IMT School for Advanced Studies, Lucca

Lucca, Italy

# Multivariate analyses of neural patterns in the human brain

PhD Program in Cognitive, Computational and Social Neuroscience

XXXII Cycle

Giacomo Handjaras

2019

The dissertation of Giacomo Handjaras is approved.

Program Coordinator: Pietro Pietrini, IMT School for Advanced Studies Lucca

Advisor: Pietro Pietrini, IMT School for Advanced Studies Lucca
Co-advisor: Emiliano Ricciardi, IMT School for Advanced Studies Lucca

The dissertation of Giacomo Handjaras has been reviewed by:

Prof. Patrizia Baraldi, University of Modena and Reggio Emilia, Italy

Dr. Paul Taylor, Scientific and Statistical Computing Core, National Institute of Mental Health, Bethesda, Maryland, USA

IMT School for Advanced Studies, Lucca
2019

# Table of contents

# List of figures

# Acknowledgements

This thesis incorporates material from four papers. Chapter 2 uses material from the manuscript published in *Scientific Reports* of Rampinini & Handjaras et al. (2017), coauthored with Leo, Cecchetti, Ricciardi, Marotta and Pietrini. The affiliation of all the authors is IMT School for Advanced Studies Lucca, except for Marotta which is University of Pisa. Chapter 3 is based on Rampinini et al. (2019), published in *Frontiers in human neuroscience* and coauthored with Handjaras, Leo, Cecchetti, Betta, Ricciardi, Marotta and Pietrini. The affiliation of all the authors is IMT School for Advanced Studies Lucca, except for Marotta which is University of Pisa. Chapter 4 comprised the work of Handjaras et al. (2017), published in *Neuropsychologia* and coauthored with Leo, Cecchetti, Papale, Lenci, Marotta, Pietrini and Ricciardi. The affiliation of all the authors is IMT School for Advanced Studies Lucca, except for Lenci and Marotta which is University of Pisa. Finally, Chapter 5 is based on Benuzzi et al. (2018), published in *Frontiers in behavioral neuroscience* and coauthored with Ballotta, Handjaras, Leo, Papale, Zucchelli, Molinari, Lui, Cecchetti, Ricciardi, Sartori, Pietrini and Nichelli. The affiliation of Benuzzi, Ballotta, Zucchelli, Molinari, Lui, Nichelli is University of Modena and Reggio Emilia. The affiliation of Handjaras, Leo, Papale, Cecchetti, Ricciardi, Pietrini is IMT School for Advanced Studies Lucca. The affiliation of Sartori is University of Padova.

I would like to thank Sabrina Danti, Giada Lettieri, Luca Cecchetti, Davide Bottari, Emiliano Ricciardi and Pietro Pietrini for their support in the draft of this dissertation.

# Vita and publications

Giacomo Handjaras was born in Italy on 16/05/1976. He lives in Lucca. In the early 2000s, he worked as software developer using mainly C/C++ and JAVA languages. Since 2008, he attended the MOMILAB under the supervision of Prof. Pietro Pietrini and Prof. Emiliano Ricciardi, acquiring knowledge about analysis of biosignals and neuroimaging data. In 2009, he spent few months at the Laboratory of Neurosciences at the National Institute of Health (NIH, Bethesda, MD, USA) under the supervision of Dr. Maura Furey. He took the degree of Doctor of Medicine in Pisa in 2016. He currently develops machine learning techniques to analyse MRI data using Matlab and C/C++.

## Publications during the Phd

- Avvenuti, G., **Handjaras, G.**, Betta, M., Cataldi, J., Imperatori, L. S., Lattanzi, S., ... & Siclari, F. (2019). Integrity of corpus callosum is essential for the cross-hemispheric propagation of sleep slow waves: a high-density EEG study in split-brain patients. *bioRxiv*, 756676.
- Papale, P., Betta, M., **Handjaras, G.**, Malfatti, G., Cecchetti, L., Rampinini, A., ... & Leo, A. (2019). Common spatiotemporal processing of visual features shapes object representation. *Scientific reports*, 9(1), 7601.
- Cecchetti, L., Lettieri, G., **Handjaras, G.**, Leo, A., Ricciardi, E., Pietrini, P., ... & Train the Brain Consortium. (2019). Brain Hemodynamic Intermediate Phenotype Links Vitamin B12 to Cognitive Profile of Healthy and Mild Cognitive Impaired Subjects. *Neural Plasticity*, 2019.
- Rampinini, A. C., **Handjaras, G.**, Leo, A., Cecchetti, L., Betta, M., Ricciardi, E., ... & Pietrini, P. (2019). Formant space

reconstruction from brain activity in frontal and temporal regions coding for heard vowels. *Frontiers in human neuroscience*, 13, 32.

- Lettieri, G.*, **Handjaras, G.***, Ricciardi, E., Leo, A., Papale, P., Betta, M., ... & Cecchetti, L. (2019). Emotionotopy: Gradients encode emotion dimensions in right temporo-parietal territories. *BioRxiv*, 463166. Manuscript accepted in *Nature Communications*.
- Bernardi, G., Siclari, F., **Handjaras, G.**, Riedner, B. A., & Tononi, G. (2018). Local and widespread slow waves in stable NREM sleep: evidence for distinct regulation mechanisms. *Frontiers in human neuroscience*, 12, 248.
- Benuzzi, F., Ballotta, D., **Handjaras, G.**, Leo, A., Papale, P., Zucchelli, M., ... & Sartori, G. (2018). Eight Weddings and Six Funerals: An fMRI Study on Autobiographical Memories. *Frontiers in behavioral neuroscience*, 12.
- Danti, S., **Handjaras, G.**, Cecchetti, L., Beuzeron-Mangina, H., Pietrini, P., & Ricciardi, E. (2018). Different levels of visual perceptual skills are associated with specific modifications in functional connectivity and global efficiency. *International Journal of Psychophysiology*, 123, 127-135.
- Papale, P., Leo, A., Cecchetti, L., **Handjaras, G.**, Kay, K. N., Pietrini, P., & Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eneuro*, 5(3).
- Rampinini, A. C.*, **Handjaras, G.***, Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G., & Pietrini, P. (2017). Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. *Scientific reports*, 7(1), 17029.
- **Handjaras, G.**, Leo, A., Cecchetti, L., Papale, P., Lenci, A., Marotta, G., ... & Ricciardi, E. (2017). Modality-independent

encoding of individual concepts in the left parietal cortex. *Neuropsychologia*, 105, 39-49.

* Denotes equal first author contribution

For a complete list, please refer to:
*https://scholar.google.it/citations?user=EmbvArAAAAAJ&hl=it*

# Abstract

In the last two decades, neuroscientists have tried to establish the way in which anatomically connected groups of neurons, despite displaying non synchronized neural activity, can work together according to a specific functional architecture. From a methodological perspective, the analysis of such neural organization requires the possibility to measure and integrate the information extracted from large portions of cortex. To this end, recent methodological advancements have prompted the emergence of a new approach, namely multi-voxel pattern analysis (MVPA). Most recent MVPA has also been bred with complex machine learning techniques, which allow to identify whether information is represented in a region (e.g., decoding), and how such information is coded in specific patterns of neural activity (e.g., encoding).

Here, we discuss four MVPA algorithms successfully applied in three different functional Magnetic Resonance Imaging (fMRI) studies. In the first experiment, brain activity of the left fronto-temporal cortex was analyzed using a rank-based multi-class decoding algorithm to identify which brain regions were able to discriminate the seven Italian vowels during their listening, imagery and utterance. Moreover, by means of a canonical correlation analysis, we linearly reconstructed an acoustic, frequency-based model of vowels, using the neural information extracted from the left superior temporal sulcus and the left inferior frontal gyrus. In the second experiment, four models, based on either perceptual or semantic features, were tested to predict brain activity of the left parietal cortex employing a representational similarity encoding algorithm. Finally, in the third fMRI experiment, using a multivariate technique, we were

able to recognize at the individual subject level memories of real autobiographical events, highlighting both the time frame at which the recollection occurred and the brain networks involved in such process.

Overall, these studies tackle the role of machine learning algorithms applied to multivariate patterns of brain activity, and emphasize how the combination of these methods allows an assessment where the information is encoded, spread and organized in the human brain.

# 1. Introduction

*Brief introduction to decoding and encoding*. In recent years, machine learning approaches have been successfully applied to multivariate neuroimaging data (Norman et al., 2006). Machine learning is a relatively novel branch in computer science to achieve computational learning and pattern recognition (Mitchell, 1997). While inferential statistics was conceived to provide evidence at a population level, computational statistics and machine learning aim to learn from data and to make reliable predictions on it.

This new approach has becoming predominant in functional Magnetic Resonance Imaging (fMRI), since, by combining information across multiple voxels, the sensitivity to detect an effect of interest is ultimately increased (Haynes et al., 2015). Moreover, evidence suggested that the neural correlates of stimulus perception as well as of higher cognitive functions (i.e., mental representation) may be grounded in the activity of large ensemble of neurons, sampled across a wide pattern of blood-oxygen-level dependent (BOLD) activity (Haxby et al., 2001; Kriegeskorte et al., 2008). Thus, the shift between analyses performed at single voxel level to analyses carried out on a large extent of voxels (i.e., multi-voxel pattern analysis -MVPA-) is favorable both from a methodological perspective and from a functional one. Indeed, this shift could be seen as the modern counterpart of the conceptual advancement between localism and holistic views of brain functioning during the history of neuropsychology (Norman et al., 2006).

Techniques based on MVPA can be approximately divided in two broad categories, the decoding and encoding algorithms (Haynes et al., 2015). The decoding approach attempts to map

the neural activity into the space defined by stimulus features, whereas encoding does the opposite (Naselaris et al., 2011). In other words, in the encoding approach, one measures the effect of the modulation of the experimental variables on neural activity, whilst in the decoding procedure, one aims at revealing the dimensions represented in neural activity. Even if encoding techniques strictly require the development of specific feature-based models, they are in general favourable, since they can in theory fully describe the neural space, while a decoding approach always offers a partial description. Moreover, a decoding procedure can be easily built upon a successful encoding model while the opposite is not always possible.

In this view, the decoding is generally based on classification algorithms (Pereira et al., 2009) which use information distributed across multiple voxels (as in MVPA), while the encoding adopts *a priori* models crafted by the experimenter to predict neural activity mostly at single-voxel level (Mitchell et al., 2008; Naselaris et al., 2009, Huth et al., 2016).

*Our perspective.* During my PhD, I implemented four different algorithms applied to three fMRI studies. These procedures have already been presented in the scientific literature, but here I adapted their analytical properties to our specific experimental designs and aims and, at the same time, improved their operational robustness. For these purposes, using Matlab (©TheMathWorks, Inc.), I developed:

- a decoding algorithm based on rank accuracy to handle multi-class scenarios, as described in a seminal paper by Mitchell and colleagues (Mitchell et al., 2004) (see Chapter 2);
- a canonical correlation algorithm (Hotelling, 1936), to reconstruct multi-dimensional feature-based models using

information from multiple voxels, aiming to improve the current single-voxel encoding pipelines (Naselaris et al., 2011) (see Chapter 3);
- a representational similarity analysis algorithm (Anderson et al., 2016b) applied across different models and groups (see Chapter 4);
- an improved version of the algorithm originally proposed by Mitchell and colleagues (Mitchell et al., 2008), which merges encoding and decoding procedures in an integrated framework (see Chapter 5).

In addition, all these procedures relied on permutation tests (Schreiber and Krekelberg, 2013) to obtain unbiased, robust estimation of statistical significance and were also developed and coded to limit their computational loads.

*Rank accuracy decoding algorithm*. The first algorithm developed and tested in Rampinini and Handjaras et al., 2017 (see Chapter 2) was adapted from an early work of Mitchell's group (Mitchell et al., 2004). The procedure entails a searchlight (Kriegeskorte et al., 2006) and a rank-based classifier to handle multi-class data (see Figure 1.1). The rank-based algorithm offered many advantages, since it had a chance level centered on 50% even if it was designed to handle multiple classes of stimuli, and it was fast from a computational viewpoint. Thus, rank-based algorithms allowed the use of easily interpretable measures (e.g., sensitivity, specificity) and to plot receiver operating characteristic (ROC) curves (Hand et al., 2001) to interpret the results.

The algorithm requires the acquisition of brain activity of $n$ stimuli pertaining to $m$ classes, where $n$ must be larger than $m$ (e.g., at least two stimuli for each class).

First, a spherical searchlight with a specific radius *r* (i.e., generally 6 to 10 mm) is moved throughout the volume of interest. Each time the sphere shifts in position, its center lays on a specific voxel and the patterns of neural responses elicited by the experimental stimuli are collected within the boundaries of the searchlight. Subsequently, selected response patterns are generally normalized and feed a cross-validation leave-one-stimulus-out algorithm. For each iteration, a distance measure is computed between the pattern of the left-out stimulus and the patterns related to the *m* classes, assembled by averaging the remaining stimuli within-class. Usually, to represent pattern distances, a similarity measure is used (e.g., Pearson's r correlation, Spearman's ρ, or cosine; see Kriegeskorte et al. 2008; Mitchell et al., 2008; Nili et al., 2014).

Second, the collected distances for the left-out stimulus are converted into a rank-ordered list of the potential classes from the least likely category (higher distance, lower similarity, rank *m*) to the most likely (lower distance, higher similarity, rank *1*). The rank list is then adjusted in a rank accuracy measure, so that the chance level is always 50% (corresponding to $m/2$ in the rank-ordered list), regardless of the number of classes involved. Accuracy measures of the stimuli pertaining to each class are averaged and ultimately the procedure generates an accuracy value for each class in each voxel and subject.

Third, group accuracies are then obtained by averaging the accuracy measures across subjects, thus resulting in a group accuracy value at each voxel for each class. To assess the statistical significance, group accuracy values are tested against chance by using a permutation test (Pereira et al., 2009). Briefly, the membership of the stimuli to the classes is shuffled in order to generate *k* (e.g., minimum 1,000 iterations) permuted matrices.

Each permuted matrix is then used in the same searchlight procedure described above. The permutation test generates a set of *k* null accuracies for each class in each voxel and subject. Since the permutation schema is kept fixed across subjects, group-level null accuracies are obtained by simply averaging single subject null distributions (Winkler et al., 2016). Then, a one-sided rank-order test is performed to obtain the empirical p-value for each voxel and class.

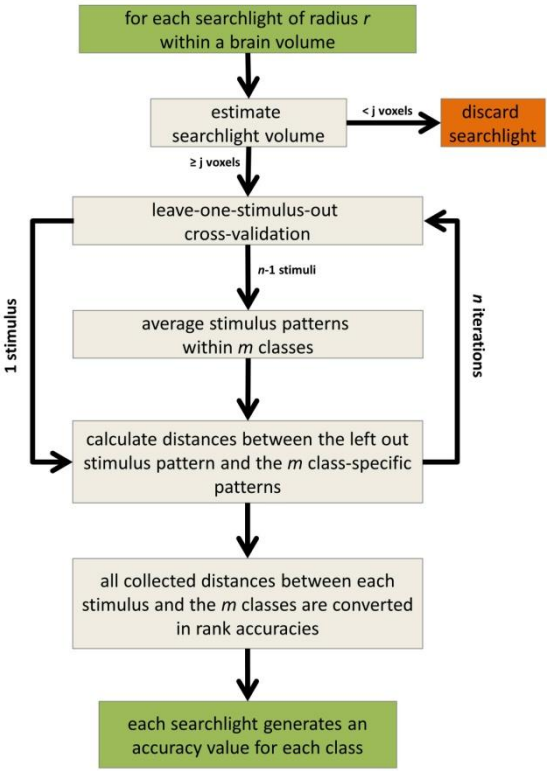


**Figure 1.1.** The flowchart diagram depicts the searchlight procedure combined with a rank-based classifier to handle multi-class data.

Fourth, for the correction of multiple tests, one can adopt a family-wise error rate (FWE) correction or a False Discovery Rate (FDR) procedure (Genovese et al., 2002). Moreover, the permutation test offers two other robust opportunities to correct the results: 1) by directly extracting a null distribution of maximal accuracies across voxels and permutations; 2) by generating a null distribution of the largest clusters obtained when thresholding the null data at a voxel-level p-value of interest (Nichols et al., 2002; Eklund et al., 2015). Then, a one-sided rank-order test is performed to obtain the threshold (at voxel level or related to the minimum cluster size) at the α-value of interest.

*Canonical Correlation Analysis to reconstruct multivariate models*. The second algorithm developed and tested in Rampinini et al., 2019 (see Chapter 3) was conceived to linearly reconstruct stimulus models from BOLD activity in specific regions of interest (ROIs). We selected Canonical Correlation Analysis (CCA; Hotelling, 1936; Bilenko & Gallant, 2016) since it was conceived to find the best associations between two multidimensional variables. In the implementation proposed by Bilenko & Gallant (2016), the authors used CCA as a hyper-alignment technique (Haxby et al., 2011), whereas here we exploited CCA to reconstruct a multidimensional model using information extracted from multiple voxels. Our approach aimed at overcoming the limitations of the current encoding pipelines which used a model to predict neural activity of single voxels.

We first defined X as a matrix $n*f$, where $n$ are the stimuli and $f$ the stimulus features, and Y as a matrix $n*v$, where $n$ are the patterns of brain activity evoked by the stimuli described in X and $v$ are the voxels of a region of interest. Indeed, CCA

provides a set of basis vectors so to maximize the correlations between the projections of the variables of interest (i.e., canonical variables of X, Y) onto these basis vectors.

The X matrix usually contains the descriptors of the stimuli (e.g., acoustic frequencies, semantic features), whereas the Y matrix instead consists of the elicited patterns of BOLD activity, normalized within each voxel. Since Y could be a non full-rank matrix, depending on the number of $v$ voxels as compared to the $n$ stimuli, Singular-Value Decomposition (SVD) is employed before performing CCA. In details, for each subject, the rank of Y was reduced by retaining the first eigenvectors to explain at least 90% of total variance (thus to obtain a $Y_r$, with $n$ rows and $d$ columns, where $d$ is imposed $\geq f$). Subsequently, within each subject, a leave-one-stimulus-out CCA is performed. Specifically, for each iteration, the canonical coefficients and variables -two matrices of $(n-1)*f$ each- are estimated. Since the canonical variables could be rotated if compared to the original matrices X and $Y_r$, within the cross-validation procedure, a procrustes analysis is performed to align the canonical variable of X to X and this linear transformation is retained. Then, for each of the left out stimuli, the canonical coefficients and the transformation matrix from the procrustes analysis are applied to the left-out exemplar of $Y_r$ to obtain a predicted canonical variable of $Y_r$ associated to the features space. As a goodness-of-fit measure, $R^2$ was computed between the group-averaged predicted canonical variable of $Y_r$ and the X matrix (see Figure 1.2).

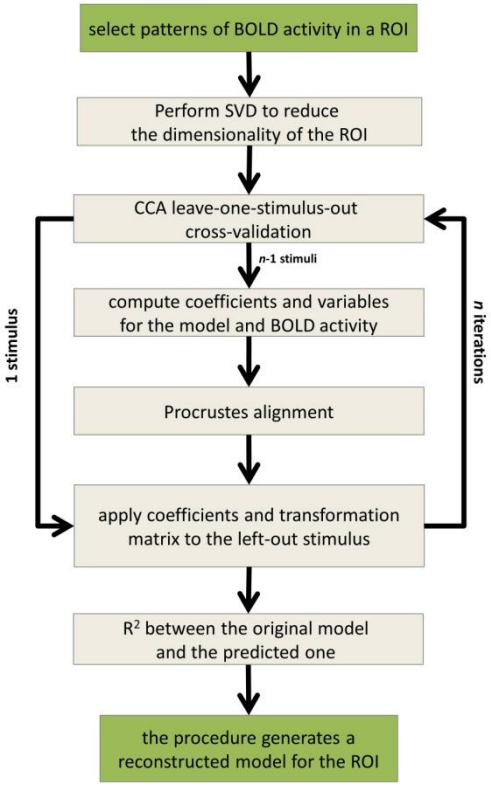


**Figure 1.2.** The flowchart diagram depicts the Canonical Correlation Analysis procedure.

The entire CCA procedure is validated by a permutation test (minimum 1,000 *k* iterations permutations): specifically, for each iteration, the labels of brain activity patterns (i.e., the rows of the Y matrix) are randomly shuffled and subjected to the leave-one-stimulus-out CCA as described above. This procedure provides a $R^2$ null distribution related to the group-level predicted canonical variables. A one-sided rank-order test is then carried out to derive the p-value associated with the original $R^2$ measure.

The main disadvantage of the CCA algorithm is the high computational load required to conduct a whole brain analysis.

For this reason, in Rampinini et al. (2019), we performed the CCA in few ROIs and correction for multiple comparisons was carried out using Bonferroni criterion.

*Representational Similarity Encoding analysis*. The third algorithm developed and tested in Handjaras et al., 2017 (see Chapter 4) was an implementation of the one recently proposed by Anderson and colleagues (2016b). The Representational Similarity Encoding (RSE) merges Representational Similarity Analysis (RSA) and model-based encoding in a unique decoding approach and it is specifically designed to compare the performances of models with different dimensionality. Indeed, model encoding suffers of overfitting issues when high-dimensional models are used as predictors of brain activity and often requires the estimation of several hyper-parameters (Haynes et al., 2015). To overcome these limitations, authors could acquire larger amount of data and adopt cross-validations techniques (Huth et al., 2016) which ultimately increased computational load. However, Anderson and colleagues (20016b) conceived a valid and fast alternative based on RSA. Representational spaces (RSs) are generally derived by measuring stimulus similarities both in the space defined by their descriptions (e.g., semantic space) and in the space defined by the elicited brain activity (i.e., neural space). These two RSs are created by simply comparing each pair of experimental conditions (i.e., stimulus features or patterns of brain activity) using similarity measures (e.g., Pearson's r correlation, Spearman's $\rho$, or cosine; see Kriegeskorte et al. 2008; Mitchell et al., 2008) or even using classical metric ones (e.g., Euclidean or Manhattan distances; see Nili et al., 2014). The results of the procedure is a symmetric matrix *n* by *n* (where *n* are the number

of stimuli) of distances (e.g., 1-r), which serves as a global descriptor of brain regions and models (Kriegeskorte et al., 2008b).

In the RSE approach, first two RSs are created, one from the model space, one from the neural activity of a specific ROI. Then, a leave-two-stimulus-out cross-validation procedure is performed. Briefly, for each iteration, two stimuli are randomly selected and the corresponding rows (i.e., similarity vectors) in the two RSs are retained. Subsequently, the elements related to the two stimuli are removed from the similarity vectors, since they contains zero (i.e., the dissimilarity of the stimulus with itself) or their reciprocal similarity. Then, reduced similarity vectors representing neural and model information for the two left-out stimuli are compared with each other (i.e., Pearson's r) and the score of similarity is converted in an accuracy measure (Mitchell et al., 2008; see Figure 1.3).

Lastly, to assess the significance of the RSE analysis, the resulting accuracy value is tested against the null distribution from a permutation test in which both the neural and behavioral matrices are shuffled (1,000 permutations minimum, one-tailed rank test).

