regions elicited by passive perception of eight CV syllables, achieving classification of voicing, manner and place of articulation of the consonants, across vowels and speakers. Similarly, Arsenault and Buchsbaum (2015) successfully correlated a behavioral confusion matrix of twelve consonants (embedded in fixed-vowel CV syllables) with a neural confusion matrix derived from multivariate analysis of the fMRI signal in a set of acoustic, though not motor regions.

While fMRI-based investigations into the commonalities shared by different stages of speech have focused mainly on spatial characteristics (except for Arsenault and Buchsbaum, 2015; Correia *et al.*, 2015), electrophysiology has recently provided clues into the commonalities of *information content representation* across regions involved in different speech tasks. (Martin *et al.*, 2014) cross-decoded the coarse signal of the *imagined* speech envelope using the spectrogram of *articulated* speech as a model, while cross-decoding of production and perception features of CV syllables has been attempted recently within the motor cortex (Cheung *et al.*, 2016). The authors crucially showed how ECoG high gamma activity extracted from the lateral ventral sensorimotor cortex (vSMC) isolated during syllable production, predicted *manner* of articulation for three consonant groups in motor electrodes during speech *perception*, thus demonstrating how the motor cortex engages in processing of acoustic features (manner) during listening, rather than articulator features (place).

We will conclude this section with Table 1, summarizing the salient results in terms of brain localization by speech task. Each region of interest is represented alongside the task directly associated with its involvement. Moreover, functional crosstalk is indicated for each region: here, further speech tasks are indicated where within-region neural information was decoded in addition to the primary task. Figure 1 additionally shows the most relevant brain regions from the reviewed literature on the left lateral surface of the cerebrum in sagittal view.

REGION	TASK	CROSSTALK
BROCA	production	perception, imagery
M1	production	
MOF	production	
PMC	production	Perception
PO	production	perception, imagery
prCG	production	
pSTG	production	
R AC	production	
vSMC	production	Perception
BROCA	imagery	
dorsal STG	imagery	
lateral STG	imagery	
M1	imagery	
PMC	imagery	
STS	imagery	
A1 (animal model)	perception	
BROCA	perception	Production
HG	perception	
L-ant SMG	perception	
L MTG	perception	
L PT	perception	
L/bilat STS	perception	Production
pSTG	perception	Production
prCG	perception	Production
R STG	perception	

Table 1. *Regions relevant for phonological classification and cross-talk between tasks.*