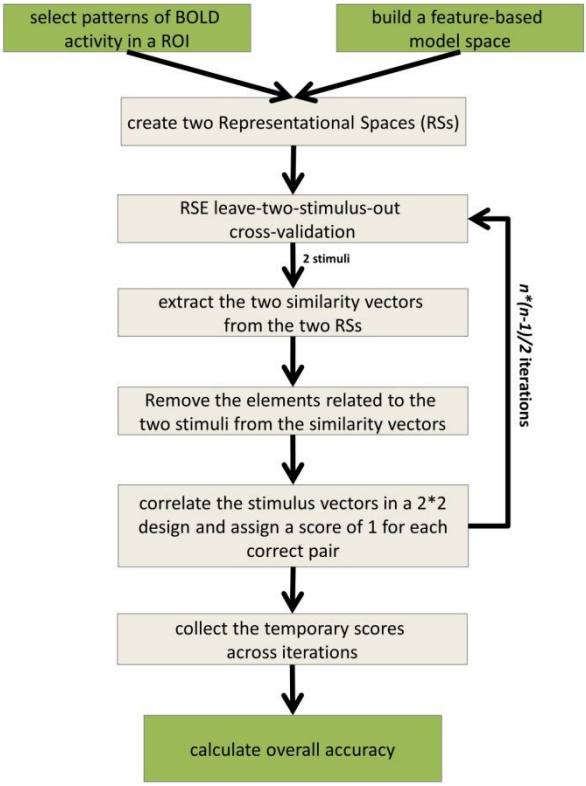


**Figure 1.3.** The flowchart diagram depicts the Representational Similarity Encoding procedure.

*Single subject MVPA using the encoding/decoding pipeline of Mitchell and colleagues (2008)*. The fourth algorithm developed and tested in Benuzzi et al., 2018 (see Chapter 4) was adapted from a pivotal paper of the Mitchell's group (Mitchell et al., 2008).

Briefly, as proposed by Mitchell and colleagues (2008), a machine learning algorithm is used to predict BOLD activity employing encoding dimensions as predictors. Specifically, a least-squares multiple linear regression analysis nested within a

leave-two-stimuli-out cross-validation procedure, generates a set of learned weights able to predict the patterns of brain activity of the two left-out stimuli. Hence, for each iteration, the model is first trained with $n$-2 out of $n$ stimuli, then only $i$ voxels that shows the highest coefficient of determination $R^2$ (e.g., 500) and with a cluster size larger than $j$ voxels (e.g., 20, to remove small isolated clusters; see below) are considered. Once trained, the resulting algorithm is used to predict the fMRI activation within the selected voxels of the two left-out stimuli. Subsequently, accuracy is calculated by means of a decoding procedure, measuring the match between the predicted and the real BOLD patterns of the two left-out stimuli using a similarity measure (see Figure 1.4).

Finally, the single-subject accuracy is tested for significance against the null distribution of accuracies generated with a permutation test by shuffling the labels of the rows of the encoding matrix (Schreiber and Krekelberg, 2013; Handjaras et al., 2015) (one-sided rank test).

The developed algorithm has one major difference with the original one. Indeed, to reduce the computational load, Mitchell et al. (2008) performed the analysis by imposing a predetermined set of voxels outside the cross-validation loop, by preselecting only the brain voxels with a high "stability score" (i.e., low standard deviation across stimuli). This choice could lead in principle to a slight overfit of the data and in general could systematically conceal several brain regions from the analysis (Akama et al., 2018). In our implementation (Benuzzi et al., 2018; Leo et al., 2016; Handjaras et al., 2016), we decided to move the selection of voxels within the cross-validation loop, since the main goal of this algorithm is to measure the discrimination ability of the encoding matrix and not to specifically isolate the

voxels responsible for that. However, this algorithm might lead to small, noisy clusters included in the training steps. To avoid this possibility, we adopted the following countermeasures: 1) a spatial filter to isolate grey matter only regions; 2) a volume correction with an arbitrary minimum cluster size to remove small isolated clusters, which hardly encode model information and likely represent false positives (e.g., overfitting of the training set). Indeed, high-level semantic information (Handjaras et al., 2016), hand-specific motor synergies (Leo et al., 2016) and autobiographical memory (Benuzzi et al., 2018) are encoded in wide patches of cortex. This size is at least two order of magnitude larger than our arbitrary minimum cluster size of twenty voxels (Huth et al., 2016; Hardwich et al., 2018; Svoboda et al., 2006).

Moreover, it should be noted that the choice of voxel space size mapping the encoding matrix is arbitrary, even if several studies estimated this parameter with similar pipelines, at least in semantic tasks (Shinkareva et al., 2011; Chang et al., 2011; Pereira et al., 2013).

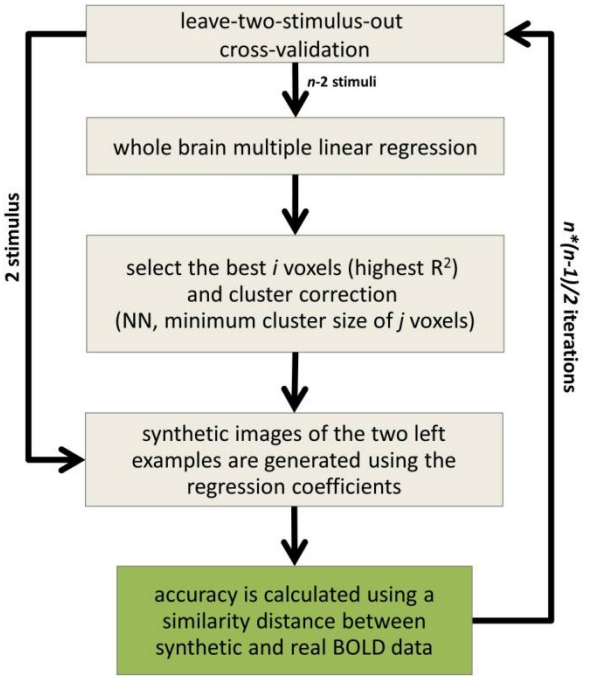


**Figure 1.4.** The flowchart diagram depicts the procedure proposed by Mitchell et al. (2008).

In addition, we introduced another slight deviation from the original methodological pipeline developed by Mitchell et al. (2008). While Mitchell and colleagues used raw fMRI signal as input for the encoding analysis, we extracted the brain hemodynamic activity related to each stimulus after a multiple regression analysis. This procedure was carried out at single-subject level to better control for head movement, baseline activity and drift effects.

Despite these limitations, this algorithm is one of the most used procedures to deal with distributed, sparse representations.

# 2. Decoding vowels using searchlight and rank accuracy algorithm


## Abstract

Classical models of language localize speech perception in the left superior temporal and production in the inferior frontal cortex. Nonetheless, neuropsychological, structural and functional studies have questioned such subdivision, suggesting an interwoven organization of the speech function within these cortices.

We tested whether sub-regions within frontal and temporal speech-related areas retain specific phonological representations during both perception and production. Using functional magnetic resonance imaging and multivoxel pattern analysis, we showed functional and spatial segregation across the left fronto-temporal cortex during listening, imagery and production of vowels. In accordance with classical models of language and evidence from functional studies, the inferior frontal and superior temporal cortices discriminated among perceived and produced vowels respectively, also engaging in the non-classical, alternative function – i.e. perception in the inferior frontal and production in the superior temporal cortex. Crucially, though, contiguous and non-overlapping sub-regions within these hubs performed either the classical or non-classical function, the latter also representing non-linguistic sounds (i.e., pure tones). Extending previous results and in line with integration theories, our findings not only demonstrate that sensitivity to speech listening exists in production-related regions and vice versa, but they also suggest that the nature of such interwoven organization is built upon low-level perceptual features.

## Introduction

According to classical models of speech processing, superior temporal and inferior frontal brain regions are consistently involved in speech perception and production, respectively (Price, 2012). However, theories dealing with the relationship between perceived and produced speech have long debated whether and to what extent perceptual and articulatory information are integrated in language acquisition and use, either assuming that perception shapes production, or that production influences perception (Galantucci et al., 2006).

The phoneme-specific specialization of the superior temporal cortex in perception, as well as the specialization of a wide prefrontal territory around Broca's area in production, are well-known in the literature of phonological competence (Bouchard et al., 2013; Chang et al., 2010). Interestingly, many recent studies revealed that brain activity specific to phonological stimuli could be indeed isolated in the classical foci pertaining to both perception and production, using functional neuroimaging or electrophysiology methods (Rampinini, 2017). In particular, the superior temporal cortex has been shown to represent the overall acoustic form of syllables (Evans et al., 2015), syllable-embedded perceived consonants or vowel categories (Zhang et al., 2016), and even tones when phonologically marked (Feng et al., 2017), while a precise account of motor involvement during production or imagery of phonemes has received less attention in the existing literature (Skipper et al., 2017).

Such rich and mixed picture sparked other questions: do distinct brain regions support different aspects of speech processing (such as perception, imagery and production of phonemes)? Do they share specific phonological representations?

In the context of theories debating an interwoven organization of speech perception and production, the Motor Theory of Speech Perception (MTSP) (Galantucci et al., 2006) argued in favour of a covert articulatory rehearsal mechanism, which would take place implicitly and automatically whenever a speaker is exposed to language, thus connecting the two ends of the perception-production continuum.

In this respect, functional neuroimaging and electrophysiological studies have recently sought to determine the relationship between the perceptual and articulatory stages of speech, seeking perception-related information in frontal areas engaged by production tasks, and production-related information in temporal areas engaged by perception tasks (Tankus et al., 2012; Correia et al., 2015; Cheung et al., 2016; Arsenault et al., 2015; Lee et al., 2012; Markiewicz et al., 2016). In these studies, multivariate analyses were exploited to reveal similarities in informational content between regions previously inferred to perform different functions (through classical activation experiments): a mixed picture of shared information and cortical space as well was assessed, thus tangentially supporting integration models such as those described.

Similarly, virtual and real lesion studies failed to validate an exact correspondence between language impairments and information represented in the frontotemporal speech network: damage in one area may, or may not, entail loss of function in the other, as even sub-regions within such well-known perimeters appear to support different functions (Schomers & Pulvermüller, 2016; Josephs, et al., 2006; Hickok et al., 2011; Ardila et al., 2016). The idea of an interwoven cortical organization of speech function is also favoured by structural studies that reveal a fine-grained cytoarchitectonic, connectivity-

and receptor-mapping-based parcellation of fronto-temporal language areas (Amunts et al., 2010; Anwander, et al., 2007; Catani, et al., 2005; Amunts & Zilles, 2012). Therefore, disentangling the nature of the perception-production interface appears far from straightforward.

According to these indications, we tested whether sub-regions within the frontal and temporal speech areas retain specific, functionally segregated phonological representations during both perception and production, and whether a possible covert rehearsal mechanism could be elicited, through articulation imagery, to simulate the production-perception interface postulated by the MTSP. To this aim, using functional Magnetic Resonance Imaging (fMRI) and multivoxel pattern analysis (MVPA), we measured the spatial overlap of the brain regions involved in stimulus-specific representations during vowel perception (listening), and production (imagined and overt articulation). Within a set of phonemes, the basic units of words, we selected vowels since they retain acoustic features (i.e., formants) that can combine together, thus to distinguish them in a discrete manner. Moreover, formant combinations emerge from unique articulatory gestures, so that their processing depends upon the same perceptuo-motor model (Hardcastle et al., 2010), differently from consonants (Obleser et al., 2010). Particularly, while consonants need to be embedded in syllables to be fully heard and articulated, vowels are self-standing phonemes with high salience. Vowels act as syllabic nuclei, prosodic aggregating centres and, ultimately, can carry stress (whereas consonants cannot), around which the phonic profile of words organizes (Hardcastle et al., 2010). Therefore, vowels offer an interesting perspective to investigate the workings of the perceptual and motor stages of speech.

Thus, building on previous knowledge on phoneme representation in the brain, we tried to provide a finer characterization of the fronto-temporal language cortex: in fact, we compared modalities of perception, production and articulation imagery. Crucially, we also assessed whether sub-regions within the frontal and temporal hubs of the speech network support high-level, fully phonological representations of vowels exclusively, rather than sharing sensitivity to lower-level acoustic stimuli (pure tones), not pertaining to categorical perception of the salient, linguistic kind.

## Materials and Methods

*Participants*. Fifteen right-handed (Edinburgh Handedness Inventory, mean laterality index 0.79±0.17) healthy, mother-tongue Italian monolingual speakers (9F; mean age 28.5±4.6 years) participated in this study, after its approval by the Ethics Committee of the University of Pisa.

*Stimuli.* The seven vowels of the Italian language ([i] [e] [ɛ] [a] [ɔ] [o] [u]) were selected as experimental stimuli, along with seven pure tones (450, 840, 1370, 1850, 2150, 2500, 2900 Hz). Pure tones are physically simpler sounds with no harmonic structure, whereas vowels, despite being periodic waves as well, are endowed with acoustic resonances at specific frequency bandwidths, determined by the vocal tract modifying the source signal produced by the laryngeal mechanism. This structure yields a continuous emission of sound with a fundamental frequency (F0) and a number of overtones called formants (e.g., F1, F2, F3), in a combination that is unique for each vowel. The seven vowels from the Italian phonemic inventory can be disambiguated by the two lower formants F1 and F2, with F0 being constant (Figure 2.1) (Hardcastle et al., 2010).
Three separate, 2s natural voice recordings of each vowel (21 stimuli) were obtained from a female Italian speaker using Praat (©Paul Boersma and David Weenink) a 44100 Hz frequency sampling rate (F0: 191±2.3Hz) and spectrograms were visually inspected for abnormalities. Pure tones were selected by dividing the minimum/maximum mean F1 range of the vowel set into seven, equally distanced bins; the resulting values were approximated to the closest Bark scale value and then converted back to Hertz, so that all tones would lie within the sensitive

perceptual bands in a psychophysical model (Zwicker, 1961). In Audacity (©Audacity Team), seven tones were thus generated using the input-frequencies associated to the Bark value obtained through the aforementioned procedure.

*Experimental procedures.* A slow event-related paradigm was implemented with Presentation (©Neurobehavioral Systems, Inc.) and comprised two perceptual tasks (tone perception and vowel listening), a vowel imagery task and a vowel production one. To increase the amplitude of individual BOLD responses during scan time, all perceived vowels and tones, as well as the execution of imagery and production, were made to last for 2 whole seconds, with the duration signalled by a green fixation cross that would turn black during resting time. All perceptual stimuli (tones or vowels) were thus administered in trials comprising 2s stimulus presentation, then followed by 8s rest. Imagery/production stimuli were administered in trials comprising 2s stimulus presentation, 8s maintenance, 2s task execution and 8s rest. For the imagery task, participants were instructed to perform mental articulation of a heard vowel with their own voice and simulating speech in their mind without ever moving; for the production task, they were instructed to speak naturally and at a normal volume, with rubber wedges and pillows secured so as to avoid head motion without constraining the chin and jaw. In the perceptual tasks (tone perception and vowel listening) subjects were instructed to lay still and listen attentively to the presented stimuli. Globally, functional scans were 47m long, divided in 10 runs. Each of the three vowel recordings was presented twice, thus to obtain 42 trials randomized within and across tasks and subjects, with each sound, either vowel or tone, being equally represented.

BOLD activity was measured using GRE-EPI sequences on a GE Signa 3 Tesla scanner (TR/TE=2500/30ms; FA=75°; 2mm isovoxel; geometry: 128x128x37 axial slices). Brain anatomy was provided by a T1-weighted FSPGR sequence (TR=8.16; TE=3.18ms; FA=12°; 1mm isovoxel; geometry: 256x256x170 axial slices). Stimuli were presented using MR-compatible on-ear headphones (30dB noise-attenuation, 40Hz to 40kHz frequency response).

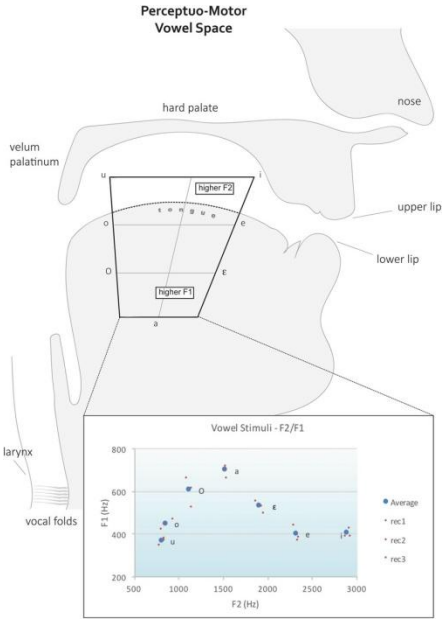


**Figure 2.1.** Vowel acoustic and motor spaces. Here, an ideal representation of the perceptuo-motor vowel space can be appreciated through a sagittal view of the head and phonatory apparatus (top). The articulators are labelled and the relationship that lip and tongue positions entertain with the first and second formant (F1 and F2) can be seen from the trapezoid shape representing the Italian vowel system. Below, the real first- and second formant measurements from our experimental stimuli are plotted in the F1/F2 space, reproducing a projection of the pictured perceptuo-motor vowel space. In this chart, averages for each vowel are represented with blue dots, while measures from single recordings are represented with smaller, red dots (see legend: rec - recording).

*fMRI pre-processing.* The AFNI software package (Cox et al., 1996) was used to pre-process functional MRI data. First, all acquired slices were temporally aligned within each volume (3dTshift), corrected for head movement (3dvolreg), spatially smoothed (3dmerge) with a 4mm FWHM Gaussian filter, and within each voxel every timepoint was divided by the mean of the time series. A multiple regression analysis was then performed on runs (3dDeconvolve), to identify stimulus-related BOLD patterns. Movement parameters and signal trends were included in this procedure as regressors of no interest. Specifically, we used TENT functions for the estimation of BOLD activity (T-values), focusing on the third time point (7.5 seconds) after the acoustic stimulus onset or task execution (imagery or production). By doing this, we aimed at limiting sensory-motor and maintenance-related information, possibly biasing the signal preceding vowel imagery and production (Leo et al, 2016; Connolly et al., 2012). BOLD activity related to the acoustic stimulation in the imagery and production tasks was discarded. Afterwards, T1 images were pre-processed in FSL (Jenkinson, et al., 2012) and nonlinearly registered (Andersson et al., 2007) to the Montreal Neurological Institute (MNI) standard space (2 mm iso-voxel; Fonov et al., 2009); then, the obtained deformation field was used to warp functional maps for each task type.

*Language-sensitive regions.* Hereon, all analyses were performed within a pre-defined topic-based meta-analytic mask of language-sensitive regions. Specifically, the mask was selected from the Neurosynth database (Yarkoni et al., 2011), version 3, topic 21 out of 200, forward inference with a $p<0.01$ (False Discovery Rate -FDR- corrected)(Genovese et al., 2002; Poldrack,

et al., 2012). Keywords included terms related to language and phonological competence, among which were "speech, auditory, sounds, processing, perception, voice, pitch, listening, production, vocal, tones, voices, phonetic, syllable, linguistic, speaker, discrimination, spectral, vowel, language". The extent of the mask was 152,744 $mm^3$ and comprised the bilateral posterior portion of the inferior and middle frontal gyrus, the left precentral sulcus, the bilateral superior temporal cortex, running more posteriorly in the left hemisphere; the left inferior temporal gyrus, supramarginal gyrus and angular gyrus, and the bilateral intraparietal sulcus and middle/inferior occipital gyrus. The mask also included the bilateral caudate nuclei, and the medial portion of the superior frontal gyrus (please refer to Figure 2.2). All analyses, both univariate and multivariate, were performed within this mask.

*Univariate Analysis*. BOLD activity was used to perform one-sample 2-tailed t-test voxel-wise (3dttest++; $p<0.05$, FDR corrected), thus comparing task activity versus rest in each modality.

*Multivariate Analysis*. To assess stimulus discrimination accuracy in each task, the T-value maps were then used in four searchlight-based classifiers (Mitchell et al., 2004; Kriegeskorte et al., 2006) (rank accuracy; cosine similarity; 6mm searchlight radius), one for each task (tone perception, vowel listening, imagery and production). A cross-validation leave-one-stimulus-out procedure was adopted to measure classification accuracy.
Each classifier was conceived to discriminate among seven classes of stimuli: the seven tones in the tone perception task and the seven vowels in the listening, imagery and production tasks.

Accuracies emerging from the tone perception classifier would be used later on, to measure sensitivity to low-level features of acoustic stimuli within clusters defined by the vowel classifiers. Finally, the procedure generated a stimulus discrimination accuracy value for each task, in each voxel and subject. Group accuracies for tone perception, vowel listening, imagery and production were obtained by averaging all single-subject accuracy values, at each voxel.

To assess significance, group accuracies were tested against chance by a permutation test (Winkler, et al. 2016; Pereira et al., 2009; Nichols et al., 2002), where all stimulus-class labels were shuffled in order to generate 1,000 permuted matrices to be used in a multi-class searchlight-based classifier identical to the one described above. The entire procedure generated a set of 1,000 single-subject null discrimination accuracies for each stimulus class, in each voxel, subject and task. Group null accuracies were obtained by averaging single-subject null accuracies in a distribution of 1,000 null accuracies for each voxel and stimulus class. Group accuracy maps were then corrected for multiple comparisons using AFNI: first, real smoothness in the data (resulting from pre-processing, anatomical and searchlight-related smoothing) was estimated (3dFWHMx); later, cluster correction was performed using Monte Carlo simulations (the latest version of 3dClustSim, Gaussian kernel, 10,000 iterations - Cox, et al., 2017). This procedure preserved clusters larger than 1,656 $mm^3$ ($p<0.05$ at voxel level with $\alpha<0.05$ for the correction for multiple comparisons). All the procedures were developed in Matlab (©TheMathWorks, Inc.), unless otherwise specified, through code developed in-house.

*Cross-task accuracies.* To assess whether vowel-sensitive clusters were specific to each task, we measured the averaged accuracies of each task within the masks defined by each of the others (e.g., accuracy of vowel listening within the vowel production mask; 3dROIstats). The same procedure was applied to the null distribution used in the aforementioned permutation test, thus to obtain cluster-based accuracies and their associated statistical significance (1,000 permutations, one-tailed rank test, $p<0.05$). Finally, significance level was adjusted using Bonferroni's correction for multiple comparisons (6 clusters by 3 tasks, $p<0.0028$ for $p_{bonf}<0.05$). The same procedure was employed to assess whether vowel-sensitive clusters represented tone-related information as well, thus to assess their specificity to non-linguistic versus linguistic stimuli; results were Bonferroni-corrected as well (6 clusters by 1 task, $p<0.0083$ for $p_{bonf}<0.05$).

## Results

*Univariate results.* To show regions activated by each of the four tasks, tone perception, vowel listening, imagery and production were subjected to one-sample, two-tailed, voxel-wise t tests against the resting condition ($p<0.05$, corrected for FDR, Genovese et al., 2002), within a topic-based meta-analytic mask of language-sensitive regions selected from the Neurosynth database.

Figure 2.2 shows the results of this procedure and the extension of the mask. Particularly, the tone perception task activated the bilateral primary auditory cortex (Heschl's gyrus, HG) extending to the superior temporal cortex especially in the left hemisphere, along with the superior part of the precentral sulcus (PrCS) at the border with the precentral gyrus (PrCG). In the vowel listening task, HG and superior temporal cortex were activated bilaterally, with more posterior activations in the left hemisphere only; in the frontal cortex, this task activated the left inferior frontal sulcus (IFS) and the opercular portion of the inferior frontal cortex, the insular cortex (INS), and the horizontal ramus of the sylvian fissure, the right *pars opercularis* of the inferior frontal gyrus (IFGpOp), and a small part of the IFS. In the vowel imagery task the frontal cortex was activated in the bilateral (though mostly left) PrCS, left IFS and PrCG, right MFG/IFS and bilateral INS; moreover, this task activated significantly the right STS, left *planum temporale* and supramarginal gyrus (SMG), bilateral, though mostly left, intraparietal sulcus (IPS), left pMTG and inferior temporal gyrus (ITG), the bilateral middle/inferior occipital gyrus (MOG/IOG), and finally, the bilateral medial portion of the superior frontal gyrus (SFG) and caudate nuclei. The vowel production task showed significantly positive BOLD

responses in the bilateral superior temporal cortex extending to the *planum temporale* in the left hemisphere only, in the bilateral INS and PrCS, in left PrCG, in the medial SFG, and in left SMG; in this task, significant negative BOLD responses were observed in the left hemisphere, particularly in the left *pars orbitalis*, the vertical ramus of the sylvian fissure, anterior portion of the medial SFG, anterior and posterior portions of the STS.

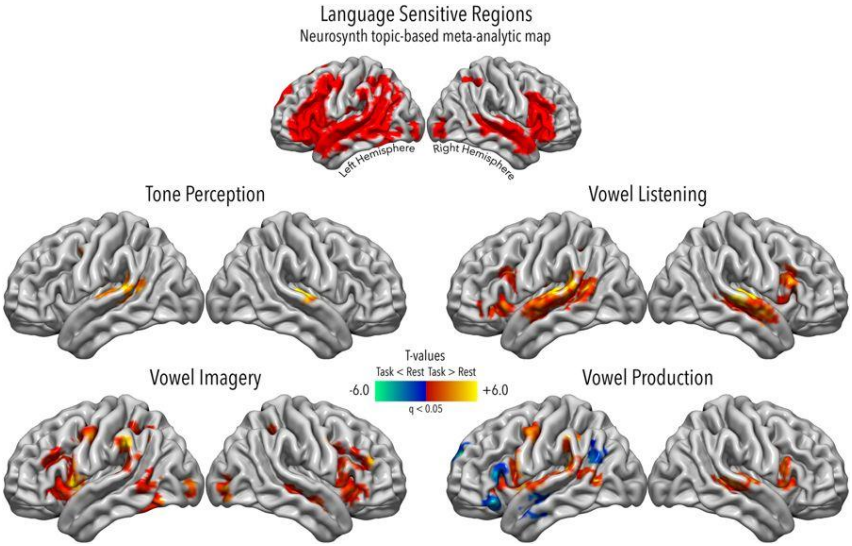


**Figure 2.2.** Univariate results. Here the results for one-sample, two-tailed t-tests are shown in each of the four tasks against the resting condition (dof=14; $p<0.05$, FDR corrected). These measures were conducted to assess which regions were activated in each task and restricted to a topic-based meta-analytic mask of language-sensitive regions from the Neurosynth database, whose extension can be appreciated in the top panel of this figure.

*Multivariate results.* A multi-class searchlight-based classifier highlighted three sets of clusters, one for each vowel task, where pattern discrimination was successful. Figure 2.3 shows clusters on the cortical volume through axial slices, while Figure 2.4

shows the accuracy maps of all experimental tasks projected onto the lateral cortical surfaces.

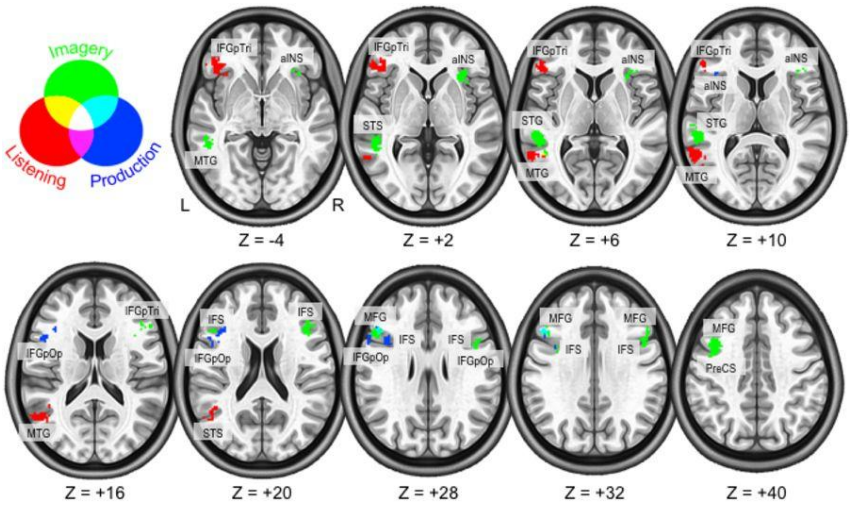


**Figure 2.3.** Here, significant searchlight classifier clusters are shown for the vowel tasks, represented on the cortical volume through axial slices. Colours were assigned by task, and any of their possible combinations were indicated as well in the circle legend. The almost complete contiguity of regions can be appreciated, as marginal overlap emerged only between imagery/production and imagery/listening. No voxels were shared by all three tasks. Labels are spelled as follows: STS - superior temporal sulcus; MTG - middle temporal gyrus; IFGpTri - inferior frontal gyrus, pars triangularis; STG - superior temporal gyrus; IFS - inferior temporal sulcus; MFG - middle frontal gyrus; IFGpOp - inferior frontal gyrus, pars opercularis; aINS - anterior insular cortex.

*Vowel listening, imagery and production dissociate in the left inferior frontal cortex.* The left inferior frontal cortex (IFG, IFS) was engaged across all experimental conditions, with the addition of the right homologue in the imagery task only. Particularly, though, clusters of voxels within these macro-regions responded specifically to each task (regions were labelled and their overlap with the result masks was interpreted

in accordance with the Harvard-Oxford Cortical Atlas). In details, during vowel listening, the *pars triangularis* of the left IFG represented vowels, crossing over anteriorly into the *pars orbitalis*. During vowel imagery, the left IFS and its right homologue intersected superiorly the middle frontal gyrus (MFG), with a relative overlap with the INS as well. During production, a slightly more posterior region within the left IFS was engaged, running inferiorly into the *pars opercularis* of the IFG, and superiorly into the MFG.

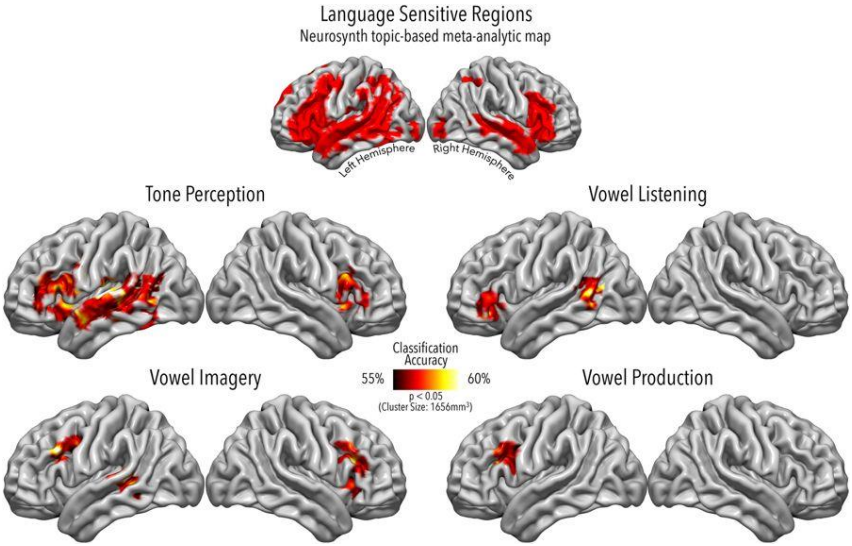


**Figure 2.4.** Accuracy maps projected onto the lateral surfaces of the brain. Here we show regions where accuracy values were significant across the searchlight area defined by the selected Neurosynth topic-based meta-analytic map (top panel) in each task (bottom panels). The extension and location of these regions was validated through cluster correction in AFNI at a minimum cluster size of 207 voxels ($p<0.05$ at voxel level with $\alpha<0.05$ for the correction for multiple comparisons).

*Vowel listening and imagery dissociate in the superior temporal cortex.* Temporal regions representing vowels revealed that the

left STG and STS running posteriorly and inferiorly towards MTG, were engaged in listening, as well as performing imagery of vowels through covert articulation. Particularly, temporal regions representing vowels during listening were the left pSTS, extending into the pMTG. Vowel imagery engaged a nearby portion of the left pMTG extending superiorly into the STG and STS. No temporal regions represented vowels significantly during overt production.

*Measuring cross-task spatial segregation and tone sensitivity.* No spatial overlap among tasks was revealed, except for a cluster of voxels in the IFS/MFG for vowel imagery and production, and a very small cluster in the MTG for vowel imagery and listening. Moreover, cross-task accuracy measurements revealed that the imagery-sensitive left pMTG-STG region also shared tone representations, as well as IFGpTri during vowel listening.
Table 2.1 summarizes cross-task accuracy results from the calculations performed in each cluster from the vowel tasks, with the associated $p$ value and standard errors (SE). Table 2.2 reports cross-task accuracies for the pure tones within the vowel clusters.

| | Task | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Vowel listening | | Vowel imagery | | Vowel production | |
| Mask | *Cluster* | Acc. ± SE | *p* | Acc. ± SE | *p* | Acc. ± SE | *p* |
| VOWEL LISTENING | *Left pSTS-MTG* | 56.7 ± 1.2 | **<0.001*** | 52.8 ± 0.9 | 0.167 | 49.8 ± 1 | 0.958 |
| | *Left IFGpTri* | 57.3 ± 0.8 | **<0.001*** | 54.1 ± 1.4 | 0.018 | 50.3 ± 0.9 | 0.879 |
| VOWEL IMAGERY | *Left pMTG-STG* | 53.1 ± 1 | 0.106 | 57.3 ± 0.9 | **<0.001*** | 51.4 ± 0.7 | 0.645 |
| | *Right IFS-MFG* | 52.2 ± 1.2 | 0.315 | 56.9 ± 1 | **<0.001*** | 52.9 ± 0.9 | 0.112 |
| | *Left IFS-MFG* | 51.7 ± 1.2 | 0.464 | 57.1 ± 1.4 | **<0.001*** | 53.2 ± 1.4 | 0.087 |
| VOWEL PRODUCTION | *Left IFS-IFGpOp* | 52.5 ± 1 | 0.215 | 53.5 ± 0.9 | 0.031 | 56.5 ± 1 | **<0.001*** |

**Table 2.1**. Cross-task accuracy measures between vowel tasks. Accuracy measures are shown here for each task in its own significant regions, but also compared to the other tasks by constraining the extraction of accuracy values for one task within the areas that were significant in each of the others. Significant values are reported in bold, and gray shading was used to highlight accuracy values within correspondent masks and tasks. Of note, accuracy values were significant only for a task within its own regions, showing no

functional overlap between modalities (accuracies were Bonferroni-corrected at $p_{bonf}<0.0028$).

| Mask | Cluster | Tone Perception | |
|---|---|---|---|
| | | Acc. ± SE | *p* |
| VOWEL LISTENING | *Left pSTS-MTG* | 53.7 ± 1.2 | 0.044 |
| | *Left IFGpTri* | 54.9 ± 1.5 | **0.004*** |
| VOWEL IMAGERY | *Left pMTG-STG* | 55.4 ± 1 | **0.002*** |
| | *Right IFS-MFG* | 53.0 ± 1.1 | 0.112 |
| | *Left IFS-MFG* | 51.8 ± 0.8 | 0.466 |
| VOWEL PRODUCTION | *Left IFS-IFGpOp* | 52.9 ± 1.2 | 0.134 |

**Table 2.2.** Cross-task accuracy measures of pure tone perception within each vowel mask. Tone perception accuracy results were constrained within the masks defined by the vowel classifier. Significant values are reported in bold. Of note, the Left IFGpTri from the vowel listening task and the Left pMTG-STG from the vowel imagery task were also able to represent tones significantly (accuracies were Bonferroni-corrected at $p_{bonf}<0.0083$).

## Discussion

In this study we combined fMRI and MVPA to assess the functional organization of vowel listening, imagery and production. We explored the representation of vowels across these three modalities, as well as determining commonalities and differences with a tone perception control task in a frequency range close to that of our speech stimuli. Specifically, patches of cortex in inferior frontal and superior temporal regions retained information to significantly discriminate the seven vowels of the Italian language in each condition. Within these areas, contiguous, and just minimally overlapping clusters were sensitive to listening, imagery and production of speech sounds. Of note, left IFGpTri and left pMTG/STG shared sensitivity to both tones and vowels.

*Functional segregation and tone sensitivity in brain regions involved in vowel listening, imagery and production.* Several functional studies have explored the representation of vowels, consonants and syllables in the fronto-temporal language areas (although more often considering one task at a time): some have highlighted their sensitivity to very fine-grained aspects of speech, such as formant structure, manner and place of articulation, and even speaker identity (Chang, et al., 2010; Formisano et al., 2008; Tankus et al., 2012; Bonte, et al., 2014), while others have highlighted the importance of a shared neural code for validating popular theories about the acquisition and processing of language (Cheung et al., 2016). Univariate results comparing each of the four tasks (tone perception, vowel listening, imagery and production) against resting condition highlighted a set of regions in line with previous findings that

revealed frontal and temporal involvement in language perception and production (Price, 2012). However, while classical univariate approaches sought to infer specific mental function by comparing regional average activations, and thus were amply exploited to investigate the spatial organization of speech, multivariate analyses show representational content similarities over regional engagement: this, together with a comprehensive comparison of speech modalities, can provide a finer characterization of the speech function across the fronto-temporal language cortex.

To provide a finer spatial and functional account of phonological processing and the production-perception interface, we ran a searchlight classifier of listened, imagined and produced vowels within a mask of neuroimaging studies of the language function. This procedure aimed at measuring the accuracy of vowel discrimination, and, most importantly, the spatial organization and possible overlap between regions controlling the three vowel tasks. Moreover, with the same procedure we attempted tone classification in frequencies close to those of our speech stimuli. Accuracies yielded by each vowel task were also measured in clusters resulting from the classifiers of all the other vowel tasks, as well as tone perception accuracies were tested in the vowel regions.

Globally, our results revealed that speech tasks are indeed processed within two classically linguistic macro-regions in the frontal and temporal cortices. Particularly, though, we did not find production of vowels confined to the inferior frontal cortex, nor perception confined to the superior temporal cortex. Instead, both the inferior frontal and superior temporal cortices represented vowel-specific information in both perception and production (imagined as well as overt). Nonetheless, the three

vowel tasks engaged well-defined, bordering sub-portions of the inferior frontal and superior temporal hubs, a picture already sustained by lesion studies and pre-operative language function testing (Long et al., 2016). Moreover, the vowel model was well represented in articulation imagery, a task whose aim was to simulate the articulatory rehearsal mechanism assumed by integration theories: even there, segregated regions revealed sensitivity to vowels in contrast with those clusters, adjoining though non-overlapping, which represented perceived and produced stimuli.

Interestingly, though, while no vowel-sensitive regions retained above-chance accuracies for other tasks, two regions represented tones significantly, that is, the IFGpTri involved in listening and the pSTG-MTG involved in imagery of vowels (of note, the region identified in imagery as being tone-sensitive is spatially closer to the primary auditory cortex than the vowel-specific region identified in vowel listening as pSTS-MTG). This result reveals that, while we have regions *within* the frontal and temporal cortices performing both production-related and perception-related functions in a segregated fashion, these areas also retain low-level non-linguistic information. Specifically, though, high-level information pertains only to the "classical" function associated to that area (production in the inferior frontal and perception in the superior temporal cortex), while the "non-classical" associated function is not language-specific (perception in the inferior frontal and articulation imagery in the superior temporal cortex). Therefore, our findings seem to suggest that the brain retains a capacity for sub-specialization within the classical language fronto-temporal hubs.

*Vowel listening, imagery and production dissociate in the left inferior frontal cortex*. Our results showed how vowel listening, as well as vowel imagery and production, engage the left inferior frontal cortex, from the IFGpOp crossing over anteriorly into the IFGpTri, superiorly into the IFS and touching the MFG. Within the right hemisphere, vowel imagery engaged the IFS, MFG and aINS. However, vowel tasks engaged the broad "Broca's territory" in a functionally segregated fashion: left IFGpOp engaged in vowel production, while the IFS engaged in vowel imagery (as well as its right homologue). Finally, a more anterior region in the IFGpTri engaged in vowel listening although it also represented tones, revealing to be non-specific for speech sounds.

A debate exists on the role of the inferior frontal cortex in processing high- rather than low-level language functions in the healthy brain as well as in lesion studies: this region has been broadly implicated in syntactic working memory (Embick et al., 2000), perceptuo-motor integration (Skipper et al., 2005) and phonetic/phonological representations (Papoutsi et al., 2009; Cheung et al., 2016). Furthermore, along the lines of a functional segregation argument, IFGpOp and IFGpTri within Broca's area have been associated, respectively, to processes pertaining to syntax and semantics (Goucha et al., 2015). Still, early evidence from Positron Emission Tomography (PET) had already suggested that Broca's area is primed by *any* phonological differences subtending semantic representations, and not by the processing of meaning *per se* (Demonet et al., 1992). Moreover, Heim and collaborators do not report additional activations in IFGpTri for semantic *versus* phonological fluency, with only the latter significantly activating IFGpOp (Heim et al., 2008).

Along these lines, some have ascribed the disrupted patterns of both complex syntactic *comprehension* and general speech *production* in Broca's aphasia to a disturbance in the hierarchical chain-processing mechanism at the basis of the phonological loop, which may be controlled by IFGpOp and possibly IFGpTri (Davis et al., 2008). Recently, it was proposed that Broca's area in particular mediates the transformation of perceptual information coming first into the superior temporal cortex, thus to be projected back to the PrCG as articulatory instructions for production (Flinker et al., 2015).

The idea that locations anterior to the PrCG perform sensorimotor transformations and relay information back to the PrCG is in agreement with our findings. Furthermore, we were able to provide a finer characterization of the functional neuroanatomy of the IFG, showing sensitivity to perceived tones and vowels in the *pars triangularis*, and to produced vowels in the *pars opercularis*. Therefore, our results suggest that the language-related inferior frontal cortex, before anything else that may be of a higher level, is concerned *at least* with the representation of perceived speech, as well as non-speech sounds.

The idea that IFGpTri supports simpler, non-linguistic representations, as we found in the cross-task accuracy measurements between vowel listening and tone perception, was previously hinted at by Reiterer and colleagues, who demonstrated IFGpTri involvement in processing tone frequency though not sound pressure, using a pitch *versus* volume discrimination task (Reiterer et al., 2008). On the other hand, Hickok and colleagues reported how IFG-lesioned patients show no *auditory* syllable discrimination deficits whatsoever (Hickok et al., 2011). Although this result may appear in disagreement

with ours, it is reasonable to speculate that the extensions and locations of lesions (as noted by the authors themselves) do not allow for a full comparison with ours and others' functional results in the healthy brain (as also advised by Ardila and colleagues, 2016).

Regarding the *pars opercularis* as the most posterior cluster showing vowel sensitivity, we found produced vowels represented discretely in IFGpOp. In its proximity, the PrCG has been associated to apraxia of speech (Dronkers, 1996), a disturbance in the articulatory aspects of production exclusively. Consistently, we were able to discriminate overtly produced vowels at the posterior border of the IFGpOp extending into the precentral sulcus. Instead, vowel imagery involved more anterior regions for the processing of intermediate phonological representations with no sensory output. These arguments appear to sustain the importance of this inferior frontal region at the perceptuo-motor interface for speech.

All in all, our results suggest that both IFGpOp and IFGpTri do perform *phonological* computations, that is, a sub-lexical kind of processing at the basis of any higher-level function (from syntax to semantics, as already mentioned), and their spatial organization is rather driven by the speech task being performed, with perception and production completely detached, and perception being non-specific for speech sounds.

In fact, some of those trying to reconcile the vast literature on inferior frontal cortex involvement in speech processing have argued that, if its engagement is a matter of perceptuo-motor interface, then the IFG *as a whole* should share activations related to different tasks in the speech loop (Iacoboni et al., 2008). This argument has been brought forward particularly by those sustaining that region sharing would constitute a

neurofunctional correlate of mainframes such as the MTSP (Galantucci et al., 2006). Our results, instead, reveal functional *dissociation* within the inferior frontal cortex for different tasks related to speech sound discrimination, and clarify at least the correlation of *both* IFGpOp and IFGpTri with phonological-level functions.

The processing of produced and imagined speech in close-by regions, as well as more anterior and more rightward activations for imagined speech, were previously reported (Shuster et al., 2005; Huang et al., 2002). In our results, we found a cluster of spatial overlap between the regions involved in produced and imagined vowels in the IFS/MFG. This location's centre of mass was associated to cognitive processes related to working memory in the Neurosynth database (highest posterior probability: 'retrieved' 0.77, 'memory retrieval' 0.76, 'wm task' 0.76). Of note, our subjects were asked to maintain and then retrieve a heard vowel thus to perform imagery or production, and the searchlight analysis was then conducted on the retrieval phase of the trials. In this sense, the small cluster of spatial overlap that we found between production and imagery could be explained as a common *focus* for the mnemonic-attentive component of the task (vowel retrieval). To reinforce this argument, cross-task accuracy measurements did not reveal shared sensitivity for produced *and* imagined vowels in this region, instead showing complete dissociation: in fact, that cluster of spatial overlap may be shared by the production and imagery-sensitive clusters for task-specific demands, and not information content representation.

Finally, the involvement of the right IFS-MFG homologue, as well as aINS, in the imagery task would be justifiable in that these regions were shown to be involved in mental/imagined

speech (Hinke et al., 1993) and aphasia recovery in left IFG/IFS-lesioned patients (Winhuisen et al., 2005).

*Vowel listening and imagery dissociate in the superior temporal cortex.* In our study, the left superior and middle temporal cortices were largely engaged by vowel listening and vowel imagery. Regarding the engagement of the superior temporal cortex in perceived speech, a large body of evidence suggests that this region retains sensitivity to complex harmonic structures and, generally, spectral features down to a stimulus-specific level, studied with both fMRI (Formisano et al., 2008) and ECoG (Chang et al., 2010; Mesgarani et al., 2014). The superior temporal cortex has been associated also to imagery of speech, arguing that the pSTG-pSTS-MTG macro-region supports *both* imagery and perception (Okada et al., 2006; Buchsbaum et al., 2001). Interestingly, though, our results showed that vowel listening and vowel imagery dissociate spatially, as in the inferior frontal cortex; moreover, pSTG-MTG retains tone-specific representations as well as imagined vowels. This reveals how, in the superior temporal cortex as well as the inferior frontal, the function classically associated to the region is language-specific, while the non-classical function shares sensitivity to lower-level stimuli.

Among those who argued in favour of an integrated model, Murakami and colleagues (2015) found that repetitive transcranial magnetic stimulation over the left superior temporal cortex can disrupt phonological fluency, in that it suppresses muscular evoked potential facilitation in the primary motor cortex. This evidence may be of help in characterizing our vowel imagery result in left pSTS-MTG, in that it may validate the idea that mechanisms springing from inferior frontal, speech-

generating areas modulate activity in speech-perceiving ones, during covert articulation (Shergill et al., 2002). It is worth mentioning again that vowels arise from a perceptuo-motor model, with formant structure being determined by unique articulator configurations. Such a model would contain both acoustic and motor information, and thus be represented equally well in superior temporal and inferior frontal areas. These findings are in agreement with previous results obtained with MVPA on functional brain imaging (Formisano et al., 2008) as well as electrocorticographic data (Chang et al., 2010) showing not only that the auditory cortex can encode vowel-specific information during perception, but also, that it can represent articulated speech sounds (Tankus et al., 2012). Particularly, though, HG, the primary auditory cortex, did not show sensitivity to single phonemes (Formisano et al., 2008), as our findings confirm, despite the exquisitely acoustic nature of the task. Nonetheless, in our results, HG was significantly activated during vowel listening (see Figure 2.3), although engaged in representing pure tones (see Figure 2.4): an extrapolation coming from MVPA is that HG is simply not representing vowels in the listening task, despite being activated, as can be seen from Figure 2.2.. Of note, as explained in the Methods section, vowels are aggregates of formants above a fundamental frequency, which are perceived as a summation of the fundamental and the overtones, but also as discrete categories. Such kind of complex stimuli with heightened (linguistic) salience might be computed outside the psychophysically low-level HG (Santoro et al., 2017), as our findings seem to suggest in comparison with simpler tones that are, indeed, represented there. Finally, findings from task-dependent decoding of speaker and vowel identity (Bonte et al., 2014) reveal that the primary auditory cortex in the left

hemisphere actually represents speaker information over vowel information, which seems reasonable when we consider the higher frequential variability of different speakers (across which is the fundamental frequency that changes), rather than the small changes in different vowels uttered by the same speaker, related to harmonic structure over the same fundamental.