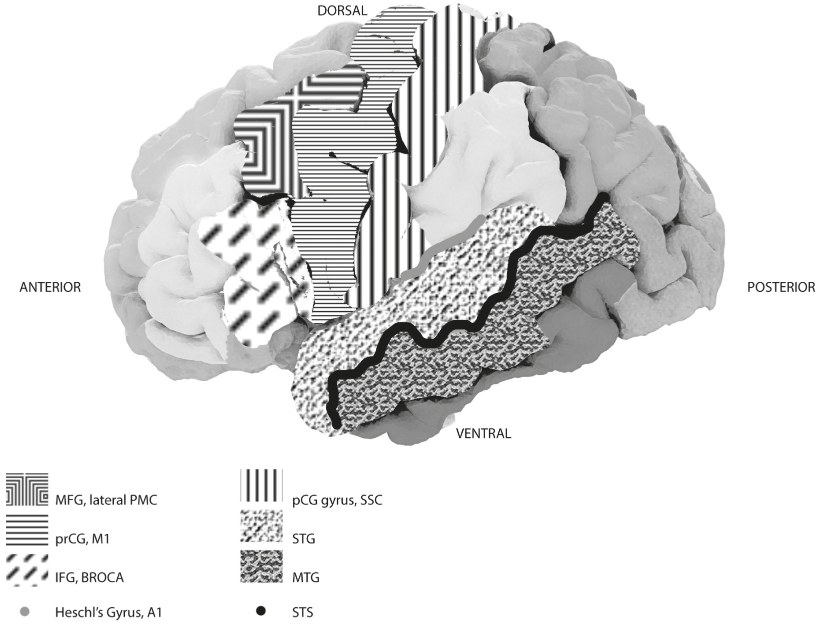

Figure 1. *Human cerebrum, lateral surface, left hemisphere.*

4. *The phoneme as an abstract and discrete unit of speech?*

An issue around the existence of the phoneme arose when its role as *pars minima vocis articulatae* started to be debated by those criticizing its historical, unchallenged status influenced by the concept of *littera*, and particularly, the fostering of the phonemic principle by Structural linguistics without sufficient empirical evidence (Albano Leoni, 2009; Marotta, 2010). In the history of phonological theory, this issue can be traced back to Bühler's criticism of the conceptual nature of phonemes (notwithstanding his acceptance of the *merkmal*) as well as Martinet's resignation in seeking a theoretical foundation to the concept of phoneme (though Martinet himself had purported the double articulation theory). In fact, phonemes seem to emerge from this debate as mere aggregates of allophones, otherwise making it impossible for speakers to derive the 'invariant' from their numerous realizations in continuous speech. Moreover, in Albano Leoni (2009) it is argued that perception unfolds through an overall acoustic shape of words, rather than in a 'phoneme by phoneme fashion' (similarly to how fluent reading proceeds), and that minimal pairs do not contribute to word distinction as well as (mainly) context, voice quality and facial expression do. Thus, segments alone would not oppose meanings. In this 'physiognomic' framework, based on holistic and fully semantic-pragmatic, more than phonologically-based comprehension, the conceptual nature of phonemes sinks lower into the abyss of speculation, and all that remains is the reality of speech production in form of a continuous acoustic signal mediated by concrete, context-based cues.

Such positions have undoubtedly enriched a debate that had come to stagnation due to lack of a fruitful confrontation with the new methodologies to investigate not just speech, but its very birth within the roots of the brain. In fact, though, the numerous observations from electrophysiological and neurofunctional data reviewed here support the hypothesis of an organization of the 'conceptual phonological space' adherent to the discrete acoustic-articulatory representation of single speech sounds, revealing the cortical correlate of phonemes as unique bundles of distinctive acoustic-articulatory features.

To reconcile these apparently opposed stances in the discussion, we may argue that while the 'acoustic shape' of words reflects the reality of perception due to physiological reasons (*i.e.* the continuous, seamless flow of the speech signal), nothing prevents the brain – which retains an astounding ability to form categories – to conceptualize the invariant version of a speech sound, serving the purpose of facilitating regional crosstalk, as well as subtending the ability to comprehend speech in adverse conditions and from a virtually infinite number of utterers in one's language, through *abstraction*. The functional sharing of regions (Broca's territory and the superior temporal lobe) and codes (phonological features) among the stages of the speech network both hint at the possibility that phonemes do indeed possess mental representations that are mutually comprehensible across the speech network and whose content is integrated by (at least) the acoustic and the motor components of the central nervous system. If the acoustic shape of words was the only representational criterion for language comprehension and articulation, many of the studies presented here would have failed in extracting phoneme-specific features for classification and cross-classification of information content within the regions involved in speech production, perception or imagery. We would argue that these two positions reconcile in that the physical flow of the speech signal, seamless by nature, is made up of holistic phenomena that supplement comprehension, while *in nuce* our brains cannot abstain from conceptualizing representations of one's own phonological inventory, arising from region- and code-sharing, and subtending abstraction as they do with many other aspects of language, such as semantics, even when certain sensory modalities are lacking (Handjaras *et al.*, 2016).

5. *A cue into the auditory-motor nature of the speech network*

Here we reviewed the current body of literature exploring phonological competence and categorization through *in vivo* brain exploration techniques, mostly exploiting the new possibilities offered by recent developments in data analysis for the neurosciences. Despite the variety of aims and methodologies, all presented studies appear to converge on the existence of an abstract, now we would say *cognitively rooted*, basis to the phoneme. Indeed, the results offered by neurofunctional data show how phonological representations can be traced in the cerebral cortex during perception, articulation, and imagery of speech sounds, suggesting a role for speech as both a sound and an action, as postulated by Scott and Johnsrude (2003).

An interesting implication concerns the relationship between the cognitive representation of the phoneme and the role of region- and code-sharing in language acquisition and processing all throughout the lifespan of healthy speakers. In fact, the motor theory of speech perception in its original formulation, as well as its subsequent revisions, revolves around the possibility that language is processed by interactions between its acoustic and motor components: specifically, its main tenet postulates that the merely passive perception of speech initiates some sort of articulatory 'resonance', engaging the memory of articulatory gestures for speech production as this mechanism trains itself covertly. Moreover, what is perceived together with the acoustic content is the articulatory gesture, so that perceiving and producing is one and the same (Galantucci *et al.*, 2006).

In this context, the discovery of mirror neurons in the non-human primate brain's area F5[5] sparked new life into a theory that had been received

---

[5] Mirror neurons have been shown in the primate's brain to fire in the context of action observation, *i.e.* when a conspecific to the observer performed the action of grasping, bringing something to the mouth, putting something on a plate, and tearing (FADIGA *et al.*, 1995; GALLESE *et al.*, 1996; RIZZOLATTI *et al.*, 1996; IACOBONI *et al.*, 1999; RIZZOLATTI *et al.*, 2001; KOHLER *et al.*, 2002). This has led to an intense debate around the existence of mirror neuron-based mechanisms in humans and what