Moreover, in Tankus and colleagues (2012), while STG was further probed to assess its ability to discriminate between a complex system of five vowels, the authors also showed how this classically auditory hub of the cortex actually represents *articulated* speech sounds as well. Nevertheless, while neurons in anterior locations such as the medial orbitofrontal cortex (MOF) and the right anterior cingulate cortex (rAC) responded to single or coupled vowels, in this study STG did not, in fact, reveal vowel specificity. In agreement with this study, we found STG activated by vowel production (Figure 2.2), but crucially it did not classify single vowels (Figure 2.4).

Moreover, pSTS-MTG, previously shown to be engaged in articulation imagery over *hearing* imagery (Tian et al., 2016), shared sensitivity to mentally articulated vowels, as well as pure tones, in our data. This is supported by a study reporting conflict between vowel imagery and tone perception in the superior temporal cortex (Kauramäki et al., 2010). As in our findings, the region showing shared sensitivity to lower- and higher-level stimuli was significantly lateralized in the left, language-dominant hemisphere. Moreover, in our results, the patterns of imagined vowels that were represented in left pSTS-MTG could not be ascribed to any acoustic feedback due to the inner nature of the task itself. In this region, tone sensitivity would therefore sustain higher-level representations pertaining to a non-classical

function associated to the location, as well as it did in the inferior frontal cortex.

*Limitations.* Our study presented the following limitations. First, the sample size (n=15), as well as the decoding accuracy (average accuracy in our ROIs reached ~57% across all the seven Italian vowels during the listening task), appeared to be relatively small. However, it should be noted that the first fMRI study which successfully discriminated listened vowels, acquired BOLD activity in seven subjects and obtained an average accuracy of ~63% between three vowels only (i.e., *a*, *i*, *u*; Formisano et al., 2008). Indeed, these three cardinal vowels are commonly represented across languages and retain the highest acoustic and articulatory differences (Hardcastle et al., 2010). Second, the mental imagery task intrinsically required participants' compliance. Third, the experimental design had a fixed inter-stimulus interval (ISI) which may not represent a procedure statistically efficient (Dale, 1999). Nevertheless, we adopted a constant ISI since our machine learning algorithm relied on stimulus decoding across multiple trials and ISI-related differences in hemodynamic responses could have affected its performance.

In conclusion, using fMRI we were able to discriminate the seven vowels of the Italian language in listening, articulation imagery, and production tasks. Globally, these three functions revealed spatial dissociation within language-related brain regions, as well as collateral sensitivity to tone representations: building on previous evidence, these findings provide a finer characterisation of the fronto-temporal language-related cortex. Notably, frontal brain regions classically associated to

production can also represent acoustic features of both linguistic and non-linguistic stimuli; similarly, temporal regions that process low-level acoustic features (pure tones) retain sensitivity to covertly produced vowels. Importantly, in line with integration theories, not only sensitivity to speech listening exists in production-related regions and vice versa, but the nature of such interwoven organisation is also built upon low-level perception.

# 3. Canonical Correlation Analysis to reconstruct acoustic features of vowels


## Abstract

Classical studies have isolated a distributed network of temporal and frontal areas engaged in the neural representation of speech perception and production. With modern literature arguing against unique roles for these cortical regions, different theories have favored either neural code-sharing or cortical space-sharing, thus trying to explain the intertwined spatial and functional organization of motor and acoustic components across the fronto-temporal cortical network. In this context, the focus of attention has recently shifted toward specific model fitting, aimed at motor and/or acoustic space reconstruction in brain activity within the language network. Here, we tested a model based on acoustic properties (formants), and one based on motor properties (articulation parameters), where model-free decoding of evoked fMRI activity during perception, imagery, and production of vowels had been successful. Results revealed that phonological information organizes around formant structure during the perception of vowels; interestingly, such a model was reconstructed in a broad temporal region, outside of the primary auditory cortex, but also in the *pars triangularis* of the left inferior frontal gyrus. Conversely, articulatory features were not associated with brain activity in these regions. Overall, our results call for a degree of interdependence based on acoustic information, between the frontal and temporal ends of the language network.

## Introduction

Classical models of language have long proposed a relatively clear subdivision of tasks between the inferior frontal and the superior temporal cortices, ascribing them to production and perception respectively (Damasio and Geschwind, 1984; Gernsbacher and Kaschak, 2003). Nevertheless, lesion studies, morphological and functional mapping of the cortex evoke a mixed picture concerning the control of perception and production of speech (Josephs et al., 2006; Hickok et al., 2011; Basilakos et al., 2015; Ardila et al., 2016; Schomers and Pulvermüller, 2016).

Particularly, classical theories propose that, on one hand, perception of speech is organized around the primary auditory cortex in Heschl's gyrus, borrowing a large patch of superior and middle temporal regions (Price, 2012); on the other hand, production would be coordinated by an area of the inferior frontal cortex, ranging from the ventral bank of the precentral gyrus toward the pars opercularis and the pars triangularis of the inferior frontal gyrus, the inferior frontal sulcus, and, more medially, the insular cortex (Penfield and Roberts, 1959).

This subdivision, coming historically from neuropsychological evidence of speech disturbances (Poeppel and Hickok, 2004), makes sense when considering that the two hubs are organized around an auditory and a motor pivot (Heschl's gyrus and the face-mouth area in the ventral precentral gyrus), although the issue of their exact involvement already surfaced at the dawn of modern neuroscience (Cole and Cole, 1971; Boller, 1978).

Eventually, the heightened precision of modern, in vivo, brain measures in physiology and pathology ended up supporting

such a complex picture, since an exact correspondence of perception/production speech deficits with the classical fronto-temporal subdivision could not be validated by virtual lesion studies (Fadiga et al., 2002; D'Ausilio et al., 2009, 2012b). Moreover, cytoarchitecture, connectivity and receptor mapping results do suggest a fine-grained parcellation of frontal and temporal cortical regions responsible for speech (Catani and Jones, 2005; Anwander et al., 2006; Fullerton and Pandya, 2007; Hagmann et al., 2008; Amunts et al., 2010; Amunts and Zilles, 2012).

Functional neuroimaging and electrophysiology have therefore recently approached the issue of mapping the exact organization of the speech function, to characterize the fronto-temporal continuum in terms of cortical space-sharing [i.e., engagement of the same region(s) by different tasks] and neural code-sharing (i.e., similar information content across regions and tasks) (Lee et al., 2012; Tankus et al., 2012; Grabski et al., 2013; Arsenault and Buchsbaum, 2015; Correia et al., 2015; Cheung et al., 2016; Markiewicz and Bohland, 2016). Considering this, such studies seemingly align to phonological theory by validating perceptuo-motor models of speech (Schwartz et al., 2012; Laurent et al., 2017), where phonemes embed motor and acoustic information. In fact, vowels are indeed represented by a model based on harmonic properties (formants) modulated by tongue-lip positions: such a model is by all means based on acoustics, but it is also tightly linked to articulation (Ladefoged and Disner, 2012).

Previous fMRI attempts have been made to reconstruct formant space in the auditory cortex (Formisano et al., 2008; Bonte et al., 2014) with a model restricted to a subsample of vowels lying most distant in a space defined by their harmonic

structure. Electrocorticographic recordings have also shown similar results and demonstrated the fine-tuning of the temporal cortex to harmonic structure (Chang et al., 2010; Mesgarani et al., 2014; Chakrabarti et al., 2015). In fact, the possibility of mutual intelligibility along the production-perception continuum, if demonstrated through shared encoding of neural information, might enrich the debate around the neurofunctional correlates of the motor theory of speech perception (MTSP; Liberman et al., 1967), and, more generally, action-perception theories (Galantucci et al., 2006).

In a previous study, a searchlight classifier on fMRI data obtained during listening, imagery and production of the seven Italian vowels, revealed that both the temporal and frontal hubs are sensitive to perception and production, each engaging in their classical, as well as non-classical function (Rampinini et al., 2017). Particularly, though, vowel-specific information was decoded in a spatially and functionally segregated fashion: in the inferior frontal cortex, adjoining regions engaged in vowel production, motor imagery and listening along a postero-anterior axis; in the superior temporal cortex, the same pattern was observed when information relative to perception and motor imagery of vowels was mapped by adjoining regions. Moreover, results from a control task of pure tone perception highlighted the fact that tone sensitivity was also present in the superior temporal and inferior frontal cortices, suggesting a role for these regions in processing low-level, non-strictly linguistic information.

Despite evidence of functional and spatial segregation across the fronto-temporal speech cortex down to the phonological level, a question remained unsolved: which features in the stimuli better describe brain activity in these regions? To

investigate this issue, we sought to reconstruct formant and motor spaces from brain activity within each set of regions known to perform listening, imagery and production of the seven Italian vowels, using data acquired in our previous fMRI study and a multivariate procedure based on canonical correlation (Bilenko and Gallant, 2016).

## Materials and Methods

*Formant Model*. The seven vowels of the Italian language were selected as experimental stimuli (IPA: [i] [e] [ɛ] [a] [ɔ] [o] [u]). While pure tones do not retain any harmonic structure, vowels are endowed with acoustic resonances, due to the modulation of the glottal signal by the vocal tract acting as a resonance chamber. Modulation within the phonatory chamber endows the glottal signal (F0), produced by vocal fold vibration, with formants, i.e., harmonics rising in average frequency as multiples of the glottal signal. Along the vertical axis, first-formant (F1) height correlates inversely with tongue height: therefore, the lower one's tongue, the more open the vowel, the higher frequency of the first formant. The second formant (F2) instead correlates directly with tongue advancement toward the lips. Formant space for the Italian vowels makes it so that each vowel is described by the joint and unique contribution of its first and second formant (Albano Leoni and Maturi, 1995): when first and second formant are represented one as a function of the other, their arrangement in formant space resembles a trapezoidal shape.

Three recordings of each vowel (21 stimuli, each lasting 2 s) were obtained using Praat (©Paul Boersma and David Weenink) from a female, Italian mother-tongue speaker (44100 Hz frequency sampling rate; F0: 191 ± 2.3 Hz). In Praat, we generated spectrograms for each vowel so as to obtain formant listings for F1 and F2, with a time step of 0.01 ms and a frequency step of 0.05 Hz. Average F1 and F2 were obtained by mediating all sampled values within-vowel and are reported, together with the corresponding standard deviations, in Table 3.1 and Figure 3.3. These values were converted from Hertz to

Bark and subsequently normalized defining the formant model for this study.

| *VOWEL* | *F1 (Hz)* | *F2 (Hz)* |
|---|---|---|
| *i* | 305±21.1 | 2170±25.7 |
| *e* | 303±35.9 | 1736±30.7 |
| *e* | 400±27.1 | 1428±47.4 |
| *a* | 525±28.9 | 1139±7.1 |
| *ɔ* | 455±68.1 | 836±34.9 |
| *o* | 338±23.4 | 637±71.6 |
| *u* | 278±16.2 | 604±27.0 |

**Table 3.1.** Average F1 and F2 values and standard deviations for each stimulus

*Articulatory Model.* Structural images of the original speaker's head were used to construct a model based on measurements of the phonatory chamber as in Laukkanen et al. (2012), while the speaker pronounced the vowels. Structural imaging of the speaker uttering three repetitions of each vowel was obtained in a separate session from auditory recording. The speaker was instructed to position her mouth for the selected vowel right before the start of each scan, so as to image steady-state articulation. Scanning parameters were aimed at capturing relevant structures in the phonatory chamber; at the same time, each sequence needed to last as long as the speaker could maintain constant, controlled airflow while keeping motion to a minimum: with this goal, scanning time for each vowel lasted 21 s. Structural T1-weighted images were acquired on a Siemens Symphony 1.5 Tesla scanner, equipped with a 12-channel head coil (TR/TE = 195/4.76 ms; FA = 70˚; matrix geometry: 5 × 384 ×

384, sagittal slices, partial coverage, voxel size 5 mm × 0.6 mm × 0.6 mm, plus 1 mm gap).

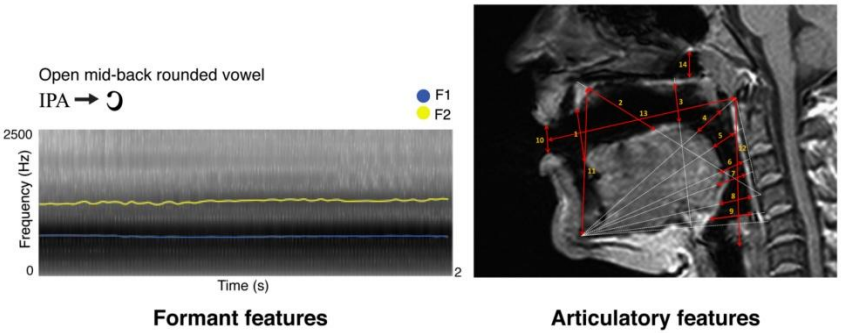


**Figure 3.1.** Here we show a sample vowel by its formant (left) and articulatory (right) representations, as described in Materials and Methods. Formant features represent F1 in blue and F2 in yellow (sampled time step = 0.025 s for display purposes; frequency step unaltered). On the top right, MRI-based articulatory features for the same vowel are indicated by red arrows, with numbers matching the anatomical description of the same measure in Materials and Methods.

Three independent raters performed the MRI anatomical measurements. Particularly, fourteen distances were measured in ITK-SNAP (Yushkevich et al., 2006) as follows: (1) we measured from the tip of the tongue to the anterior edge of the alveolar ridge; (2) we connected the anterior edge of the hard palate to the anterior upper edge of the fourth vertebra, and in that direction we measured from the anterior part of the hard palate to the dorsum of the tongue; (3) we connected the lowermost edge of the jawbone contour to the upper edge of the fifth vertebra, and in that direction we measured from the posterior dorsum of the tongue, to the posterior edge of the hard palate, at a 90˚ angle with the direction line; (4) we connected the lowermost edge of the jawbone contour to the anterior edge of the Arch of Atlas, and in that direction we measured from the anterior tongue body to the soft palate; (5) we connected the

lowermost edge of the jawbone contour to half the distance between the anterior edge of the arch of Atlas and the upper edge of the third vertebra, and in that direction we measured from the posterior tongue body to the back wall of the pharynx; (6) we connected the lowermost edge of the jawbone contour to the upper edge of the third vertebra, and in that direction we measured from the upper tongue root to the back wall of the pharynx; (7) we connected the lowermost edge of the jawbone contour to the longitudinal midpoint of the third vertebra, and in that direction we measured from the lowermost tongue root to the lowermost back wall of the pharynx; (8) we connected the lowermost edge of the jawbone contour to the anterior upper edge of the fourth vertebra and in that direction we measured from the epiglottis to the back wall of the pharynx; (9) we connected the lowermost edge of the jawbone contour and the anterior lower edge of the fourth vertebra, and in that direction we measured from the root of the epiglottis to the back wall of the pharynx; (10) we measured lip opening by connecting the lips at their narrowest closure point; (11) we measured jaw opening by connecting the lowermost edge of the jawbone contour to the anterior end of the hard palate; (12) we measured the vertical extension of the entire vocal tract by tracing the distance between the posterior end of the vocal folds to the anterior lower arch of Atlas; (13) we measured the horizontal extension of the entire vocal tract by tracing the distance between the anterior arch of Atlas to the narrowest closure point between the lips; (14) in the naso-pharynx, we traced the distance between the highest point of the velum platinum and the edge of the sphenoid bone. As an example, Figure 3.1 reports the spectrogram of a vowel obtained in Praat and the MRI

measurements of the phonatory chamber for the same vowel, according to Laukkanen et al. (2012).

Each rater produced a matrix of 21 rows (i.e., seven vowels with three repetitions each) and 14 columns (i.e., the fourteen anatomical distances). For each rating matrix, a representational dissimilarity matrix (RDM, cosine distance) was obtained, and subsequently the accordance (i.e., Pearson's correlation coefficient) between the three RDMs was calculated to assess inter-rater variability. Furthermore, the three RDMs were averaged and non-metric multidimensional scaling was performed to reduce the original 14-dimensional space into two dimensions, thus approximating the dimensionality of the formant model. Finally, the two-dimensional matrix was normalized and aligned to the formant model (procrustes analysis using the rotational component only), to define the articulatory model as reported in Figure 3.3.

*Subjects*. Fifteen right-handed (Edinburgh Handedness Inventory; laterality index 0.79 ± 0.17) healthy, mother-tongue Italian monolingual speakers (9F; mean age 28.5 ± 4.6 years) participated in the fMRI study, approved by the Ethics Committee of the University of Pisa.

*Stimuli*. The seven vowels of the Italian language recorded during the experimental session, for the calculation of the formant model, were used as experimental stimuli (IPA: [i] [e] [ɛ] [a] [ɔ] [o] [u]). Moreover, by dividing the minimum/maximum average F1 range of the vowel set into seven bins, we also selected seven pure tones (450, 840, 1370, 1850, 2150, 2500, 2900 Hz), whose frequencies in Hertz were converted first to the

closest Bark scale value, and then back to Hertz: this way, pure tones were made to fall into psychophysical sensitive bands for auditory perception. Then, pure tones were generated in Audacity (©Audacity Team; see Rampinini et al., 2017 for further details).

*Experimental Procedures*. Using Presentation, we implemented a slow event-related paradigm (©Neurobehavioral Systems, Inc.) comprising two perceptual tasks defined as tone perception and vowel listening, a vowel articulation imagery task and a vowel production task. In perceptual trials, stimulus presentation lasted for 2 s and was followed by 8 s rest. Imagery/production trials started with 2 s stimulus presentation, then followed by 8 s maintenance phase, 2 s task execution (articulation imagery, or production of the same heard vowel) and finally 8 s rest. Globally, functional scans lasted 47 m, divided into 10 runs. All vowels and tones were presented twice to each subject, and their presentation order was randomized within and across tasks and subjects.
Functional imaging was carried out through GRE-EPI sequences on a GE Signa 3 Tesla scanner equipped with an 8-channel head coil (TR/TE = 2500/30 ms; FA = 75°; 2 mm isovoxel; geometry: 128 × 128 × 37 axial slices). Structural imaging was provided by T1-weighted FSPGR sequences (TR/TE = 8.16/3.18 ms; FA = 12°; 1mm isovoxel; geometry: 256x256x170 axial slices). MR-compatible on-ear headphones (30 dB noise-attenuation, 40 Hz to 40 kHz frequency response) were used to achieve auditory stimulation.

*fMRI Pre-processing*. Functional MRI data were preprocessed using the AFNI software package, by performing temporal

alignment of all acquired slices within each volume, head motion correction, spatial smoothing (4 mm FWHM Gaussian filter) and normalization. We then identified stimulus-related BOLD patterns by means of multiple linear regression, including movement parameters and signal trends as regressors of no interest (Rampinini et al., 2017). In FSL (Smith et al., 2004; Jenkinson et al., 2012) T-value maps of BOLD activity related to auditory stimulation (vowels, tones) or task execution (imagery, production) were warped to the Montreal Neurological Institute (MNI) standard space, according to a deformation field provided by the non-linear registration of T1 images of the same standards.

*Previously Reported Decoding Analysis.* In our previous study, this dataset was analyzed to uncover brain regions involved in the discrimination of the four sets of stimuli. Using a multivariate decoding approach based on four searchlight classifiers (Kriegeskorte et al., 2006; Rampinini et al., 2017), we identified, within a pre-defined mask of language-sensitive cortex from the Neurosynth database (Yarkoni et al., 2011), a set of regions discriminating among seven classes of stimuli: the seven tones in the tone perception task and the seven vowels in the listening, imagery and production tasks ($p < 0.05$, corrected for multiple comparisons; see Figure 3.2). Moreover, accuracies emerging from the tone perception classifier had been used to measure sensitivity to low-level features of acoustic stimuli within regions identified by the vowel classifiers.

*Reconstructing Formant and Motor Features From Brain Activity.* While a multivariate decoding approach had successfully detected brain regions representing vowels, it lacked

the ability to recognize the specific, underlying information encoded in those regions, as previous evidence from fMRI had hinted (Formisano et al., 2008; Bonte et al., 2014). We therefore tested here whether the formant and articulatory models were linearly associated to brain responses in the sets of regions representing listened, imagined and produced vowels, as well as pure tones. To this aim, instead of adopting a single-voxel encoding procedure (Naselaris et al., 2011), we selected Canonical Correlation Analysis (CCA; Hotelling, 1936; Bilenko and Gallant, 2016) as a multi-voxel technique which provided a set of canonical variables maximizing the correlation between the two input matrices, X (frequencies of the first two formants of our recorded vowels or, alternatively, the two dimensions extracted from the vocal tract articulatory parameters) and Y (brain activity in all the voxels of a region of interest). Specifically, in the formant model, the X matrix described our frequential, formant-based model in terms of F1 and F2 values of the vowel recordings (three for each vowel, as described in the *Stimuli paragraph*), whereas, in the articulatory model, the X matrix described the phonatory chamber measurements extracted from structural MRI acquired during vowel articulation. The Y matrix instead consisted of the elicited patterns of BOLD activity, normalized within each voxel of each region. Since Y was a non full-rank matrix, Singular-Value Decomposition (SVD) was employed before CCA. In details, for each brain region and subject, the rank of Y was reduced by retaining the first eigenvectors to explain at least 90% of total variance. Subsequently, for each region and within each subject, a leave-one-stimulus-out CCA was performed (Bilenko and Gallant, 2016) thus to obtain two predicted canonical components derived from BOLD activity maximally associated

to the two two-dimensional models. Afterward, predicted dimensions were aligned to the models (procrustes analysis using the rotational component only), and aggregated across subjects in each brain region. As a goodness-of-fit measure, $R^2$ was computed between group-level predicted dimensions and the models. For the formant model, the predicted formants were converted back to Hertz and mapped in the F1/F2 space (Figure 3.3).

The entire CCA procedure was validated by a permutation test (10,000 permutations): specifically, at each iteration, the labels of brain activity patterns (i.e., the rows of the Y matrix, prior to SVD) were randomly shuffled and subjected to a leave-one-stimulus-out CCA in each subject. This procedure provided a null $R^2$ distribution related to the group-level predicted dimensions. A one-sided rank-order test was carried out to derive the p-value associated with the original $R^2$ measure (Tables 3.2–3.5). Subsequently, p-values were corrected for multiple comparisons by dividing the raw p-values by number of tests (i.e., six regions and three tasks, 18 tests).

All the CCA procedure was developed using MATLAB R2016b (MathWorks Inc., Natick, MA, USA), whereas the canonical correlation function (canoncorr) relied on the Matlab implementation.

## Results

*Previous Results.* In a previous study, we sought to decode model-free information content from regions involved in vowel listening, imagery and production, and in tone perception (Rampinini et al., 2017). Using four searchlight classifiers of fMRI data, we extracted a set of regions performing above-chance classification of seven vowels or tones in each task. As depicted in Figure 2, vowel listening engaged the pars triangularis of the left inferior frontal gyrus (IFGpTri), extending into the pars orbitalis. Vowel imagery engaged the bilateral inferior frontal sulcus (IFS) and intersected the middle frontal gyrus (MFG), slightly overlapping with the insular cortex (INS) as well. Production engaged the left IFS though more posteriorly into the sulcus, extending into the pars opercularis of the IFG (IFGpOp), and the MFG. In the temporal cortex, vowel listening engaged the left posterior portion of the superior temporal sulcus and middle temporal gyrus (pSTS-pMTG). Vowel imagery as well engaged a bordering portion of the left pMTG extending superiorly into the superior temporal gyrus (STG) and superior temporal sulcus (STS), while no temporal regions were able to disambiguate vowels significantly during overt production. A small cluster of voxels in the IFS/MFG was shared by vowel imagery and production, as well as another very small one in the middle temporal gyrus (MTG) was shared by imagery and listening. Further testing revealed that the imagery-sensitive left pMTG-STG region also represented pure tones, as well as IFGpTri during vowel listening, while the shared clusters in the IFS-MFG and MTG did not share tone representations.

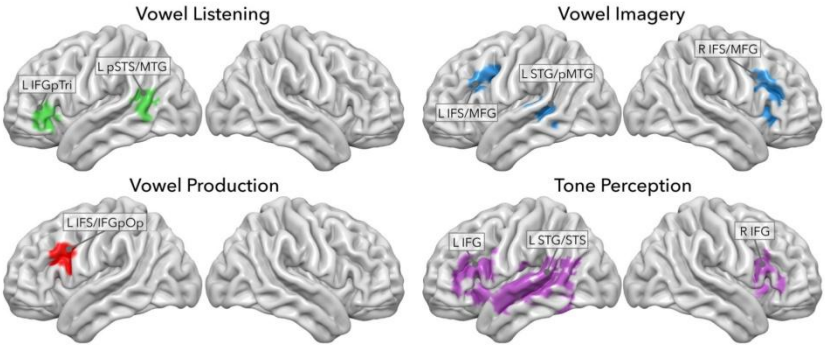


**Figure 3.2.** Searchlight classifier results from Rampinini et al. (2017). Each panel shows regions where model-free decoding was successful in each task.

*Model Quality Assessment.* The articulatory model was constructed by three independent raters, who exhibited an elevated inter-rater accordance (mean = 0.94, min = 0.91, max = 0.96). As depicted in Figure 3.3, both models retain low standard errors between repetitions of the same vowel. Despite the high collinearity between the two models ($R^2$ = 0.90), some discrepancies in the relative distance between vowels can be appreciated in Figure 3.3.

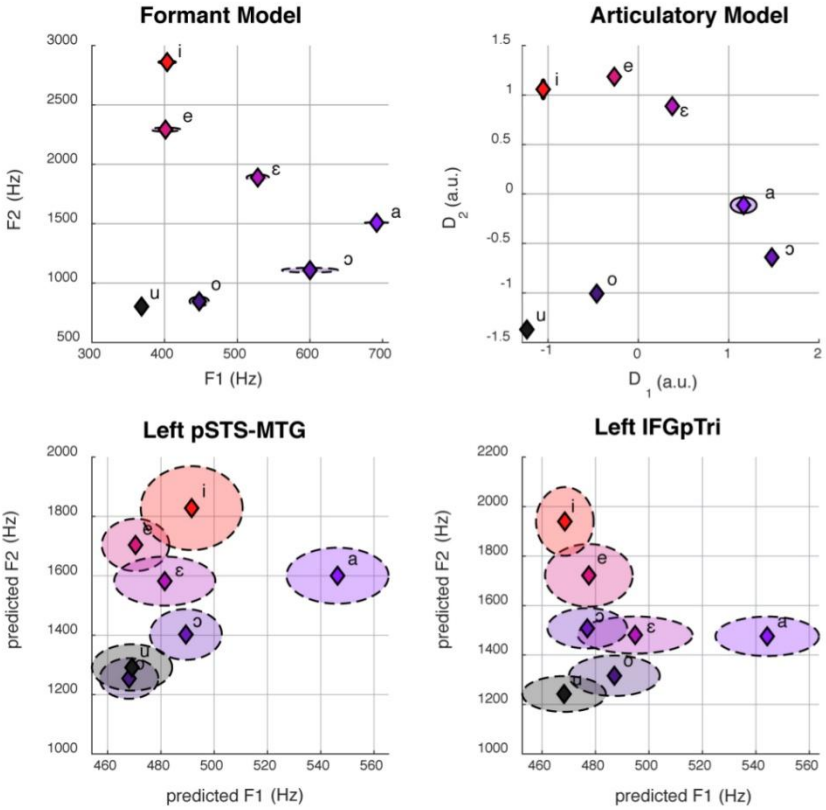


**Figure 3.3.** Here we show formant space (top left) and articulatory space (top right). The bottom panel shows the reconstruction of formant space (bottom left and right) from group-level brain activity in the left pSTS-MTG (left, $R^2 = 0.40$) and IFGpTri (right, $R^2 = 0.39$) through CCA. Dashed ellipses represent standard errors. Articulatory space reconstruction is not reported for lack of statistical significance.

*Current Results*. Here, we employed CCA to assess whether formant and articulatory models, derived from the specific acoustic and articulation properties of our stimuli, could explain brain activity in frontal and temporal regions during vowel listening, articulation imagery, and production. We correlated the formant and articulatory models to brain activity in a region-to-task fashion, i.e., vowel listening activity in vowel listening

regions, imagery activity in imagery regions, and production activity in production regions; moreover, we correlated the models to brain activity from each task, in regions pertaining to all the other tasks (e.g., we tested vowel listening brain data for correlation with the formant and articulatory models not only in vowel listening regions, but also in imagery and production regions). Moreover, brain activity evoked by vowel listening was correlated with the two models in tone perception regions.

*Formant Model.* Globally, the correlation between formant model and brain activity was significant at group level for vowel listening data, in vowel listening regions (uncorrected p = 0.0001; Bonferroni-corrected p < 0.05). As reported in Table 3.2, the left pSTS-MTG yielded an $R^2$ of 0.40 (CI 5th–95th: 0.24–0.52) and left IFGpTri yielded an $R^2$ of 0.39 (CI 5th–95th: 0.20–0.53). For these two regions a reconstruction of vowel waveforms from brain activity was also accomplished (see Supplementary Material in Rampinini et al., 2019). The correlation between formant model and brain data did not reach significance in any other tasks and regions after correction for multiple comparisons. In tone perception regions (i.e., left STG/STS, left IFG and right IFG, see Figure 3.2), the correlation between formant model and brain data did not reach significance (Table 3.3).

| *Region* | | BRAIN ACTIVITY | | |
|---|---|---|---|---|
| | | **VOWEL LISTENING** | VOWEL IMAGERY | VOWEL PRODUCTION |
| ***Vowel Listening*** | ***Left pSTS-MTG*** | **$R^2$=0.402, $p$=0.0001** | $R^2$=0.210, $p$=0.0876 | $R^2$=0.011, $p$=0.7599 |
| | ***Left IFGpTri*** | **$R^2$=0.391, $p$=0.0001** | $R^2$=0.165, $p$=0.1826 | $R^2$=0.125, $p$=0.3244 |
| | *Left pMTG-STG* | $R^2$=0.159, $p$=0.2418 | $R^2$=0.291, $p$=0.0222 | $R^2$=0.113, $p$=0.4285 |
| *Vowel Imagery* | *Right IFS-MFG* | $R^2$=0.234, $p$=0.0706 | $R^2$=0.248, $p$=0.0572 | $R^2$=0.334, $p$=0.0074 |
| | *Left IFS-MFG* | $R^2$=0.133, $p$=0.2845 | $R^2$=0.096, $p$=0.3985 | $R^2$=0.310, $p$=0.0124 |
| *Vowel Production* | *Left IFS-IFGpOp* | $R^2$=0.090, $p$=0.4492 | $R^2$=0.090, $p$=0.4551 | $R^2$=0.262, $p$=0.0359 |

**Table 3.2.** CCA results in regions from vowel listening, imagery and perception (lines), between brain activity in each task (columns) and the formant model.

| Region | | Brain Activity |
|---|---|---|
| | | Vowel Listening |
| *Tone Perception* | *Left STS* | $R^2$=0.169, $p$=0.3077 |
| | *Left IFG* | $R^2$=0.079, $p$=0.6086 |
| | *Right IFG* | $R^2$=0.185, $p$=0.1852 |

**Table 3.3.** CCA results in tone perception regions, between vowel listening brain data and the formant model at group level.

*Articulatory Model.* Globally, the correlation between articulatory model and brain data did not survive correction for multiple comparisons in any tasks or regions (Table 3.4). More importantly, comparison of the formant and motor bootstrap distributions revealed that the acoustic model fit significantly better than the motor model with brain activity in both left pSTS-MTG and left IFGpTri ($p < 0.05$; pSTS-MTG CI 5th–95th: 0.01–0.17; IFGpTri CI 5th–95th: 0.04–0.18; Figure 3.4). Articulatory model correlation with vowel listening brain activity in tone perception regions did not reach statistical significance (Table 3.5).

| Region | | Brain Activity | | |
|---|---|---|---|---|
| | | Vowel Listening | Vowel Imagery | Vowel Production |
| *Vowel Listening* | *Left pSTS-MTG* | $R^2$=0.317, $p$=0.0067 | $R^2$=0.250, $p$=0.0399 | $R^2$=0.106, $p$=0.4195 |
| | *Left IFGpTri* | $R^2$=0.283, $p$=0.0179 | $R^2$=0.068, $p$=0.5224 | $R^2$=0.090, $p$=0.4515 |
| | *Left pMTG-STG* | $R^2$=0.091, $p$=0.4905 | $R^2$=0.256, $p$=0.0649 | $R^2$=0.128, $p$=0.3626 |
| *Vowel Imagery* | *Right IFS-MFG* | $R^2$=0.182, $p$=0.1658 | $R^2$=0.320, $p$=0.0099 | $R^2$=0.299, $p$=0.0189 |
| | *Left IFS-MFG* | $R^2$=0.130, $p$=0.2617 | $R^2$=0.107, $p$=0.3546 | $R^2$=0.292, $p$=0.0159 |
| *Vowel Production* | *Left IFS-IFGpOp* | $R^2$=0.120, $p$=0.3426 | $R^2$=0.072, $p$=0.4825 | $R^2$=0.269, $p$=0.0209 |

**Table 3.4.** CCA results in regions from vowel listening, imagery and perception (lines), between brain activity in each task (columns) and the articulatory model.

| *Region* | | BRAIN ACTIVITY |
|---|---|---|
| | | VOWEL LISTENING |
| *Tone Perception* | *Left STS* | $R^2$=0.120, $p$=0.4952 |
| | *Left IFG* | $R^2$=0.085, $p$=0.5834 |
| | *Right IFG* | $R^2$=0.184, $p$=0.1858 |

**Table 3.5.** CCA results in tone perception regions, between vowel listening brain data and the articulatory model at group level.

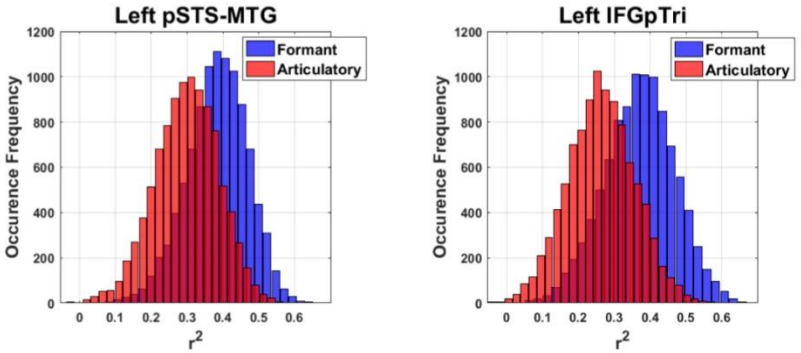


**Figure 3.4.** Bootstrap-based performance comparison between the articulatory and formant models, in regions surviving Bonferroni correction (C.I.: 5–95th of the distribution obtained by computing their difference).

## Discussion

Model-free decoding of phonological information from our previous study, provided a finer characterization of how production and perception of low-level speech units (i.e., vowels) do organize across a wide patch of cortex (Rampinini et al., 2017). Here, we extended those results by testing a frequential, formant-based model and a motor, articulation-based model on brain activity elicited during listening, imagery and production of vowels. As a result, we demonstrated that harmonic features (formant model) correlate with brain activity elicited by vowel listening, in the superior temporal sulcus and gyrus as shown in previous fMRI evidence (Formisano et al., 2008; Bonte et al., 2014). Importantly, here we show that a sub-region of the inferior frontal cortex, the pars triangularis, is tuned to formants during vowel listening. None of the other tasks reflected the formant model significantly, other than IFGpTri-listening and pSTS-MTG-listening. Moreover, despite the high collinearity between the two models, the performance of the articulatory model was never superior to that of the formant model.

*Model Fitting and the Perception-Production Continuum.* The organization of speech perception and production in the left hemisphere has long been debated in the neurosciences of language. In fact, the fronto-temporal macro-region seems to coordinate in such a way that, on one hand, the inferior frontal area performs production-related tasks, as expected from its "classical" function (Dronkers, 1996; Skipper et al., 2005; Davis et al., 2008; Papoutsi et al., 2009), while also being engaged in perception tasks (Reiterer et al., 2008; Iacoboni, 2008; Flinker et

al., 2015; Cheung et al., 2016; Rampinini et al., 2017); in turn, the superior temporal area, classically associated to perception (Evans and Davis, 2015; Zhang et al., 2016; Feng et al., 2017), seems to engage in production as well, despite the topic having received less attention in literature (Okada and Hickok, 2006; Arsenault and Buchsbaum, 2015; Evans and Davis, 2015; Rampinini et al., 2017; Skipper et al., 2017). Finally, sensitivity to tones seems to engage sparse regions across the fronto-temporal speech cortex (Reiterer et al., 2008; Rampinini et al., 2017). This arrangement of phonological information, despite being widely distributed along the fronto-temporal continuum, seems characterized by spatial and functional segregation (Rampinini et al., 2017). Our previous results suggested interesting scenarios as to what "functional specificity" means: in this light, we hypothesized that a model fitting approach would provide insights on the representation of motor or acoustic information in those regions. Therefore, in this study, we assessed whether formant and/or articulatory information content is reflected in brain activity, in regions involved in listening and production tasks, already proven to retain a capacity for vowel discrimination.

It is common knowledge in phonology that a perceptuo-motor model, i.e., a space where motor and acoustic properties determine each other within the phonatory chamber, describes the makeup of vowels (Stevens and House, 1955; Ladefoged and Disner, 2012; Schwartz et al., 2012). This premise could have led to one of the following: two scenarios. First, formant and articulatory information could have been detected in brain activity on an all-out shared basis; therefore, data from all tasks could have reflected both models in their own regions and those from all other tasks, confirming that the acoustic and motor ends

of the continuum indeed weigh the same in terms of cortical processing. Second, a specific task-to-region configuration could have been detected, where information in listening and production regions reflected the formant and articulatory model, respectively. An all-out sharing of formant and articulatory information (i.e., the first scenario) would have pointed at an identical perceptuo-motor model being represented in regions involved in different tasks. A specific task-to-region scenario, instead, would have pointed at a subdivision of information that completely separates listened vowels from imagined or produced ones. Yet again, experimental phonology has long argued in favor of an elevated interdependence between the formant and articulatory models (Stevens and House, 1955; Moore, 1992; Dang and Honda, 2002), which is not new to neuroscience either, with evidence showing perception-related information in the ventral sensorimotor cortex and production-related information in the superior temporal area (Arsenault and Buchsbaum, 2015; Cheung et al., 2016). Thus, it seemed reasonable to hypothesize a certain degree of mutual intelligibility between the frontal and temporal hubs, even maintaining that the two ends of the continuum retain their own specificity of function (Hickok et al., 2011; D'Ausilio et al., 2012a). To what extent though, it remained to be assessed.

In our results, vowel listening data reflected the formant model in a temporal and in a frontal region, providing a finer characterization of how tasks are co-managed by the temporal and frontal ends of the perception-production continuum, in line with the cited literature. Particularly, formant space was reconstructed in pSTS-MTG evoked by vowel listening, as expected from previous literature (Obleser et al., 2006; Formisano et al., 2008; Mesgarani et al., 2014), but also in IFGpTri, again in

the listening task. Yet, the formant model was insufficient to explain brain activity in imagery and production. These results are in agreement with previous associations of the superior temporal cortex with formant structure (Formisano et al., 2008). Moreover, they suggest that frontal regions engage in perception, specifically encoding formant representations. However, such behavior would be modulated by auditory stimulation, despite the historical association of this region to production. Finally, our results show that phonological information, such as that provided by formants, cannot be merely retrieved from tone-processing brain regions.

These results, while contrasting an "all-out shared" scenario for the neural code subtending vowel representation, and not fully confirming a specific "task to region" one, seem to suggest a third, more complex idea: a model based on acoustic properties is indeed shared between regions engaging in speech processing, but not indiscriminately (Grabski et al., 2013; Conant et al., 2018). Instead, its fundamentally acoustic nature is reflected by activity in regions engaging in a listening task, and with higher-level stimuli only (vowels, and not tones). These may contain and organize around more relevant information, like specific motor synergies (Gick and Stavness, 2013; Leo et al., 2016) of the lip-tongue complex (Conant et al., 2018): nonetheless, current limitations in the articulatory model restrict this argument, since, in our data, no production region contained articulatory information sufficient to survive statistical correction. Such discussion might, however, translate from neuroscience to phonology, by providing a finer characterization of vowel space, where apparently kinematics and acoustics do not weigh exactly the same in the brain, despite determining each other in the physics of articulation, as it is commonly taught (Stevens and

House, 1955; Moore, 1992; Dang and Honda, 2002; Ladefoged and Disner, 2012).

*Formants Are Encoded in Temporal and Frontal Regions.* Previous fMRI and ECoG studies already reconstructed formant space in the broad superior temporal region (Obleser et al., 2006; Formisano et al., 2008; Mesgarani et al., 2014). In line with this, we show that even a subtle arrangement of vowels in formant space holds enough information to be represented significantly in both the left pSTS-MTG and IFGpTri, during vowel listening. This presumably indicates that the temporal cortex tunes itself to the specific formant combinations of a speaker's native language, despite its complexity. Moreover, the formant model was explained by auditory brain activity (vowel listening) in regions emerging from the listening task only: one may expect such behavior from regions classically involved in auditory processes, i.e., portions of the superior temporal cortex, as reported by the cited literature; instead, vowel listening also engaged the inferior frontal gyrus in our previous study (Rampinini et al., 2017), and in these results, as well, the formant model was reflected there. This suggests that a region typical to production, as the IFG is, also reflects subtle harmonic properties during vowel listening. Coming back to the hypotheses outlined in the Introduction, these results hint at a degree of code-sharing which is subtler than an all-out scenario or a specific task-to-region one: IFGpTri may perform a non-classical function, only as it "listens to" the sounds of language, retrieving acoustic information in this one specific case. The sensitivity of IFG to acoustic properties is indirectly corroborated by a study from Markiewicz and Bohland (2016), where lifting the informational weight of harmonic structure disrupted the decoding accuracy of vowels

therein. The involvement of frontal regions seems consistent with other data supporting, to a certain degree, action-perception theories (Wilson et al., 2004; D'Ausilio et al., 2012a,b). On the other hand, while an interplay between temporal and frontal areas - already suggested by Luria (1966) -, is supported by computational models (Laurent et al., 2017), as well as by brain data and action-perception theories, the involvement of frontal regions in listening may be modulated by extreme circumstances -as noisy or masked speech- (Adank, 2012; D'Ausilio et al., 2012b), learned stimuli over novel ones (Laurent et al., 2017), or task difficulty (Caramazza and Zurif, 1976). In this sense, IFGpTri representing auditory information may contribute to this sort of interplay. Nonetheless, our results do not provide an argument for the centrality, nor the causality of IFGpTri involvement in perception.

*Articulatory Model Fitting With Brain Activity*. In phonology, the formant model is described as arising from vocal tract configurations unique to each vowel (Stevens and House, 1955; Moore, 1992; Albano Leoni and Maturi, 1995; Dang and Honda, 2002; Ladefoged and Disner, 2012). However, it has to be recognized that practical difficulties in simultaneously combining brain activity measures with linguo- and palatograms have strongly limited a finer characterization of the cerebral vowel space defined through motor markers. Indeed, to this day, the authors found scarce evidence comparing articulation kinematics with brain activity (Bouchard et al., 2016; Conant et al., 2018). Considering the articulatory model, in our data we observed how it simply never outperformed the acoustic model: in fact, it did not survive correction for multiple comparisons, even in production regions. Considering this, the formant model

holds a higher signal-to-noise ratio, coming from known spectro-temporal properties, while the definition of an optimal articulatory model is still open for discussion (Atal et al., 1978; Richmond et al., 2003; Toda et al., 2008). In fact, high-dimensionality representations have frequently been derived by those reconstructing the phonatory chamber by modeling muscles, soft tissues, joints and cartilages (Beautemps et al., 2001). Such complexity is usually managed, as we did here, by means of dimensionality reduction (Beautemps et al., 2001), to achieve whole representations of the phonatory chamber. Although a vowel model described by selecting the first two formants cannot equal the richness and complexity of our articulatory model, the brain does not seem to represent the latter either, in the pars triangularis, or in the pSTS-MTG. Of note, a simpler, two-column articulatory model based on measures maximally correlating with F1 and F2 yielded similar results ($p > 0.05$, Bonferroni-corrected). On the other hand, we point out that our articulatory model was built upon a speaker's vocal tract that, ultimately, was not the same as that of each single fMRI subject. Therefore, even though the formant and articulatory models do entertain a close relationship (signaled by elevated collinearity in our data), caution needs to be exerted in defining them as interchangeable, as shown by literature and in our results with model fitting, which favored an acoustic model in regions emerging from acoustic tasks as reported elsewhere (Cheung et al., 2016).

*Formants and Tones Do Not Overlap*. The superior temporal cortex has long been implicated in processing tones, natural sounds and words using fMRI (Specht and Reul, 2003). Moreover, it seems especially probed by exquisitely acoustic

dimensions such as timbre (Allen et al., 2018), harmonic structure (Formisano et al., 2008), and pitch, even when extracted from complex acoustic environments (De Angelis et al., 2018). There is also evidence of the inferior frontal cortex being broadly involved in language-related tone discrimination and learning (Asaridou et al., 2015; Kwok et al., 2016), as well as in encoding timbre and spectro-temporal features in music (Allen et al., 2018), in attention-based representations of different sound types (Hausfeld et al., 2018) and, in general, in low-level phonological tasks, whether directly (Markiewicz and Bohland, 2016) or indirectly related to vowels (Archila-Meléndez et al., 2018). This joint pattern of acoustic information exchange by the frontal and temporal cortices may be mediated by the underlying structural connections (Kaas and Hackett, 2000) and the existence, in primates, of an auditory "what" stream (Rauschecker and Tian, 2000) specialized in resolving vocalizations (Romanski and Averbeck, 2009). Such mechanism might facilitate functional association between the frontal and temporal cortices when, seemingly, input sounds retain a semantic value for humans (recognizing musical instruments, tonal meaning oppositions, or extracting pitch from naturalistic environments for selection of relevant information).

Coherently, we used tones lying within psychophysical sensitivity bands, within the frequencies of the first formant, a harmonic dimension important for vowel disambiguation, which appeared to be represented across frontal and temporal cortices (Rampinini et al., 2017). Specifically, the left STS and the bilateral IFG represented pure tones, although separate from vowels in our previous study, and here, consistently, no tone-specific region held information relevant enough to reconstruct formant, nor articulatory space. Therefore, this result hinted at the

possibility of more specific organization within these hubs of sound representation.