with mixed success in cognitive psychology and was waiting for some form of experimental validation through *in vivo* investigations: Broca's area in humans seems to serve as a cytoarchitectural and functional homologue to the premotor location of mirror neurons in primates, so that, since the dawn of the 'mirror neuron age' in the cognitive sciences, it has been postulated that the development of language is based on an action recognition mechanism in humans similar to that observed in primates for hand and mouth gestures (Rizzolatti and Arbib, 1998). Therefore, the motor theory of speech perception as well as the more recent *articulatory filter hypothesis* (Vihman, 1991; 1993; Westermann and Miranda, 2004) have been brought forward and pushed to convergence by a possible explanation of their workings springing from the anatomy and function of the brain itself, something that was less accessible in the years of Liberman and Mattingly (Ziegler, 2008). In particular, while the motor theory of speech perception has benefited from the possibility of mirror neurons populating the brains (and giving rise to the minds) of humans, the articulatory filter hypothesis was born later and from the recent, fertile environment of the computational neurosciences: the idea behind this model completes the motor theory of speech perception in that it clarifies the *learned* rather than *innate* nature of the action imitation mechanism subtending phonological learning, pinning it to a specific age in ontogeny, that is the *babbling phase* in pre-linguistic infants. In fact, at this age, children start producing open syllables (Jakobson, 1941; Vihman, 1991; Guasti, 2004) that progressively align to the phonological system of their mother tongue by learning to link articulatory gestures to their 'auditory consequences' (Westermann and Miranda, 2004: 393), therefore refining the phonological inventory to its fullness by means of mapping between a motor and an acoustic module.

Relevantly to our discussion around the cognitive nature of the phoneme, both the motor theory of speech perception and the articulatory filter hypothesis fall within the boundaries of the phonemic principle and pin it

---

the significance of this system might be in the phylogenic context of language development through hand- and then mouth-based gestures (CORBALLIS, 2010), as well as social cognition (GALLESE and GOLDMAN, 1998; IACOBONI, 2008): evidence for the presence of mirror neurons in humans has been debated vividly ever since; sometimes seen as the basis of the 'great leap forward' that brought humans apart from non-human primates (RAMACHANDRAN, 2000), sometimes directly opposed when bringing together the diverse and layered load of experimental evidence (TURELLA *et al.*, 2009), it remains unclear whether a mirror system exists in humans whatsoever (as conflicting examples, see KILNER *et al.*, 2009; LINGNAU *et al.*, 2009), and its significance in the context of language development in phylogeny is argued (HURFORD, 2004).

to cognition, with experimental evidence from the presented studies of functional neurophysiology and brain mapping showing *where* and *how finely* the brain actually computes these 'mapping operations' between acoustic and articulatory parameters, to represent, process and categorize the phonological inventory of the mother tongue as bundles of features that can be described in their dual nature of perceptions *and* actions.

Nevertheless, the existence of mirror neurons in humans for language processing, as well as the consistency of these acoustic-motor mapping theories are still matter of debate due to both conflicting evidence (as to the former issue) and the very practical nature of the studies that investigated phonological processing in the brain (as to the latter). While the 'mirror neuron issue' falls beyond the scope of this work and was discussed as a tentative explanation for the anatomical findings and functional correlates of motor-acoustic mapping through covert rehearsal of speech sounds, the results presented so far do sustain the motor theory of speech perception and the articulatory filter hypothesis, pinning them to the segmental roots of language.

Yet, it is still to be clarified how this motor resonance during perception, and its perceptual counterpart during production, actually emerge within the (several) regions that are engaged in speech at this very basic level. In fact, further investigations building upon the acquired knowledge of the phoneme as an entity rooted in cognition may, eventually, reveal that the neural code (*i.e.*, the organizing principle) subtending phonological representations is shared by speech perception, imagery and articulation. Most of the data reviewed here emerged from experiments whose premises were not meant to demonstrate either the phonemic principle or any kind of motor theory-oriented code sharing (with the exception of Cheung *et al.*, 2016).

In this direction, Rampinini *et al.* (2015) preliminarily explored the processing of the Italian complete vocalic system with functional magnetic resonance imaging in 15 native Italian speakers who had to listen to, and then rehearse covertly or articulate overtly each of the seven Italian vowels one at a time; subjects also heard a set of seven pure nonlinguistic tones. The whole set of experimental conditions was constructed to assess any differences in phoneme representation in the brain across the stages of speech, from pure sound perception, through phonological perception, covert rehearsal and then to overt articulation. The tasks simulated the steps through which speech is transformed from a perceptual to a motor act, according to the motor theory of speech perception and articulatory filter hypothesis. In this study, vowel-specific information, recalling the trapezoidal scheme describing the acoustic-articulatory space, was represented similarly in a language-specific network of regions from the superior temporal to the inferior frontal areas of the left hemisphere. This cortical network, and particularly Broca's area in the inferior frontal gyrus, classically engaged in the processing of language at different and also more complex levels (Clos *et al.*, 2013; Price, 2012; Vigneau *et al.*, 2006), showed to be sensitive to phonological representations, a very fine, though basic kind of linguistic processing. Moreover, Broca's area showed sensitivity to nonlinguistic sounds (pure tones), even though it lies well outside the primary and secondary acoustic hubs of the brain. The finding reinforces the argument of its sensitivity to features typical to phonemes (i.e., frequencies), justified by the structural connectivity of this region with the superior temporal cortex (Amunts and Zilles, 2012).

These results suggest the expansion of the motor theory of speech perception into an acoustic theory of speech production, hinting at the engagement of these regions in processing purely acoustic cues since, speculatively, as newborns and infants we listen before speaking. Moreover, these preliminary data showed how both the prefrontal and temporal speech hubs perform some sort of crosstalk, since the former was consistently engaged in speech perception (even to the point of processing pure tones, as explained, as well as heard vowels), while the latter was shown to be engaged in the active stages of speech, though significantly in speech imagery only, despite being a receptive region.

These results may contribute to validating the functional importance of regional crosstalk across the widespread speech network, assuming that a shared code should organize 'neural phonology', without the need for the exact same regions to process multiple tasks, rather highlighting the value of reciprocal 'intelligibility' through a shared code. The role of imagery/covert rehearsal as a functional bridge in the context of the mirror neuron/motor theory hypothesis was first evaluated together with the weighing of purely acoustic cues (i.e., frequency and harmonic structure), with respect to the idea that the motor and acoustic systems for speech processing constantly interact regionally and qualitatively across the frontotemporal language areas. Speculatively, one can argue that the representation of sounds outside the linguistic inventory of the speaker would not yield the same 'neural picture', whereas the representation of allophones might as well do, since they might be implicitly linked to their unmarked, fully phonological realization.

6. *Conclusions*

In conclusion, a language-specific sound inventory couched on the phoneme as a bundle of auditory and articulatory features endowed with neural correlates can widely enrich the discussion around the phonemic principle. At the same time, it appears to crucially sustain language acquisition theories pinning phonological competence to a proper neuroanatomical and neurofunctional basis.

The availability of literature investigating the neural network underlying speech with various techniques and at different scales has brought a growing though sparse body of evidence that:

a. Phonological sensitivity is widespread across the cortex;
b. It depends on basic, as well as high-order processes;
c. It is driven by the synergic activity of different brain regions.

The empirical evidence reviewed here seems to suggest that the phoneme is yet a linguistic entity rooted in cognition, and not only an epiphenomenon of the mechanics of our mouths and ears through social interactions. Moreover, the reviewed literature suggests further investigations into any plausible models driving neural phonology, with interesting implications for those theoretical mainframes suggesting a motor-articulatory resonance in speech perception as well as a perceptual-acoustic one in speech production. Therefore, the hypothesis that human mirror neurons might be involved in speech processing thanks to their neuroanatomical locations, covering the premotor cortex also engaged in language, could be further supported.

The current discussion highlights the relevance of a constant and evidence-based dialogue between linguistics and neuroscience, in order to foster and forward new knowledge.

Acknowledgments

The authors would like to thank Luca Cecchetti and Paolo Papale for providing the template from which Figure 1 was adapted.

Bibliography

Albano Leoni, F. (2009), *Dei suoni e dei sensi. Il volto fonico delle parole*, Il Mulino, Bologna.

Albano Leoni, F. (2011), *Discutendo sulla (presunta) morte del fonema,* in «Studi e Saggi Linguistici», 49, pp. 205-220.

Amunts, K. and Zilles, K. (2012), *Architecture and organizational principles of Broca's region*, in «Trends in cognitive sciences», 16, 8, pp. 418-426.

Alku, P., Sivonen, P., Palomäki, K. and Tiitinen, H. (2001), *The periodic structure of vowel sounds is reflected in human electromagnetic brain responses*, in «Neuroscience letters», 298, 1, pp. 25-28.

Arsenault, J.S. and Buchsbaum, B.R. (2015), *Distributed neural representations of phonological features during speech perception*, in «The Journal of Neuroscience», 35, 2, pp. 634-642.

Bloomfield, L. (1933), *Language*, Holt, New York.

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*, in «The Journal of Neuroscience», 34, 13, pp. 4548-4557.

Bonte, M., Ley, A., Scharke, W. and Formisano, E. (2016), *Developmental refinement of cortical systems for speech and voice processing*, in «NeuroImage», 128, pp. 373-384.

Bouchard, K.E. and Chang, E.F. (2014), *Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex*, in «The Journal of Neuroscience», 34, 38, pp. 12662-12677.

Bouchard, K.E., Mesgarani, N., Johnson, K. and Chang, E.F. (2013), *Functional organization of human sensorimotor cortex for speech articulation*, in «Nature», 495, 7441, pp. 327-332.

Broca, P. (1861), *Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau*, in «Bulletins de la Société d'anthropologie de Paris», 2, 1, pp. 235-238.