In our previous study, the *pars triangularis* sub-perimeter coding for heard vowels also showed high accuracy in detecting tone information: in light of this, here we hypothesized the existence of a lower-to-higher-level flow of information, from sound to phoneme. Thus, when formant space was reconstructed from brain activity in the *pars triangularis* coding for heard vowels, we interpreted this result as the need for some degree of sensitivity to periodicity (frequency of pure tones) to represent harmonics (summated frequencies). Therefore, we suggest that harmony and pitch do interact, but the path is one-way from acoustics toward phonology (i.e., to construct meaningful sound representations in one's own language), and not vice versa.

Interestingly, we may be looking at formant specificity as, yet again, a higher-level property retained by few selected voxels within the *pars triangularis*, spatially distinct and responsible for harmonically complex, language-relevant sounds, implying that formant space representation is featured by neurons specifically coding for phonology.

In summary, in the present study we assessed the association of brain activity with formant and articulatory spaces during listening, articulation imagery, and production of seven vowels in fMRI data. Results revealed that, as expected, temporal regions represented formants when engaged in perception; surprisingly, though, frontal regions as well encoded formants, but not vocal tract features, during vowel listening. Moreover, formant representation seems to be featured by a sub-set of voxels responsible specifically for higher level, strictly linguistic coding, since adjoining tone-sensitive regions did not retain formant-related information.

# 4. Representational Similarity Encoding analysis applied to semantic knowledge


## Abstract

The organization of semantic information in the brain has been mainly explored through category-based models, on the assumption that categories broadly reflect the organization of conceptual knowledge. However, the analysis of concepts as individual entities, rather than as items belonging to distinct superordinate categories, may represent a significant advancement in the comprehension of how conceptual knowledge is encoded in the human brain.

Here, we studied the individual representation of thirty concrete nouns from six different categories, across different sensory modalities (i.e., auditory and visual) and groups (i.e., sighted and congenitally blind individuals) in a core hub of the semantic network, the left angular gyrus, and in its neighboring regions within the lateral parietal cortex. Four models based on either perceptual or semantic features at different levels of complexity (i.e., low- or high-level) were used to predict fMRI brain activity using representational similarity encoding analysis. When controlling for the superordinate component, high-level models based on semantic and shape information led to significant encoding accuracies in the intraparietal sulcus only. This region is involved in feature binding and combination of concepts across multiple sensory modalities, suggesting its role in high-level representation of conceptual knowledge. Moreover, when the information regarding superordinate categories is retained, a large extent of parietal cortex is engaged. This result indicates the

need to control for the coarse-level categorial organization when performing studies on higher-level processes related to the retrieval of semantic information.

## Introduction

The organization of semantic information in the human brain has been primarily explored through models based on categories. This domain-specific approach relies on the assumption, supported by neuropsychological and neuroimaging observations, that the categories of language (e.g., faces, places, body parts, tools, animals) broadly reflect the organization of conceptual knowledge in the human brain (Kemmerer, 2016; Mahon and Caramazza, 2009).

However, rather than being limited to differentiating among a small number of broad superordinate categories, a deeper comprehension of conceptual knowledge organization at a neural level should characterize the semantic representation of individual entities (Charest et al., 2014; Clarke and Tyler, 2015; Mahon and Caramazza, 2011). In fact, despite the strong evidence in favor of a categorial organization of conceptual knowledge in the brain (Gainotti, 2010; Pulvermuller, 2013), category-based models tend to be over-simplified and often do not take into account those perceptual and semantic features (e.g., shape, size, function, emotion) involved in the finer-grained discrimination of individual concepts (Clarke and Tyler, 2015; Kemmerer, 2016). Typically, semantic studies limit at controlling those variables within broader and heterogeneous categories, thus restricting the emergence of individual item processing (Baldassi et al., 2013; Bona et al., 2015; Bracci and Op de Beeck, 2016; Ghio et al., 2016; Kaiser et al., 2016; Proklova et al., 2016; Vigliocco et al., 2014; Wang et al., 2016). Furthermore, broader categories are often affected by a high degree of collinearity, as stimuli belonging to highly dissimilar categories according to a sensory-based description (e.g., faces and places), may also be

very dissimilar according to their semantic characterization. Thus, the labeling of certain brain regions might rely either on perceptual or semantic features (Carlson et al., 2014; Fernandino et al., 2016; Jozwik et al., 2016; Khaligh-Razavi and Kriegeskorte, 2014).

In addition, the transition from lower-level sensory-based representations towards higher-level conceptual representations is still ill defined. For instance, how entities that are similar for one or more perceptual features (e.g., shape: a tomato and a ball) are represented in the brain as semantically different remains to be understood (Bi et al., 2016; Clarke and Tyler, 2015; Kubilius et al., 2014; Rice et al., 2014; Tyler et al., 2013; Wang et al., 2016; Wang et al., 2015; Watson et al., 2016).

To assess the extent to which the category-based organization relies on sensory information, our group recently adopted a property generation paradigm in sighted and congenitally blind individuals to demonstrate that the representation of semantic categories relies on a modality-independent brain network (Handjaras et al., 2016). Furthermore, the analysis of individual cortical regions showed that only a few of them (i.e., inferior parietal lobule and parahippocampal gyrus) contained distinct representations of items belonging to different semantic categories across presentation modalities (i.e., pictorial, verbal visual and verbal auditory forms or verbal auditory form in congenitally blind individuals) (Handjaras et al., 2016).

In the present study, we intended to describe the representation, across different presentation modalities, of each of the thirty concrete nouns from six different categories, using part of the same dataset of Handjaras and colleagues (2016). Instead of encoding semantic information using a category-based model, here we characterized the representation of the

individual entities using a recent method for fMRI data analysis, called representational similarity encoding (Anderson et al., 2016b), to combine representational similarity analysis and model-based encoding. In this methodological approach, two representational spaces were created, one from a priori model and one from the neural activity of a specific brain region. Then, a machine learning procedure learned to associate specific rows (i.e., similarity vectors) between the two representational spaces, ultimately generating an overall accuracy measure.

Moreover, the conceptual representation was evaluated by focusing on the entities within each category (e.g., fruits: apple vs. cherry). This within-category encoding is therefore resistant to the effect of category membership and represents an adequate perspective to study how single concepts are processed in the brain. To disentangle the role of perceptual or semantic features and of their complexity (i.e., low- or high-level), we aimed at predicting brain activity using similarity encoding with four models: two semantic models that considered either the complete set of language-based features or a subset of these features related to perceptual properties only (Lenci et al., 2013), and two perceptual models, which provided higher-level descriptions of object shape, or merely focused on low-level visual features (Oliva & Torralba, 2001; van Eede et al., 2006).

We focused the single-item encoding analysis on the angular gyrus and its neighboring regions within the left parietal cortex. The angular region has been solidly associated to a wide gamut of semantic tasks, and its activity during retrieval and processing of concrete nouns or combination of concepts (Binder et al., 2009; Price et al., 2015; Seghier, 2013) makes this region a strong candidate for semantic processing at a finer, single-item level. More importantly, neighboring regions to the angular gyrus

within the left lateral parietal cortex have been involved, to a different extent, in semantic processing, thus indicating the need for a more comprehensive characterization of conceptual representations within the parietal lobe (Binder et al., 2009; Jackson et al., 2016; Price, 2012). Therefore, the analyses were performed in a larger map of the left lateral parietal cortex that centered on the angular gyrus, as defined on both anatomical and functional criteria. The definition of different regions of interest (ROIs) assessed the different degree of involvement of specific regions in processing of individual concepts, and how such a processing is influenced by sensory modality.

## Materials and Methods

A representational similarity encoding (Anderson et al., 2016b) was applied to data collected in a fMRI experiment, in which sighted and blind participants were instructed to mentally generate properties related to a set of concrete nouns, as described in details in our previous study (Handjaras et al., 2016). In brief, participants were divided in four groups according to the stimulus presentation modality (i.e., pictorial, verbal visual and verbal auditory forms for sighted individuals and verbal auditory form for congenitally blind individuals). Two semantic models were built on the set of concrete nouns and two alternative perceptual models were derived from the pictorial form of the stimuli. For each of the semantic and perceptual model, there was a descriptor for high-level features and one for lower-level information. The four models were then used to encode the specific brain activity pattern of each concept, in each group of subjects.

*Brief summary of the Handjaras et al. (2016) fMRI protocol and preprocessing.* Brain activity was measured in fMRI with a slow event-related paradigm (gradient echo echoplanar images GRE-EPI, GE SIGNA at 3T, equipped with an 8-channel head coil, TR 2.5s, FA: 90°, TE 40ms, FOV = 24 cm, 37 axial slices, voxel size 2x2x4 mm) in 20 right-handed Italian volunteers during a property generation task after either visual or auditory presentation of thirty concrete nouns of six semantic categories (i.e., vegetables, fruits, mammals, birds, tools, vehicles) (please refer to Supplementary Materials for the list of nouns). Two semantic categories (e.g., natural and artificial places) from Handjaras et al. (2016) were excluded here due to a specific

limitation of the shape-based perceptual model which required segmented stimuli (e.g., objects). Participants were divided into four groups accordingly to the stimulus presentation format: five sighted individuals were presented with a pictorial form of the forty nouns (M/F: 2/3 mean age ± SD: 29.2±12.8 yrs), five sighted individuals with a verbal visual form (i.e., written Italian words) (M/F: 3/2 mean age ± SD: 36.8±11.9 yrs), five sighted individuals with a verbal auditory form (i.e., spoken Italian words) (M/F: 2/3 mean age ± SD: 37.2±15 yrs) and five congenitally blind with a verbal auditory form (M/F: 2/3 mean age ± SD: 36.4±11.7 yrs). High resolution T1-weighted spoiled gradient recall images were obtained to provide detailed brain anatomy.

During the visual presentation modality, subjects were presented either with images representing the written word (verbal visual form) or color pictures of concrete objects (pictorial form). Stimulus presentation lasted 3 seconds and was followed by a 7s-inter stimulus interval (ISI). During the auditory presentation modality, subjects were asked to listen to about 1s-long words – referring to the same concrete nouns above – followed by 9s ISI. During each 10s-long trial, participants were instructed to mentally generate a set of features related to each concrete noun. Each run had two 15s-long blocks of rest, at its beginning and end, to obtain a measure of baseline activity. The stimuli were presented four times, using, for each repetition, a different image (for pictorial stimuli) or speaker (for auditory stimuli). The presentation order was randomized across repetitions and the stimuli were organized in five runs.

The AFNI software package (Cox, 1996) was used to preprocess functional imaging data. All volumes from the different runs were temporally aligned, corrected for head

movement, spatially smoothed (4 mm) and scaled. Subsequently, a multiple regression analysis was performed to obtain t-score response patterns of each stimulus, which were included in the subsequent analyses. Each stimulus was modeled using five tent functions which covered the entire interval from its onset up to 10 seconds, with a time step of 2.5 seconds. Only the t-score response patterns of the fourth tent function (7.5 seconds after stimulus onset), averaged across the four repetitions, were used as estimates of the BOLD response for each stimulus (Handjaras et al., 2015; Leo et al., 2016). Afterwards, FMRIB's Nonlinear Image Registration tool (FNIRT) was used to register the fMRI volumes to standard space (MNI-152) and to resample the acquisition matrix to a 2 mm iso-voxel (Andersson et al., 2007; Smith et al., 2004).

*Regions of interest*. For our measurement of single-item semantic information, we first defined a mask of the left angular gyrus both using the Automated Anatomical Labeling (AAL) Atlas (Tzourio-Mazoyer et al., 2002) and from a functional meta-analysis using the Neurosynth database (Yarkoni et al., 2011).
Due to the fact that recent evidence shows that semantic processing, albeit mostly centered on the angular gyrus, does involve neighboring regions as well (Binder et al., 2009; Jackson et al., 2016; Price, 2012), we expanded the area of interest to include a larger extent of left parietal cortex, using a mask divided into subregions which could be analyzed separately. First, the functional mask extracted from the Neurosynth database was superimposed to the functional brain atlas by Craddock et al. (2012). A parcellation to 200 ROIs was chosen using the temporal correlation between voxels time-courses as similarity metric; this criterion ensures high anatomic homology

and interpretability (Craddock et al., 2012). At last, eight ROIs were defined in the left lateral parietal cortex, which overlapped, at least partially, with the left angular gyrus defined via Neurosynth meta-analysis (Figure 4.1, 4.3 and Table 4.1).

The bilateral Heschl gyri (HG) and the bilateral calcarine and pericalcarine cortex (Cal) were selected as control regions to assess whether the different presentation modalities could affect primary sensory regions. The HG and Cal regions were defined using the Jülich histological atlas of the FMRIB Software Library (Eickhoff et al., 2007; Smith et al., 2004). In addition, to control for the role of high-level perceptual features, we used the Neurosynth database and the mask obtained from its meta-analytic map to define the left lateral occipital complex (LOC), a region involved in shape processing (Malach et al., 1995). The organization and spatial location of the regions of interest are represented in Table 4.1 and Figure 4.1.

| ID ROI | ID Craddock | Volume | X | Y | Z |
|---|---|---|---|---|---|
| 1 | 6 | 5824 | −56 | −43 | 23 |
| 2 | 22 | 6912 | −27 | −78 | 33 |
| 3 | 27 | 5888 | −24 | −68 | 50 |
| 4 | 49 | 6336 | −43 | −68 | 38 |
| 5 | 99 | 5632 | −52 | −50 | 41 |
| 6 | 110 | 6912 | −49 | −60 | 20 |
| 7 | 130 | 7424 | −57 | −49 | 6 |
| 8 | 188 | 6208 | −43 | −76 | 17 |
| Ang AAL | # | 9176 | −45 | −61 | 36 |
| Ang NS | # | 11424 | −47 | −64 | 33 |

**Table 4.1.** Here are reported Volume (in μL), X, Y and Z coordinates (LPI) in MNI space (in mm) for the center of mass of each region. L Ang AAL and L Ang NS refer to the functional mask of the angular gyrus extracted from the Neurosynth database (Yarkoni et al., 2011) and the anatomical definition of the angular gyrus using the Automated Anatomical Labeling (AAL) Atlas (Tzourio-Mazoyer et al., 2002) respectively. ID ROI indicates the number of each region of Figure 4.3 with the corresponding identification number (ID Craddock) from the atlas by Craddock et al. (2012).

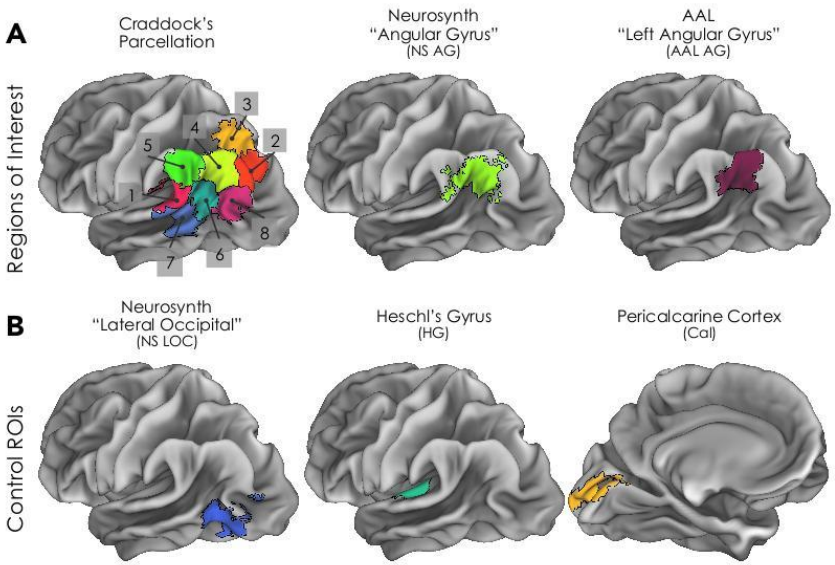


**Figure 4.1.** As regions of interest, the left lateral parietal cortex was parcellated using the brain atlas by Craddock et al. (2012), while the functional and the anatomical masks of the angular gyrus were extracted from the Neurosynth database (Yarkoni et al., 2011) and the Automated Anatomical Labeling (AAL) Atlas respectively (Tzourio-Mazoyer et al., 2002) (Panel A). As control regions, we defined the left lateral occipital complex (LOC) using the Neurosynth database, and the bilateral Heschl gyri (HG) and the bilateral calcarine and pericalcarine cortex (Cal) using the Jülich histological atlas (Eickhoff et al., 2007) (Panel B). These regions were also detailed in Table 4.1.

*Semantic models.* The Blind Italian Norming Data (BLIND) set, validated in an independent Italian sample of blind and sighted participants, was used to define the semantic model for the similarity encoding (Lenci et al., 2013). The concrete nouns of the BLIND study were a set of normalized stimuli that belong to various biological and artificial semantic categories, most of which are shared with previous norming studies (Connolly et al., 2007; Kremer and Baroni, 2011; McRae et al., 2005). In the BLIND study, sighted and congenitally blind participants were presented with concept names and were asked to verbally list the features that describe the entities the words refer to. The features

produced by the subjects were not limited to sensory attributes of the stimuli (e.g., shape, size, color) but also included high-level properties, such as associated events and abstract features (Lenci et al., 2013). The collected features were extracted, pooled across subjects to derive averaged representations of the nouns, using subjects' production frequency as an estimate of feature salience (Handjaras et al., 2016; Lenci et al., 2013; Mitchell et al., 2008). This procedure provided a feature space of 812 dimensions (properties) for sighted and 743 for blind participants. As depicted in Figure 4.2, the collected features were used to assemble two semantic models for both sighted and blind individuals: one based on the whole feature space (i.e., high-level semantic model), one restricted to the perceptual features only (i.e., Property of Perceptual Type, PPE), corresponding to those qualities that can be directly perceived, such as magnitude, shape, taste, texture, smell, sound and color (i.e., low-level semantic model) (Wu & Barsalou, 2009; Lenci et al., 2013).

Subsequently, representational spaces (RSs) were derived from the semantic models using correlation dissimilarity index (one minus Pearson's r), obtaining four group-level dissimilarity matrices (i.e., for sighted and blind subjects) (Figure 4.2).

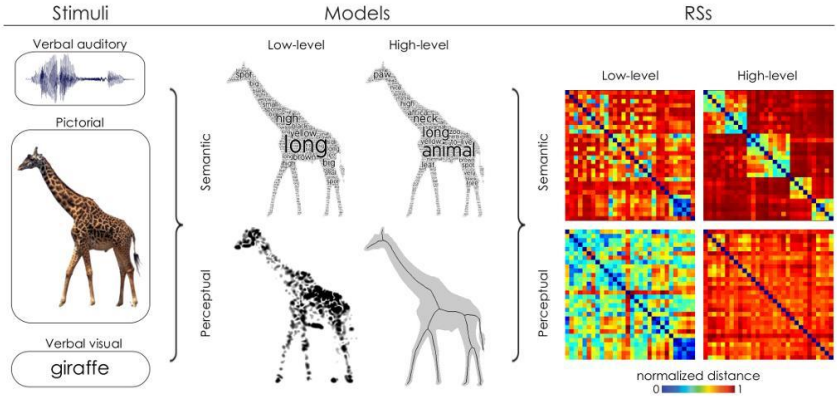


**Figure 4.2.** Figure depicts, on the left, the different presentation modalities used to evoke conceptual representations (pictorial, verbal visual and verbal auditory forms for sighted individuals and verbal auditory form for congenitally blind individuals). In the middle, the four models used for the encoding analyses are defined. Two semantic models, illustratively represented using word clouds, were built on the features generated in a behavioral experiment based on a property-generation task (Lenci et al., 2013): the high-level model was based on the whole set of linguistic features while the low-level one was defined on a subset of these features restricted to perceptual properties. Moreover, two perceptual models were obtained from the pictorial form of the stimuli: the high-level perceptual model was built on the shape features of the images through shock-graphs (Sebastian et al., 2004), while the low-level one was the GIST based on Gabor filters (Oliva and Torralba, 2001). For example, according to the high-level semantic model a screwdriver was very similar to a hammer, while according to the high-level shape-based perceptual model, a screwdriver was more similar to a pencil than to a hammer. The Representational Spaces (RSs) extracted from the four models are depicted on the right. Dissimilarity measures are reported in details in the Methods section.

*Perceptual models*. A high-level perceptual model was obtained from the shape features of the thirty images. First, all the pictorial stimuli were manually segmented and binarized. A skeletal representation of each stimulus was then computed by performing the medial axis transform (Blum, 1973). The dissimilarity between each pair of skeletal representations was then computed using the ShapeMatcher algorithm

(http://www.cs.toronto.edu/~dmac/ShapeMatcher/index.html ; van Eede et al., 2006) which builds the shock-graphs of each object and then estimates their pairwise distance by computing the deformation needed in order to match their shapes (Sebastian et al., 2004). The distances were then averaged across the four repetitions of each pictorial stimulus, which corresponded to four different pictures, to produce a shape-based RS. This high-level perceptual description was used as a model to predict brain activity, similarly to what is performed on fMRI data by other authors (Leeds et al., 2013).

Furthermore, to assess whether the patterns of neural response could be predicted also by differences in low-level image statistics of the different pictorial stimuli, we built a RS based on visual features (Oliva & Torralba, 2001; Rice et al., 2014). A global description of the spatial frequencies of each color image seen by the subjects during the pictorial presentation modality was estimated using the GIST model (Oliva and Torralba, 2001). Briefly, a GIST descriptor was computed by sampling the responses to Gabor filters with four different sizes and eight orientations; the GIST descriptor of each item was obtained by averaging the GIST descriptors of the four stimuli representing the item. The GIST descriptor of each item were then normalized and compared to each other using correlation dissimilarity index, generating a RS which was used as a low-level, perceptual model.

For each RS of the four models, the within-category information was extracted, normalized within each category scaling to the maximum distance and compared across models ($p<0.05$, two tailed test, Bonferroni corrected for the number of comparisons, i.e., 15) (Table 4.2). Subsequently, within-category information of each model was used for the similarity encoding.

*Representational similarity encoding analysis.* The similarity encoding was recently proposed to merge representational similarity analysis and model-based encoding (Anderson et al., 2016b). In this approach, two RSs, one derived from neural and one from semantic or perceptual data, are compared each other using a leave-two-stimulus-out strategy: the two left out vectors from both matrices are matched using the correlation coefficient hence to generate an accuracy measure. This approach is resistant to overfitting issues and does not require parameters estimation (for further details, please refer to Anderson et al, 2016b).

The RSs from fMRI data were computed within each ROI and subject, using the correlation distance. For each presentation modality, the five single-subject RSs were averaged and the resulting group-level RSs were compared to the models RS as specified above. The analysis was limited to the five concrete nouns within each of the six categories, thus performing only 60 comparisons (i.e., within-category individual item encoding) instead of all the 435 comparisons (i.e., among-categories individual item encoding).

The standard error of the accuracy value was estimated using a bootstrapping procedure (1,000 iterations) (Efron & Tibshirani, 1994). Finally, to assess the significance of the encoding analysis, the resulting accuracy value was tested against the null distribution from a permutation test in which both the neural and behavioral matrices were shuffled (1,000 permutations, one-tailed rank test).

Moreover, within each ROI, accuracies of each presentation modality were averaged. The significance level was calculated by averaging null distributions obtained with a fixed permutation

schema across presentation modalities (Nichols et al., 2002). The averaged accuracy was subsequently tested with a one-tailed rank test (1,000 permutations).

Accuracies across presentation modalities were reported in Table 4.3, 4.4, 4.5 and 4.6, while the averaged accuracy across presentation modalities was represented onto a brain mesh in Figure 4.3. All the p-values of the accuracies in Table 4.3, 4.4, 4.5 and 4.6 were reported as uncorrected for multiple comparisons. Results from the left parietal cortex were corrected for Bonferroni when applicable (by adjusting the raw p-values evaluating the eight ROIs from the Craddock Atlas).