Broca, P. (1865), *Sur le siège de la faculté du langage articulé*, in «Bulletins de la Société d'anthropologie de Paris», 6, 1, pp. 377-393.

Brown, S., Ngan, E. and Liotti, M. (2008), *A larynx area in the human motor cortex*, in «Cerebral Cortex», 18, 4, pp. 837-845.

Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M. and Hickok, G. (2011), *Conduction aphasia, sensory-motor integration, and phonological short-term memory-an aggregate analysis of lesion and fMRI data*, in «Brain and Language», 119, 3, pp. 119-128.

Buchsbaum, B.R., Hickok, G. and Humphries, C. (2001), *Role of left posterior superior temporal gyrus in phonological processing for speech perception and production*, in «Cognitive Science», 25, 5, pp. 663-678.

Callan, D.E., Callan, A.M., Honda, K. and Masaki, S. (2000), *Single-sweep EEG analysis of neural processes underlying perception and production of vowels*, in «Cognitive brain research», 10, 1, pp. 173-176.

Chakrabarti, S., Krusienski, D.J., Schalk, G. and Brumberg, J.S. (2013), *Predicting mel-frequency cepstral coefficients from electrocorticographic signals during continuous speech production,* Abstract presented at *Proceedings of the Sixth International IEEE/EMBS Neural Engineering Conference*, San Diego, CA.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M. and Knight, R.T. (2010), *Categorical speech representation in human superior temporal gyrus*, in «Nature neuroscience», 13, 11, pp. 1428-1432.

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

Chomsky, N. and Halle, M. (1968), *The sound pattern of English*, Harper & Row, New York.

Clos, M., Amunts, K., Laird, A.R., Fox, P.T. and Eickhoff, S.B. (2013), *Tackling the multifunctional nature of Broca's region meta-analytically: co-activation-based parcellation of area 44*, in «NeuroImage», 83, pp. 174-188.

Corballis, M.C. (2010), *Mirror neurons and the evolution of language*, in «Brain and Language», 112, 1, pp. 25-35.

Correia, J.M., Jansma, B.M.B. and Bonte, M. (2015), *Decoding articulatory features from fMRI responses in dorsal speech regions*, in «The Journal of Neuroscience», 35, 45, pp. 15015-15025.

DaSalla, C.S., Kambara, H., Sato, M. and Koike, Y. (2009), *Single-trial classification of vowel speech imagery using common spatial patterns*, in «Neural Networks», 22, 9, pp. 1334-1339.

Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D. and Poldrack, R.A. (2014), *What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis*, in «NeuroImage», 97, pp. 271-283.

Engineer, C.T., Perez, C.A., Chen, Y. H., Carraway, R. S., Reed, A.C., Shetake, J.A., Jakkamsetti, V., Chang, K.Q. and Kilgard, M.P. (2008), *Cortical activity patterns predict speech discrimination ability*, in «Nature neuroscience», 11, 5, pp. 603-608.

Fadiga, L., Craighero, L., Buccino, G. and Rizzolatti, G. (2002), *Speech listening specifically modulates the excitability of tongue muscles: a TMS study*, in «European Journal of Neuroscience», 15, 2, pp. 399-402.

Fadiga, L., Fogassi, L., Pavesi, G. and Rizzolatti, G. (1995), *Motor facilitation during action observation: a magnetic stimulation study*, in «Journal of neurophysiology», 73, 6, pp. 2608-2611.

Formisano, E., De Martino, F., Bonte, M. and Goebel, R. (2008), *"Who" Is Saying" What"? Brain-Based Decoding of Human Voice and Speech*, in «Science», 322, 5903, pp. 970-973.

Galantucci, B., Fowler, C.A. and Turvey, M.T. (2006), *The motor theory of speech perception reviewed*, in «Psychonomic bulletin & review», 13, 3, pp. 361-377.

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996), *Action recognition in the premotor cortex*, in «Brain», 119, 2, pp. 593-609.

Gallese, V. and Goldman, A. (1998), *Mirror neurons and the simulation theory of mind-reading*, in «Trends in cognitive sciences», 2, 12, pp. 493-501.

Guasti, M.T. (2004), *Language acquisition: The growth of grammar*, MIT Press, Boston.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J. and Lounasmaa, O.V. (1993), *Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain*, in «Reviews of modern Physics», 65, 2, pp. 413.

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G. and Pietrini, P. (2016), *How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge*, in «NeuroImage», 135, pp. 232-242.

Hickok, G. (2009), *The functional neuroanatomy of language*, in «Physics of Life Reviews», 6, 3, pp. 121-143.

Hickok, G. and Poeppel, D. (2000), *Towards a functional neuroanatomy of speech perception*, in «Trends in Cognitive Science», 4, 4, pp. 131-138.

Huettel, S.A., Song, A.W. and McCarthy, G. (2004), *Functional magnetic resonance imaging*, Sinauer Associates, Sunderland.

Hurford, J.R. (2004), *Language beyond our grasp: what mirror neurons can, and cannot, do for the evolution of language*, in «Evolution of communication systems», pp. 297-314.

Hymes, D.H. and Fought, J.G. (1981), *American structuralism*, Walter de Gruyter, Berlin.

Hyvärinen, A. and Oja, E. (2000), *Independent component analysis: algorithms and applications*, in «Neural networks», 13, 4, pp. 411-430.

Iacoboni, M. (2008), *Imitation, Empathy, and Mirror Neurons*, in «Annual Review of Psychology», 60, 1, pp. 653-670.

Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C. and Rizzolatti, G. (1999), *Cortical mechanisms of human imitation*, in «Science», 286, 5449, pp. 2526-2528.

Ikeda, S., Shibata, T., Nakano, N., Okada, R., Tsuyuguchi, N., Ikeda, K. and Kato, A. (2014), *Neural decoding of single vowels during covert articulation using electrocorticography*, in «Frontiers in human neuroscience», 8, 125.

Indefrey, P. and Levelt, W.J.M. (2004), *The spatial and temporal signatures of word production components*, in «Cognition», 92, 1, pp. 101-144.

Jacquemot, C., Pallier, C., Lebihan, D., Dehaene, S. and Dupoux, E. (2003), *Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study*, in «The Journal of Neuroscience», 23, 29, pp. 9541-9546.

Jakobson, R. (1941), *Child Language, Aphasia and Linguistic Universals*, Mouton, The Hague.

Jakobson, R., Fant, G. and Halle, M. (1952), *Preliminaries to speech analysis. The distinctive features and their correlates*, The MIT Press, Boston.

Jakobson, R. and Halle, M. (1956, [2010^2]), *Fundamentals of language*, in van Schooneveld, C.H. (2010, ed.), *Janua linguarum series minor. Studia memoriae Nicolai van Wijk dedicata*. Vol. 1, de Gruyter Mouton, Berlin/Boston.

Kanas, V.G., Mporas, I., Benz, H. L., Sgarbas, K.N., Bezerianos, A. and Crone, N.E. (2014), *Joint spatial-spectral feature space clustering for speech activity detection from ECoG signals*, in «IEEE Transactions on Biomedical Engineering», 61, 4, pp. 1241-1250.

Kilner, J.M., Neal, A., Weiskopf, N., Friston, K. J. and Frith, C.D. (2009), *Evidence of mirror neurons in human inferior frontal gyrus*, in «The Journal of Neuroscience», 29, 32, pp. 10153-10159.

Kim, J., Lee, S. K. and Lee, B. (2014), *EEG classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition*, in «Journal of neural engineering», 11, 3, 036010.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002), *Hearing sounds, understanding actions: action representation in mirror neurons*, in «Science», 297, 5582, pp. 846-848.

Kriegeskorte, N. and Bandettini, P. (2007), *Analyzing for information, not activation, to exploit high-resolution fMRI*, in «NeuroImage», 38, 4, pp. 649-662.

Kriegeskorte, N., Goebel, R. and Bandettini, P. (2006), *Information-based functional brain mapping*, in «Proceedings of the National Academy of Sciences», 103, 10, pp. 3863-3868.

Lane, H. (1965), *The motor theory of speech perception: A critical review*, in «Psychological Review», 72, 4, p. 275.

Lemm, S., Blankertz, B., Dickhaus, T. and Müller, K. R. (2011), *Introduction to machine learning for brain imaging*, in «NeuroImage», 56, 2, pp. 387-399.

Leuthardt, E.C., Gaona, C., Sharma, M., Szrama, N., Roland, J., Freudenberg, Z., Solis, J., Breshears, J. and Schalk, G. (2011), *Using the electrocorticographic speech network to control a brain–computer interface in humans*, in «Journal of neural engineering», 8, 3, 036004.

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

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T. and Medler, D.A. (2005), *Neural substrates of phonemic perception*, in «Cerebral Cortex», 15, 10, pp. 1621-1631.

Lingnau, A., Gesierich, B. and Caramazza, A. (2009), *Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans*, in «Proceedings of the National Academy of Sciences», 106, 24, pp. 9925-9930.

Lotte, F., Brumberg, J.S., Brunner, P., Gunduz, A., Ritaccio, A.L., Guan, C. and Schalk, G. (2015), *Electrocorticographic representations of segmental features in continuous speech*, in «Frontiers in human neuroscience», 9, p. 97.

Manca, A.D. and Grimaldi, M. (2016), *Vowels and Consonants in the Brain: Evidence from Magnetoencephalographic Studies on the N1m in Normal-Hearing Listeners*, in «Frontiers in Psychology», 7, p. 1413.

Marini, A. (2008), *Manuale di neurolinguistica*, Carocci, Roma.