The model definition and the similarity encoding approaches were accomplished by using Matlab (Matworks Inc., Natick, MA, USA), while Connectome Workbench was used to render the brain meshes in Figure 4.1, 4.3, and 4.4B.

In addition, an alternative procedure based on the discrimination of each individual concrete noun irrespective of their membership to one of the six semantic categories (i.e., among-categories individual item encoding) was performed using the high-level semantic model only: this procedure aimed at measuring the impact of the categorial organization on the classification accuracy (see Supplementary Materials of Handjaras et al., 2017).

## Results

The combined procedure to identify the angular gyrus on an anatomical and functional bases, and to parcellate the surrounding portion of left lateral parietal cortex using the brain atlas by Craddock et al. (2012), resulted in eight ROIs that comprised a wide extension of cortex from the posterior and middle part of intraparietal sulcus (IPS) to superior temporal lobule, angular and supramarginal gyri, as well as superior temporal gyrus, as depicted in Figure 4.1, and detailed in Table 4.1.

The within-category RSs obtained from the four models were compared to each other to assess models' collinearity ($p<0.05$, Bonferroni corrected). Results were reported in Table 4.2.

| | high-level semantic model in sighted | high-level semantic model in blind | low-level semantic model in sighted | low-level semantic model in blind | high-level perceptual model | low-level perceptual model |
|---|---|---|---|---|---|---|
| high-level semantic model in sighted | | $r=0.68$, $p<0.0001^*$ | $r=0.23$, $p=0.0714$ | $r=0.16$, $p=0.2231$ | $r=0.28$, $p=0.0301$ | $r=-0.14$, $p=0.3005$ |
| high-level semantic model in blind | | | $r=0.05$, $p=0.7253$ | $r=0.37$, $p=0.0037$ | $r=0.19$, $p=0.1391$ | $r=-0.01$, $p=0.9214$ |
| low-level semantic model in sighted | | | | $r=0.33$, $p=0.0095$ | $r=0.10$, $p=0.4430$ | $r=-0.06$, $p=0.6521$ |
| low-level semantic model in blind | | | | | $r=0.19$, $p=0.1459$ | $r=0.23$, $p=0.0740$ |
| high-level perceptual model | | | | | | $r=-0.07$, $p=0.6126$ |
| low-level perceptual model | | | | | | |

**Table 4.2.** Table reports the Pearson's r correlation coefficient between each model.
* Indicates a significant correlation ($p<0.05$, Bonferroni corrected).

The blind and the sighted within-category high-level semantic models were highly correlated ($r=0.68$, $p<0.05$, Bonferroni corrected). This is consistent with the high correlation value of the whole semantic RS between blind and sighted participants ($r=0.94$) previously reported (Handjaras et al., 2016). The other models retained relative lower, not significant correlations ($p>0.05$, Bonferroni corrected).

| ID ROI | Sighted Pictorial | Sighted Verbal Visual | Sighted Verbal Auditory | Blind Verbal Auditory | Average across modalities |
|---|---|---|---|---|---|
| 1 | 57.5 ± 4.2%, p = 0.150 | 51.7 ± 5.3%, p = 0.384 | 45.0 ± 3.8%, p = 0.746 | 47.5 ± 4.4%, p = 0.602 | 50.4 ± 2.7%, p = 0.459 |
| 2 | 62.5 ± 3.8%, p = 0.034 | 60.0 ± 4.4%, p = 0.059 | 69.2 ± 3.6%, p < 0.001* | 63.3 ± 3.1%, p = 0.034 | 63.8 ± 1.9%, p < 0.001* |
| 3 | 67.5 ± 4.2%, p = 0.008 | 56.7 ± 4.1%, p = 0.168 | 54.2 ± 4.4%, p = 0.254 | 61.7 ± 4.1%, p = 0.049 | 60.0 ± 2.9%, p = 0.005* |
| 4 | 53.3 ± 4.7%, p = 0.297 | 55.8 ± 4.7%, p = 0.184 | 60.8 ± 4.2%, p = 0.062 | 50.0 ± 4.8%, p = 0.491 | 55.0 ± 2.3%, p = 0.089 |
| 5 | 59.2 ± 4.8%, p = 0.107 | 40.8 ± 3.8%, p = 0.873 | 50.0 ± 4.6%, p = 0.468 | 44.2 ± 3.5%, p = 0.756 | 48.5 ± 4.0%, p = 0.651 |
| 6 | 64.2 ± 4.2%, p = 0.024 | 50.8 ± 3.7%, p = 0.402 | 41.7 ± 4.4%, p = 0.863 | 52.5 ± 4.7%, p = 0.347 | 52.3 ± 4.6%, p = 0.240 |
| 7 | 56.7 ± 5.0%, p = 0.159 | 60.0 ± 4.0%, p = 0.074 | 40.0 ± 3.7%, p = 0.910 | 45.8 ± 5.3%, p = 0.683 | 50.6 ± 4.7%, p = 0.415 |
| 8 | 54.2 ± 4.1%, p = 0.283 | 58.3 ± 3.9%, p = 0.098 | 48.3 ± 4.1%, p = 0.575 | 59.2 ± 4.2%, p = 0.090 | 55.0 ± 2.5%, p = 0.073 |
| Ang AAL | 49.2 ± 4.9%, p = 0.530 | 46.7 ± 4.4%, p = 0.640 | 61.7 ± 4.1%, p = 0.052 | 57.5 ± 4.5%, p = 0.146 | 53.8 ± 3.5%, p = 0.137 |
| Ang NS | 58.3 ± 4.4%, p = 0.114 | 51.7 ± 5.5%, p = 0.373 | 55.0 ± 4.1%, p = 0.227 | 53.3 ± 4.2%, p = 0.308 | 54.6 ± 1.4%, p = 0.108 |
| HG | 47.5 ± 3.3%, p = 0.605 | 56.7 ± 4.8%, p = 0.172 | 51.7 ± 4.4%, p = 0.372 | 50.8 ± 4.5%, p = 0.444 | 51.7 ± 1.9%, p = 0.316 |
| Cal | 45.0 ± 3.8%, p = 0.734 | 56.7 ± 4.4%, p = 0.179 | 59.2 ± 4.5%, p = 0.075 | 56.7 ± 4.1%, p = 0.170 | 54.4 ± 3.2%, p = 0.120 |
| LOC | 58.3 ± 3.7%, p = 0.121 | 59.2 ± 3.8%, p = 0.087 | 40.0 ± 3.2%, p = 0.902 | 49.2 ± 4.3%, p = 0.509 | 51.7 ± 4.5%, p = 0.325 |

**Table 4.3.** Within-category individual item encoding accuracies for the high-level semantic model. Here are reported the accuracies in each ROI of the encoding procedure in each presentation modality (mean±standard error) for the semantic model based on the whole linguistic feature space. For Ang AAL, Ang NS, LOC, HG and Cal, please refer to Figure 4.1. * Indicates a successful encoding at $p<0.05$, Bonferroni corrected for the eight ROIs from the brain atlas by Craddock et al. (2012).

The within-categories encoding analysis, performed in the left lateral parietal cortex, indicated a significant ability to discriminate individual concrete nouns using the high-level models (semantic and shape-based perceptual) in the posterior part of the IPS (ROI 2) and in the middle portion of the IPS, extending to the superior parietal lobule (ROI 3). Specifically, in ROI 2, we found an accuracy (average accuracy across presentation modalities ± standard error) of 63.8±1.9% for the semantic high-level model, 59.0±5.2% for the shape-based perceptual model (both $p<0.05$, Bonferroni corrected), while the low-level models resulted in a not significant accuracy: 54.8±5.1% for the semantic model based on the perceptual features only and 42.1±3.9% for the GIST-based perceptual one (both $p>0.05$).

| ID ROI | Sighted Pictorial | Sighted Verbal Visual | Sighted Verbal Auditory | Blind Verbal Auditory | Average across modalities |
|---|---|---|---|---|---|
| 1 | 44.2 ± 4.4%, p = 0.771 | 40.0 ± 5.8%, p = 0.903 | 53.3 ± 4.9%, p = 0.307 | 52.5 ± 4.5%, p = 0.327 | 47.5 ± 3.2%, p = 0.743 |
| 2 | 61.7 ± 4.6%, p = 0.037 | 44.2 ± 3.5%, p = 0.754 | 68.3 ± 4.7%, p = 0.006* | 61.7 ± 3.9%, p = 0.052 | 59.0 ± 5.2%, p < 0.001* |
| 3 | 58.3 ± 5.3%, p = 0.110 | 58.3 ± 4.5%, p = 0.111 | 65.0 ± 4.1%, p = 0.015 | 59.2 ± 4.3%, p = 0.095 | 60.2 ± 1.6%, p = 0.003* |
| 4 | 54.2 ± 4.1%, p = 0.274 | 55.8 ± 4.5%, p = 0.185 | 55.0 ± 4.3%, p = 0.256 | 48.3 ± 3.6%, p = 0.572 | 53.3 ± 1.7%, p = 0.203 |
| 5 | 53.3 ± 4.3%, p = 0.288 | 40.0 ± 4.5%, p = 0.892 | 55.8 ± 3.8%, p = 0.185 | 56.7 ± 4.1%, p = 0.154 | 51.5 ± 3.9%, p = 0.328 |
| 6 | 47.5 ± 4.6%, p = 0.617 | 50.0 ± 4.9%, p = 0.474 | 53.3 ± 4.9%, p = 0.317 | 46.7 ± 3.2%, p = 0.634 | 49.4 ± 1.5%, p = 0.548 |
| 7 | 62.5 ± 3.7%, p = 0.036 | 38.3 ± 4.3%, p = 0.945 | 57.5 ± 4.1%, p = 0.148 | 40.8 ± 5.0%, p = 0.878 | 49.8 ± 6.0%, p = 0.499 |
| 8 | 54.2 ± 3.8%, p = 0.269 | 57.5 ± 4.1%, p = 0.139 | 56.7 ± 4.7%, p = 0.158 | 51.7 ± 4.1%, p = 0.381 | 55.0 ± 1.3%, p = 0.095 |
| Ang AAL | 52.5 ± 4.4%, p = 0.344 | 45.0 ± 4.7%, p = 0.701 | 58.3 ± 4.1%, p = 0.121 | 50.0 ± 4.5%, p = 0.457 | 51.5 ± 2.8%, p = 0.345 |
| Ang NS | 49.2 ± 5.0%, p = 0.535 | 55.0 ± 4.0%, p = 0.215 | 60.8 ± 4.6%, p = 0.068 | 47.5 ± 4.3%, p = 0.617 | 53.1 ± 3.0%, p = 0.202 |
| HG | 35.8 ± 4.8%, p = 0.972 | 50.0 ± 4.3%, p = 0.479 | 62.5 ± 4.1%, p = 0.036 | 50.0 ± 4.6%, p = 0.446 | 49.6 ± 5.4%, p = 0.542 |
| Cal | 61.7 ± 4.2%, p = 0.041 | 48.3 ± 4.7%, p = 0.558 | 45.8 ± 5.3%, p = 0.699 | 56.7 ± 4.4%, p = 0.152 | 53.1 ± 3.7%, p = 0.207 |
| LOC | 58.3 ± 3.6%, p = 0.109 | 53.3 ± 4.5%, p = 0.287 | 47.5 ± 3.8%, p = 0.609 | 65.0 ± 3.5%, p = 0.016 | 56.0 ± 3.7%, p = 0.040 |

**Table 4.4.** Within-category individual item encoding accuracies for the high-level perceptual model. Here are reported the accuracies in each ROI of the encoding procedure in each presentation modality (mean±standard error) for the perceptual model based on shape features. For Ang AAL, Ang NS, LOC, HG and Cal, please refer to Figure 4.1. * Indicates a successful encoding at p<0.05, Bonferroni corrected for the eight ROIs from the brain atlas by Craddock et al. (2012).

| ID ROI | Sighted Pictorial | Sighted Verbal Visual | Sighted Verbal Auditory | Blind Verbal Auditory | Average across modalities |
|---|---|---|---|---|---|
| 1 | 56.7 ± 4.1%, p = 0.163 | 44.2 ± 4.8%, p = 0.765 | 54.2 ± 4.1%, p = 0.287 | 48.3 ± 4.5%, p = 0.558 | 50.8 ± 2.8%, p = 0.413 |
| 2 | 58.3 ± 4.7%, p = 0.108 | 45.8 ± 4.9%, p = 0.697 | 67.5 ± 4.6%, p = 0.005* | 47.5 ± 3.8%, p = 0.631 | 54.8 ± 5.1%, p = 0.090 |
| 3 | 63.3 ± 4.6%, p = 0.029 | 64.2 ± 5.4%, p = 0.014 | 60.0 ± 4.3%, p = 0.077 | 58.3 ± 3.6%, p = 0.133 | 61.5 ± 1.4%, p = 0.004* |
| 4 | 57.5 ± 4.6%, p = 0.154 | 53.3 ± 4.3%, p = 0.291 | 60.0 ± 4.6%, p = 0.079 | 48.3 ± 4.0%, p = 0.565 | 54.8 ± 2.6%, p = 0.091 |
| 5 | 56.7 ± 4.1%, p = 0.161 | 50.8 ± 5.1%, p = 0.411 | 42.5 ± 4.3%, p = 0.851 | 41.7 ± 4.5%, p = 0.869 | 47.9 ± 3.6%, p = 0.742 |
| 6 | 54.2 ± 4.2%, p = 0.271 | 40.8 ± 4.3%, p = 0.874 | 45.8 ± 4.5%, p = 0.685 | 41.7 ± 4.5%, p = 0.851 | 45.6 ± 3.1%, p = 0.895 |
| 7 | 52.5 ± 4.2%, p = 0.360 | 46.7 ± 4.4%, p = 0.652 | 45.8 ± 5.2%, p = 0.694 | 50.0 ± 4.6%, p = 0.470 | 48.8 ± 1.5%, p = 0.650 |
| 8 | 54.2 ± 4.1%, p = 0.253 | 68.3 ± 4.4%, p = 0.005* | 53.3 ± 4.8%, p = 0.306 | 52.5 ± 4.9%, p = 0.334 | 57.1 ± 3.8%, p = 0.034 |
| Ang AAL | 64.2 ± 4.2%, p = 0.025 | 46.7 ± 5.2%, p = 0.651 | 55.8 ± 3.9%, p = 0.180 | 48.3 ± 4.1%, p = 0.575 | 53.8 ± 4.0%, p = 0.130 |
| Ang NS | 58.3 ± 4.3%, p = 0.114 | 45.8 ± 4.6%, p = 0.706 | 57.5 ± 4.7%, p = 0.139 | 44.2 ± 3.3%, p = 0.769 | 51.5 ± 3.7%, p = 0.335 |
| HG | 52.5 ± 4.5%, p = 0.319 | 44.2 ± 4.6%, p = 0.775 | 60.0 ± 4.1%, p = 0.069 | 45.8 ± 4.3%, p = 0.682 | 50.6 ± 3.6%, p = 0.410 |
| Cal | 56.7 ± 4.2%, p = 0.150 | 40.8 ± 4.6%, p = 0.895 | 60.0 ± 4.4%, p = 0.081 | 65.8 ± 3.8%, p = 0.021 | 55.8 ± 5.3%, p = 0.063 |
| LOC | 67.5 ± 3.4%, p = 0.008 | 62.5 ± 4.2%, p = 0.041 | 32.5 ± 3.9%, p = 0.994 | 50.0 ± 4.4%, p = 0.496 | 53.1 ± 7.8%, p = 0.203 |

**Table 4.5.** Within-category individual item encoding accuracies for the low-level semantic model. Here are reported the accuracies in each ROI of the encoding procedure in each presentation modality (mean±standard error) for the semantic model based on perceptual features only. For Ang AAL, Ang NS, LOC, HG and Cal, please refer to Figure 4.1.
* Indicates a successful encoding at p<0.05, Bonferroni corrected for the eight ROIs from the brain atlas by Craddock et al. (2012).

Similarly, in ROI 3, encoding analysis led to a significant accuracy for the high-level models (60.0±2.9% for the semantic and 60.2±1.6% for the perceptual one, both p<0.05, Bonferroni corrected) and the low-level semantic-based model (61.5±1.4%, p<0.05, Bonferroni corrected), while the low-level perceptual one was at chance level (47.1±3.7%, p>0.05, Bonferroni corrected).

These results were reported in details in Table 4.3, 4.4, 4.5 and 4.6 and Figure 4.3.

| ID ROI | Sighted Pictorial | Sighted Verbal Visual | Sighted Verbal Auditory | Blind Verbal Auditory | Average across modalities |
|---|---|---|---|---|---|
| 1 | 55.8 ± 4.7%, p=0.186 | 59.2 ± 4.4%, p=0.086 | 44.2 ± 3.8%, p=0.785 | 41.7 ± 4.4%, p=0.866 | 50.2 ± 4.3%, p=0.493 |
| 2 | 30.8 ± 4.2%, p=0.997 | 48.3 ± 4.5%, p=0.569 | 42.5 ± 4.0%, p=0.836 | 46.7 ± 5.3%, p=0.646 | 42.1 ± 3.9%, p=0.993 |
| 3 | 46.7 ± 4.1%, p=0.669 | 57.5 ± 3.8%, p=0.124 | 40.8 ± 4.3%, p=0.874 | 43.3 ± 4.2%, p=0.812 | 47.1 ± 3.7%, p=0.817 |
| 4 | 44.2 ± 4.2%, p=0.770 | 50.0 ± 4.0%, p=0.511 | 36.7 ± 3.5%, p=0.958 | 53.3 ± 3.8%, p=0.284 | 46.0 ± 3.7%, p=0.888 |
| 5 | 47.5 ± 5.5%, p=0.606 | 49.2 ± 3.8%, p=0.522 | 47.5 ± 4.5%, p=0.601 | 40.8 ± 4.5%, p=0.894 | 46.2 ± 1.8%, p=0.848 |
| 6 | 52.5 ± 4.4%, p=0.366 | 58.3 ± 4.6%, p=0.117 | 52.5 ± 3.9%, p=0.343 | 55.8 ± 5.0%, p=0.204 | 54.8 ± 1.4%, p=0.100 |
| 7 | 45.8 ± 4.3%, p=0.695 | 50.8 ± 4.1%, p=0.419 | 33.3 ± 3.8%, p=0.988 | 50.8 ± 3.9%, p=0.438 | 45.2 ± 4.1%, p=0.911 |
| 8 | 54.2 ± 5.1%, p=0.287 | 46.7 ± 5.5%, p=0.672 | 45.8 ± 4.5%, p=0.688 | 60.0 ± 4.7%, p=0.077 | 51.7 ± 3.4%, p=0.349 |
| Ang AAL | 60.8 ± 3.8%, p=0.067 | 50.0 ± 4.0%, p=0.503 | 43.3 ± 3.8%, p=0.793 | 48.3 ± 4.8%, p=0.567 | 50.6 ± 3.7%, p=0.432 |
| Ang NS | 52.5 ± 3.4%, p=0.344 | 60.8 ± 3.5%, p=0.059 | 43.3 ± 3.7%, p=0.817 | 55.8 ± 4.3%, p=0.200 | 53.1 ± 3.7%, p=0.211 |
| HG | 53.3 ± 5.1%, p=0.312 | 50.0 ± 5.2%, p=0.488 | 43.3 ± 4.2%, p=0.817 | 50.8 ± 4.4%, p=0.407 | 49.4 ± 2.1%, p=0.578 |
| Cal | 59.2 ± 4.4%, p=0.089 | 42.5 ± 5.2%, p=0.82 | 49.2 ± 4.4%, p=0.521 | 45.8 ± 3.6%, p=0.702 | 49.2 ± 3.6%, p=0.600 |
| LOC | 54.2 ± 4.5%, p=0.265 | 45.0 ± 4.3%, p=0.744 | 45.0 ± 4.5%, p=0.764 | 66.7 ± 4.4%, p=0.010 | 52.7 ± 5.1%, p=0.252 |

**Table 4.6.** Within-category individual item encoding accuracies for the low-level perceptual model. Here are reported the accuracies in each ROI of the encoding procedure in each presentation modality (mean±standard error) for the perceptual model based on GIST. For Ang AAL, Ang NS, LOC, HG and Cal, please refer to Figure 4.1.
* Indicates a successful encoding at $p<0.05$, Bonferroni corrected for the eight ROIs from the brain atlas by Craddock et al. (2012).

The two intraparietal ROIs were the only ones that reached significant accuracy across presentation modalities, as the analysis in the other regions of the left parietal cortex, and in the angular gyrus defined both on anatomical or functional constraints, did not reach the significance threshold for any model.

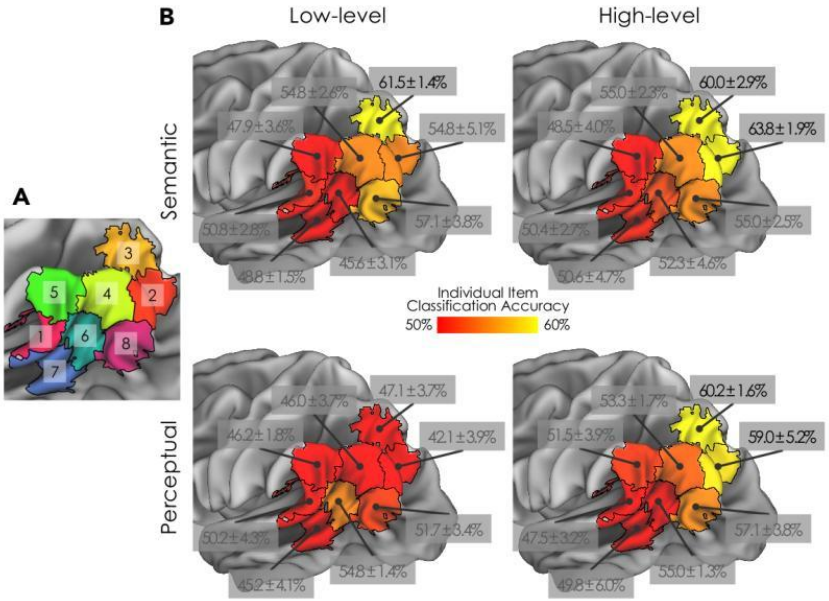


**Figure 4.3.** Encoding results. Figure depicts the mean accuracy across presentation modalities of the representational similarity encoding analysis of the four models in the left lateral parietal cortex. The significant accuracy values ($p<0.05$, Bonferroni corrected) are reported in bold font, the other values were not significant. Detailed results are reported for each ROI in Tables 4.3–4.6.

In addition, the same analysis was performed on two primary sensory control regions, bilateral Heschl gyri (HG) and pericalcarine cortex (Cal) and in the left lateral occipital complex (LOC). Overall, the accuracy across presentation modalities in these ROIs did not reach the threshold for significance ($p>0.05$, uncorrected for multiple comparisons) apart for the high-level shape-based perceptual model, which achieved a significant discrimination in left LOC (56.0±3.7%, uncorrected $p=0.040$).

Here, the similarity encoding procedure aimed at discriminating individual items within each category thus to control for possible biases related to the categorial organization. However, to obtain accuracies comparable to results from previous studies

(Anderson et al., 2016b; Mitchell et al., 2008), we performed the encoding analysis exploring the whole RS (i.e., among-categories procedure), without restricting to the within-category information. Results for the high-level semantic model only were depicted in Figure 4.4B. Briefly, the high-level semantic model yielded an overall increase of the accuracy values in the eight ROIs of the left lateral parietal cortex (i.e., +13.5±3.0% on average), when using models which were affected by categorial organization. Moreover, all the ROIs in the left parietal cortex resulted to be significant using the among-categories procedure ($p<0.05$, Bonferroni corrected).