Marotta, G. (2010), *Sulla (presunta) morte del fonema*, in «Studi e Saggi Linguistici», 48, pp. 283-304.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H.J., Crone, N.E., Rieger, J., Schalk, G., Knight, R.T. and Pasley, B.N. (2014), *Decoding spectrotemporal features of overt and covert speech from the human cortex*, in «Frontiers in neuroengineering», 7, p. 14.

Mesgarani, N., Cheung, C., Johnson, K. and Chang, E.F. (2014), *Phonetic feature encoding in human superior temporal gyrus*, in «Science», 343, 6174, pp. 1006-1010.

Mesgarani, N., David, S.V., Fritz, J.B. and Shamma, S.A. (2008), *Phoneme representation and classification in primary auditory cortex*, in «The Journal of the Acoustical Society of America», 123, 2, pp. 899-909.

Miller, K.J., Shenoy, P., Miller, J.W., Rao, R.P.N. and Ojemann, J.G. (2007), *Real-time functional brain mapping using electrocorticography*, in «NeuroImage», 37, 2, pp. 504-507.

Mugdan, J. (2011), *On the Origins of the Term Phoneme*, in «Historiographia Linguistica», 38, 1-2, pp. 85-110.

Mugdan, J. (2014), *More on the origins of the term phoneme*, in «Historiographia Linguistica», 41, 1, pp. 185-187.

Mugler, E.M., Patton, J.L., Flint, R.D., Wright, Z.A., Schuele, S.U., Rosenow, J., Shih, J.J., Krusienski, D.J. and Slutzky, M.W. (2014), *Direct classification of all American English phonemes using signals from functional speech motor cortex*, in «Journal of neural engineering», 11, 3, pp. 035015.

Naselaris, T., Kay, K.N., Nishimoto, S. and Gallant, J.L. (2011), *Encoding and decoding in fMRI*, in «NeuroImage», 56, 2, pp. 400-410.

Niedermeyer, E. and da Silva, F.H.L. (2005), *Electroencephalography: basic principles, clinical applications, and related fields*, Lippincott Williams & Wilkins, Philadelphia.

Occelli, F., Suied, C., Pressnitzer, D., Edeline, J.M. and Gourévitch, B. (2016), *A Neural Substrate for Rapid Timbre Recognition? Neural and Behavioral Discrimination of Very Brief Acoustic Vowels*, in «Cerebral Cortex», 26, 6, pp. 2483-2496.

Ogawa, S., Lee, T.M., Kay, A.R. and Tank, D.W. (1990), *Brain magnetic resonance imaging with contrast dependent on blood oxygenation*, in «Proceedings of the National Academy of Sciences», 87, 24, pp. 9868-9872.

Opitz, B., Müller, K. and Friederici, A.D. (2003), *Phonological processing during language production: fMRI evidence for a shared production-comprehension network*, in «Cognitive Brain Research», 16, 2, pp. 285-296.

Paoloni, A. (2012), *Una nota sul dibattito relativo alla nozione di fonema,* in «Studi e Saggi Linguistici», 50, 2, pp. 165-172.

Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T. and Chang, E.F. (2012), *Reconstructing speech from human auditory cortex*, in «PLoS Biology», 10, 1, e1001251.

Pei, X., Barbour, D.L., Leuthardt, E.C. and Schalk, G. (2011), *Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans*, in «Journal of neural engineering», 8, 4, 046028.

Penfield, W. and Jasper, H. (1954), *Epilepsy and the functional anatomy of the human brain*, in «Science», 119, 3097, pp. 645-646.

Price, C.J. (2000), *The anatomy of language: contributions from functional neuroimaging*, in «Journal of Anatomy», 197, 3, pp. 335-359.

Price, C.J. (2010), *The anatomy of language: a review of 100 fMRI studies published in 2009*, in «Annals of the New York Academy of Sciences», 1191, pp. 62-88.

Price, C.J. (2012), *A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading*, in «NeuroImage», 62, 2, pp. 816-847.

Ramachandran, V.S. (2000), *Mirror neurons and imitation learning as the driving force behind "the great leap forward" in human evolution*, Edge, New York.

Rampinini, A., Handjaras, G., Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G. and Pietrini, P. (2015). *Decoding speech: multivariate classification of the Italian vocalic system through fMRI*, in «Neuropsychological trends», 18, pp. 141-142.

Rimol, L.M., Specht, K., Weis, S., Savoy, R. and Hugdahl, K. (2005), *Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study*, in «NeuroImage», 26, 4, pp. 1059-1067.

Rizzolatti, G. and Arbib, M.A. (1998), *Language within our grasp*, in «Trends in neurosciences», 21, 5, pp. 188-194.

Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996), *Premotor cortex and the recognition of motor actions*, in «Cognitive brain research», 3, 2, pp. 131-141.

Rizzolatti, G., Fogassi, L. and Gallese, V. (2001), *Neurophysiological mechanisms underlying the understanding and imitation of action*, in «Nature Reviews Neuroscience», 2, 9, pp. 661-670.

Ruescher, J., Iljina, O., Altenmüller, D.M., Aertsen, A., Schulze-Bonhage, A. and Ball, T. (2013), *Somatotopic mapping of natural upper-and lower-extremity movements and speech production with high gamma electrocorticography*, in «NeuroImage», 81, pp. 164-177.

Sapir, E. (1925), *Sound patterns in language*, in «Language», 1, 2, pp. 37-51.

Scott, S. K. and Johnsrude, I. S. (2003), *The neuroanatomical and functional organization of speech perception*, in «Trends in neurosciences», 26, 2, pp. 100-107.

Skipper, J.I., Nusbaum, H.C. and Small, S.L. (2005), *Listening to talking faces: motor cortical activation during speech perception*, in «NeuroImage», 25, 1, pp. 76-89.

Stankiewicz, E. (1972), *A Baudouin de Courtenay anthology: the beginnings of structural linguistics*, Indiana University Press, Bloomington.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. and Cooper, F.S. (1970), *Motor theory of speech perception: A reply to Lane's critical review*, in «Psychological Review», 77, 3, pp. 234-249.

Tankus, A., Fried, I. and Shoham, S. (2012), *Structured neuronal encoding and decoding of human speech features*, in «Nature communications», 3, p. 1015.

Trubetzkoy, N. S. (1969), *Principles of phonology*, University of California Press, Berkeley.

Turella, L., Pierno, A.C., Tubaldi, F. and Castiello, U. (2009), *Mirror neurons in humans: consisting or confounding evidence?*, in «Brain and language», 108, 1, pp. 10-21.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B. and Tzourio-Mazoyer, N. (2006), *Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing*, in «NeuroImage», 30, 4, pp. 1414-1432.

Vihman, M. (1991), *Ontogeny of phonetic gestures: Speech production*, in Studdert-Kennedy, M. and Mattingly, I.G. (1991, eds.). *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*, Erlbaum, Hillsdale, pp. 69-84.

Vihman, M. (1993), *Variable paths to early word production*, in «Journal of Phonetics», 21, 1-2, pp. 61-82.

Villringer, A. and Dirnagl, U. (1994), *Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging*, in «Cerebrovascular and brain metabolism reviews», 7, 3, pp. 240-276.

Wang, L., Zhang, X., Zhong, X. and Zhang, Y. (2013), *Analysis and classification of speech imagery EEG for BCI*, in «Biomedical Signal Processing and Control», 8, 6, pp. 901-908.

Wang, R., Perreau-Guimaraes, M., Carvalhaes, C. and Suppes, P. (2012), *Using phase to recognize English phonemes and their distinctive features in the brain*, in «Proceedings of the National Academy of Sciences», 109, 50, pp. 20685-20690.

Watkins, K. and Paus, T. (2004), *Modulation of motor excitability during speech perception: the role of Broca's area*, in «Journal of cognitive neuroscience», 16, 6, pp. 978-987.

Weiss, S. and Mueller, H.M. (2003), *The contribution of EEG coherence to the investigation of language*, in «Brain and language», 85, 2, pp. 325-343.

Wernicke, C. (1874), *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*, Max Cohn & Weigert, Breslau.

Westermann, G. and Miranda, E.R. (2004), *A new model of sensorimotor coupling in the development of speech*, in «Brain and language», 89, 2, pp. 393-400.

Wilson, S.M., Saygin, A.P., Sereno, M.I. and Iacoboni, M. (2004), *Listening to speech activates motor areas involved in speech production*, in «Nature neuroscience», 7, 7, pp. 701-702.

Wise, R.J.S., Greene, J., Büchel, C. and Scott, S.K. (1999), *Brain regions involved in articulation*, in «The Lancet», 353, 9158, pp. 1057-1061.

Yoshimura, N., Nishimoto, A., Belkacem, A.N., Shin, D., Kambara, H., Hanakawa, T. and Koike, Y. (2016), *Decoding of covert vowel articulation using electroencephalography cortical currents*, in «Frontiers in neuroscience», 10, p. 175.

Zatorre, R.J. and Evans, A.C. (1992), *Lateralization of phonetic and pitch discrimination in speech processing*, in «Science», 256, 5058, p. 846.

Ziegler, W. (2008), *Neurophonetics*, in Ball, M.J., Perkins, M.R., Müller, N. and Howard, S. (2008, eds.), *The Handbook of Clinical Linguistics*, Blackwell, Oxford.

Alessandra Rampinini
MoMiLab, IMT School for Advanced Studies Lucca
Piazza S. Francesco 19
55100 Lucca (Italy)
alessandra.rampinini@imtlucca.it

Emiliano Ricciardi
MoMiLab, IMT School for Advanced Studies Lucca
Piazza S. Francesco 19
55100 Lucca (Italy)
emiliano.ricciardi@imtlucca.it