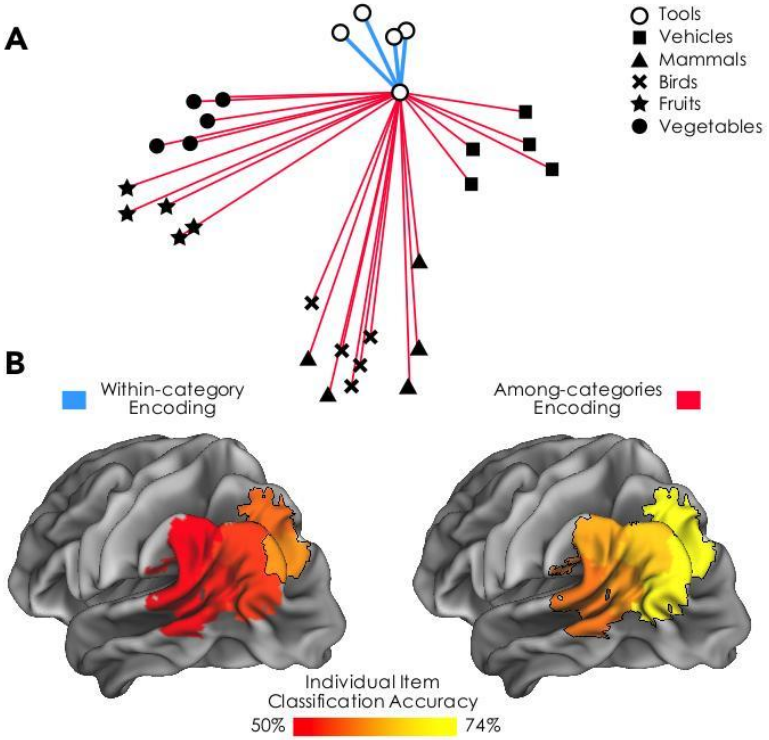


**Figure 4.4.** Comparison between the within-category and among-categories procedures. Panel A: a multidimensional scaling of the high-level semantic RS in sighted subjects. Within- and among- distances for a single item were represented with blue and red lines respectively. Overall, the mean of the within-distances represents about the 55% of the mean of the distances between all the possible pairs of semantic items belonging to different categories in the RS. Panel B left: overall accuracies for the within-category procedure. Panel B right: the overall accuracies for the among-categories procedure in the left lateral parietal cortex. The among-categories procedure yielded an overall increase of the accuracy values of +13.5±3.0% in the left parietal cortex, and all the eight ROIs from the Craddock's atlas resulted to be significant ($p<0.05$, Bonferroni corrected). The borders of the regions that reported an above chance accuracy are marked with a solid line.

## Discussion

To pursue a more comprehensive description of conceptual knowledge organization, this study investigated the specific representation of individual semantic concepts in the angular gyrus and in the neighboring cortical regions within the left lateral parietal cortex, as the extant literature strongly links this area to semantic processing. Patterns of brain activity related to thirty concrete nouns belonging to different categories were analyzed through similarity encoding. Our within-category procedure focused on the differences between items belonging to the same category, representing therefore a reliable description of single-item processing, rather than reflecting the superordinate information. In addition, we used four models – two based on linguistic features extracted by a property generation task, and two based on visual computational models applied to pictorial stimuli – to identify brain regions that encode semantic or perceptual properties of single items and to assess whether these representations were more tied to low-level or high-level features.

*Similarities and differences of the encoding models*. The significant correlation between the high-level semantic models in sighted and congenitally blind individuals, as obtained using the within-category approach, confirms the similarity between their representations. Akin results have been previously obtained from the correlation of the whole semantic RS, without controlling for the role of category membership (Handjaras et al., 2016). Therefore, the current finding suggests that the similar high-level semantic representations between the two groups do not merely originate from a common categorial ground

(Connolly et al., 2007). Conversely, no significant correlation was achieved when comparing all the semantic models (i.e., low- or high-level) with all other perceptual models, suggesting that the language- and sensory-based descriptions adopted in this study covered different features of the thirty concrete nouns. Of note, the low-level semantic model, albeit based on the subsample of features covering specific sensory information (e.g., shape or color) did not correlate significantly with the high-level semantic model, showing that the selection of features yielded an alternative description of the concrete nouns. Similarly, this low-level semantic model did not correlate between sighted and blind individuals, indicating that it retains specific linguistic features shaped by sensory (i.e., visual or non visual) information (Lenci et al., 2013).

*Parietal regions encode perceptual and semantic representations.* When selectively focusing on the left angular gyrus only – either anatomically or functionally defined – neither the high-level, nor the low-level models achieved significant accuracy. On the other hand, in the parcellated map that included also the surrounding parietal areas, the within-category procedure yielded a successful encoding of the thirty concrete nouns in the intraparietal regions for the high-level models, both semantic and shape-based.

The left lateral parietal region is a key part of the frontoparietal network and is typically associated with attentional tasks focusing on specific features of a stimulus, i.e. feature-based attention (Liu et al., 2011; Liu et al., 2003), or on specific objects in complex environments, i.e. object-based attention (Corbetta and Shulman, 2002). However, other studies have reported processing of object features in posterior parietal

regions of the dorsal visual pathway to guide actions or motor behavior, and even suggested a strong similarity of object representation between posterior IPS and LOC (Konen and Kastner, 2008; Mruczec et al., 2013). In our study, we report above-chance accuracy for the shape-based model in ROI 2 and 3, which comprises posterior and middle IPS and extends to superior parietal cortex. Of note, we consider the shape-based model as a high-level perceptual description of the items, since it relies on shock-graphs that are robust to object rotation and scaling (Van Eede et al., 2006). Therefore, our finding is in line with a very recent study showing that posterior IPS is not critical for perceptual judgments on object size or orientation (Chouinard et al., 2017),

The low-level perceptual model did not reach above-chance accuracy thresholds neither in the lateral parietal cortex, nor in the primary sensory (though achieving 59.2±4.4%; $p = 0.089$ in Cal for the pictorial modality in sighted individuals) and lateral occipital areas chosen as control regions. This finding suggests that parietal regions do not encode low-level information and that our GIST-based perceptual model allows to control for low-level visual features. Of note, this is in accordance with a previous fMRI report, which shows that IPS is recruited for object processing irrespective of spatial frequency modulation (Mahon et al., 2013).

When considering semantic representations, we achieved above-chance accuracies in ROI 2 and ROI 3 for the high-level model, while the low-level one was significant in ROI 3 only. Our findings are consistent with the evidence that left posterior parietal areas are usually activated during experimental tasks involving retrieval and combination of concepts (Seghier and Price, 2012), and single-word processing during sentence reading

can even predict response patterns in this area (Anderson et al., 2016a). Hence, both the functional role of the left lateral parietal cortex in semantic processing and autobiographical memory (Seghier and Price, 2012) and its anatomical location and connections (Binder et al., 2009; Friederici, 2009; Price, 2012) strengthen the hypothesis that the angular gyrus and its surrounding regions may represent a key hub to access high-level content of sensory information. This area is also the putative human homologue of the lateral inferior parietal area of the monkey that processes individual items to match them with the superordinate categories they belong to (Freedman and Assad, 2006). Overall, these studies suggest a coding of high-level features in the left intraparietal area, accounting for its role in memory retrieval, combination of concepts and other language-related functions (Price, 2012).

In this study, left LOC showed above-chance accuracy for the high-level perceptual model only. This finding is therefore consistent with the literature suggesting the encoding of object features in this area (Malach et al., 1995; Downing et al., 2007; Konen and Kastner, 2008; Peelen et al, 2014; Papale et al., 2017; Papale et al., 2019). In addition, the below-chance accuracy of the high-level semantic model suggests that the role of this region could be more related to the processing of shape-based information. The results in LOC for the shape-based model are mainly driven by blind individuals and are in line with previous studies that identified LOC ability to process object features across different modalities (Peelen et al, 2014; Handjaras et al., 2016; Amedi et al., 2007).

*Category-related properties strongly impact on single-item semantic encoding.* To account for the impact of the categorial

organization of semantic information on single-item discrimination, the analysis was also performed with an among-categories approach, thus comparing the activity patterns between all the possible pairs of concrete nouns. The results, reported in Supplementary Materials in Handjaras et al. (2017), show an increased accuracy in the Angular Gyrus (defined either anatomically or functionally) and in all the regions of the parcellated map. As consequence, all the ROIs in the left parietal cortex reached the significance threshold using the among-categories procedure.

To further describe the impact of superordinate information within the high-level semantic model, we measured the ratio of the distances between items from the same category and the distances between all the possible pairs of semantic items belonging to different categories, as depicted for illustrative purposes in Figure 4.4A. The resulting value of about 0.55 suggests that superordinate categories play a sizable role: this contribute points out that the individual-item semantic encoding may be driven by the differences among superordinate categories, as the increased accuracy values in all ROIs for the among-categories encoding confirm (Figure 4.4B). This occurrence may arise from broader differences between stimuli, which can be related to the role of superordinate categories per se or by coarse-level distinctions (e.g., living vs. not-living).

The relationship between individual semantic items and brain activity patterns during semantic processing have been recently questioned (Barsalou, 2017). In this account, the development of semantic tiles (i.e., the clusters of voxels homogeneously encoding groups of words, as described by Huth et al., 2016) may be shaped by concurrent coarse-level properties which emerge as principal components of the items and subsequently

guide their clustering (Huth et al., 2016; Barsalou, 2017). In other words, superordinate categories emerge from major differences between stimuli and can be therefore collinear with global properties of the stimuli (e.g., animacy, concreteness, function). Recently, some authors attempted to encode global properties in brain areas associated with semantic processes, reporting above-chance discrimination for biological categories (Connolly et al., 2012) and natural behaviors (Nastase et al., 2016) in wide cortical patches encompassing multiple brain areas. On the contrary, some individual and well-defined properties of objects (i.e., manipulability: Mahon et al., 2013) or animals (i.e., dangerousness: Connolly et al., 2016) were specifically decoded from brain activity in IPS. In light of this, the large extent of parietal cortex achieved in our study by the among-categories encoding of individual items should be interpreted as a lack in specificity, due to the major role played here by superordinate information and its associated global properties. Whether these global properties, widely distributed on the human cortex, retain an essential role in conceptual representations of individual items is still matter of debate (Barsalou et al., 2017). We speculate that areas like the Angular gyrus may process superordinate features only, therefore representing concepts at a higher level of abstraction through a hierarchical conjunctive coding (Barsalou, 2016; Binder, 2016). These results highlight the need to control for category-driven differences – as we did in our within-category individual item encoding – as this represents the best possible way to disentangle the role of coarse and fine differences between concepts in semantic studies.

*The role of the property-generation task. methodological considerations and limitations.* The results from the high- and

low-level models in the IPS suggest that this region is not simply recruited by sensory-specific information in a bottom-up manner (Ibos and Freedman, 2016), but, conversely, encodes higher-level feature-based representations. This is consistent with previous reports (Scolari et al., 2015) and with the overlapping activation of intraparietal cortex during semantic processing, previously observed in sighted and congenitally blind individuals during single word processing (Noppeney et al., 2003). Since results were above chance in both sighted and congenitally blind individuals, we posit that the left IPS encodes representations, independent from sensory modality and not related to visual imagery (Ricciardi et al., 2014a; Ricciardi et al., 2014b; Ricciardi and Pietrini, 2011).

Of note, lateral and posterior parietal areas have been traditionally associated with feature binding tasks, during which object features processed in separate maps are spatially and temporally integrated to produce a unified perceptual and cognitive experience (Robertson, 2003; Scolari et al., 2015; Shafritz et al., 2002; Treisman and Gelade, 1980). Additional evidence of the binding role of parietal areas were provided by neuropsychological studies that showed patients with lesions in posterior parietal regions which fail to conjoin different visual features related to the same object (Friedman-Hill et al., 1995; Robertson et al., 1997; Treisman and Gelade, 1980). Even if we may suppose the binding of perceptual and semantic features to be fundamental for a finer-grained description of individual items, we cannot exclude that the within-category encoding in latero-posterior parietal cortex could be more related to the property generation task, rather than to conceptual processing. Indeed, the property generation task, similar to a feature binding task, relied on the association of properties to concrete nouns.

We assume that the nature of the task, combined with an analysis aimed at evidencing the differences between the representations of single nouns, could account for the recruitment of the intraparietal cortex (Bonnici et al., 2016; Handjaras et al., 2016; Pulvermuller, 2013). The extent of the association between the activity in posterior parietal regions and the task used should be investigated by future studies, in which single-item semantic processing is analyzed through different tasks which do not require an active manipulation of the words.

*Limitations*. Some additional limitations of our study also should be highlighted. First, the analysis was conducted on a single group-level neural RS, obtained from the average of the five individual RSs for each presentation modality. While this can be considered as an estimation of a group-level representation (Carlson et al., 2014; Kriegeskorte et al., 2008), this RS does not consider differences between individual subjects (i.e., each subject's own conceptual representation), that may play a greater role in single-item semantic studies as compared to studies employing category-based models (Charest et al., 2014). Moreover, group-level RSs, although commonly used to increase signal-to-noise ratio of fMRI activity patterns (Carlson et al., 2014; Kriegeskorte et al., 2008) – a mandatory requirement to perform single item encoding – do not take into account the random-effect model. This limitation affects the generalizability of these findings. In addition, the within-category encoding was performed only on a small number of examples, as each category contained only five different items. Further studies may benefit greatly from more accurate models that compare a greater number of concrete nouns while controlling for their category membership. Finally, the analyses were performed on a single

parcellation of the left parietal cortex, chosen a priori on the basis of an atlas based on resting-state functional activity (Craddock et al., 2012). For this reason, we cannot exclude that different parcellation criteria (e.g., the choice of a different atlas or a different number of ROIs) can yield different results in the encoding analysis, mainly due to the dependence of the accuracy on the size and signal-to-noise ratio of the chosen ROIs.

In addition, the sample size for each experimental group (n=5) might represent a criticism. While this number may appear relatively small for an univariate fMRI study, this is not the case for studies employing a RS pipeline, as the current one (Kriegeskorte et al. 2008; Kriegeskorte et al. 2013; Ejaz et al., 2015). Notably, the first paper using this technique (Kriegeskorte et al. 2008), compared RSs obtained from two monkeys and four human subjects. In RS analysis, rather than the number of subjects, the total number of acquired trials represents the key factor to obtain stable RS. In addition, in a previous study (Handjaras et al., 2016), we tested the effect size stability using this experimental setup. We acquired data from a larger sample of subjects (n=10) employing the pictorial presentation modality. Subsequently, we measured the encoding accuracy when including in the analysis 1 to 10 subjects (Handjaras et al., 2016; Supplementary Figure 12). Results demonstrated that the encoding accuracy remained stable (mean accuracy in 5 subs 77.3±6.4%; mean accuracy in the larger sample of 10 subs: 77.2±5.2%, p=n.s.), supporting the robustness of the RS methodological approach.

Another potential limitation regards the choice of averaging the encoding performances across different groups. Our previous study using the same data has reported that the semantic information in the left lateral parietal cortex is

consistent across all presentation modalities (Handjaras et al., 2016). In addition, a recent study has reported highly similar activity patterns for pictorial and word-based representation of natural scenes in posterior IPS, showing that brain patterns elicited by pictures can be decoded by a classifier trained on words, and vice-versa (Kumar et al., 2017). This confirms that the presentation modality does not play an important role in driving semantic processing in this region.

In conclusion, this study shows that the processing of high-level features – both semantic and perceptual (i.e., shapes) engages to different degrees individual sub-regions of the left lateral parietal cortex, showing higher accuracy in the intraparietal sulcus, whose activity was predicted using a high-level models that accounted for the differences between individual concepts. Conversely, high accuracy in a large extent of parietal cortex comprising the angular gyrus and its neighboring regions can be achieved only when the information regarding superordinate categories is retained. Overall, these results indicate the need to control for the coarse-level categorial organization when performing studies on higher-level processes related to the retrieval of semantic information, such as language and autobiographical memory.

# 5. Single subject decoding of autobiographical events


## Abstract

"Autobiographical memory" (AM) refers to remote memories from one's own life. Previous neuroimaging studies have highlighted that voluntary retrieval processes from AM involve different forms of memory and cognitive functions. Thus, a complex and widespread brain functional network has been found to support AM. The present functional magnetic resonance imaging (fMRI) study used a multivariate approach to determine whether neural activity within the AM circuit would recognize memories of real autobiographical events, and to evaluate individual differences in the recruitment of this network. Fourteen right-handed females took part in the study. During scanning, subjects were presented with sentences representing a detail of a highly emotional real event (positive or negative) and were asked to indicate whether the sentence described something that had or had not really happened to them. Group analysis showed a set of cortical areas able to discriminate the truthfulness of the recalled events: medial prefrontal cortex, posterior cingulate/retrosplenial cortex, precuneus, bilateral angular, superior frontal gyri, and early visual cortical areas. Single-subject results showed that the decoding occurred at different time points. No differences were found between recalling a positive or a negative event. Our results show that the entire AM network is engaged in monitoring the veracity of AMs. This process is not affected by the emotional valence of the experience but rather by individual differences in cognitive strategies used to retrieve AMs.

## Introduction

The expression Autobiographical memory (AM) refers to remote memories from one's own life which are characterized by a sense of subjective time, autonoetic awareness (Tulving, 2002), and feelings of emotional re-experience (Tulving, 1983; Tulving and Markowitsch, 1998). AM is part of episodic memory (i.e., the conscious recollection of experienced events), as opposed to semantic memory-i.e., the conscious recollection of factual information and general knowledge about the world (Tulving, 2002). Neuropsychological and neuroimaging data support this notion of multiple systems of memory, each specialized in processing distinct types of information (Vargha-Khadem et al., 1997; Cipolotti and Maguire, 2003) and subserved by distinct, functionally independent neural networks (Gabrieli, 1998; Cabeza and Nyberg, 2000; Tulving, 2002).

As a matter of fact, neuropsychological studies support the functional dissociation between these memories: patients with medial temporal lobe lesions are defective in AM recall, but not in semantic memory tasks (Vargha-Khadem et al., 1997; Tulving and Markowitsch, 1998; Gadian et al., 2000). Conversely, patients with semantic dementia, who show damage in fronto-temporal regions, are impaired in semantic memory tasks (Neary et al., 1999), whereas their AM is relatively spared (Snowden et al., 1994; McKinnon et al., 2006).

More recently, neuroimaging studies have disentangled the functional characteristics of the neural networks mediating specific memory systems. The left inferior prefrontal cortex and left posterior temporal areas are in general recruited during semantic retrieval (Vandenberghe et al., 1996; Wiggs et al., 1999; Graham et al., 2003), whereas right dorsolateral prefrontal areas

subserve episodic retrieval (Cabeza et al., 2004; Düzel et al., 2004; Gilboa, 2004). With respect to AM, functional neuroimaging studies focused on voluntary retrieval processes that involve different forms of memory and cognitive functions. In particular, recovering an autobiographical event requires a prolonged and effortful memory search about one's own life, combined with the retrieval of specific episodic knowledge about its contextual information. The retrieved memory content typically includes emotions and visual images, and is mediated by inferential and monitoring cognitive processes (Cabeza and St Jacques, 2007).

A meta-analysis paper showed that, because of the multi-modal nature of AM retrieval and of the heterogeneity of the tasks used in literature, different regions emerge during recollection (Svoboda et al., 2006). However, a core neural network for AMs comprises the left lateral prefrontal cortex (l-PFC) for search and controlled processes; the medial prefrontal cortex (m-PFC) for self-referential processes; the hippocampus and the retrosplenial cortex for recollection; the amygdala for emotional processing; the occipital and cuneus/precuneus regions for visual imagery, and the ventromedial PFC (vm-PFC) regions for feeling-of-rightness and monitoring (Cabeza and St Jacques, 2007).

Two additional issues are relevant for AM. First, AMs often exhibit a richer emotional content as compared to episodic and semantic memories. In particular, emotional life events are recalled better than non-emotional events (Holland and Kensinger, 2010). Second, several neuroimaging studies demonstrated a significant individual variability in AMs performance (Rypma et al., 2002; Schaefer et al., 2006; Miller and Van Horn, 2007). Typically, most of these studies evaluated the modulation of brain areas commonly activated across subjects,

and only a few studies considered the individual variability across the whole brain (McGonigle et al., 2000; Feredoes and Postle, 2007; Seghier et al., 2008).

In spite of the importance of the mechanisms underlying the successful recollection from AM, only a few studies previously investigated this issue (Gilboa et al., 2004; Greenberg et al., 2005; Cabeza and St Jacques, 2007; Chen et al., 2017). Rather, many authors questioned whether brain functional patterns could differentiate between true memory, false memory (a common type of memory distortion in which individuals incorrectly believe they have already encountered a novel object or event), and deception. Regions within the prefrontal cortex have been related to these memory monitoring activities (Cabeza and St Jacques, 2007). Nonetheless, to the best of our knowledge, only one study evaluated recognition from AM (Harris et al., 2008). However, the authors used a wide range of stimuli (autobiographical, mathematical, geographical, religious, ethical, semantic, and factual) and results were presented irrespectively of the kind of memory involved.

The present single-event fMRI study was designed to determine whether neural activity within the AM network, as identified by previous neuropsychological and neuroimaging studies, would recognize memories of real autobiographical events. Moreover, we examined whether retrieval of positive and negative emotional events from AM would exert distinctive effects on brain response. Specifically, we asked subjects to recall a highly emotional personal event (either her wedding or the funeral of a close relative) in a pre-scan semi-structured interview. During scanning, subjects were presented with sentences referring to a detail of the event recalled and were asked to indicate whether the detail actually belonged (true) or

not (false) to their AMs. Using a multivariate technique (Mitchell et al., 2008), we aimed at evaluating the neural network in each individual subject independently, so that we could identify both the time points at which the successful recollection occurred and the network involved in the process. Then, results from each subject were combined to identify the brain regions involved in the common cognitive mechanism underlying AM, thus accounting for individual differences in the recollection processes.

## Materials and Methods

*Subjects.* Inclusion criteria were: right-handed healthy females with no history of neurological or psychiatric diseases; no subject took any psychiatric medication at the time of the study; age 30–45 years; having experienced either a highly positive (own wedding, being still married at the time of the experiment) or a highly negative (funeral of a loved one, who died suddenly) event in the recent past (range: 2–8 years). Consequently, 14 subjects (mean age 37 ± 7 years; mean school-age 17 ± 2) were enrolled. This final group included: personnel from the University of Modena and Reggio Emilia staff, acquaintances and relatives of the authors. Only female volunteers participated to the study, as data in the literature indicate that gender influences memory, and particularly the emotional modulatory mechanism on memory storage (Cahill, 2010). All participants gave their written informed consent after the study procedures and potential risks had been explained. The study was conducted under protocols approved by the Local Modena Ethical Committee, in accordance with the ethical standards of the 2013 Declaration of Helsinki.

*Pre-scan interview session.* From 2 to 8 days before fMRI scanning, a detailed description of highly emotional events was collected using a custom-made semi-structured interview. Indeed, the "pre-scan interview method" could be particularly useful to evaluate the common and individual neural network for retrieving AMs in neuroimaging studies. Eight participants were asked to describe a positive event (i.e., their wedding), whereas six participants to recall a negative event (i.e., the funeral of a loved one). The interview about the wedding day

consisted of 54 questions, organized in 4 different categories concerning: 1. the ceremony; 2. the wedding dress; 3. the wedding party; 4. the honeymoon. Four categories were also included in the funeral day's interview (32 questions): 1. the deceased's physical description at the time of his/her death; 2. the announcement of the death; 3. the last meeting; 4. the funeral. The answers were used to compose a true story. A second false story was written, modifying some details of the true story (e.g.: "We got married in April": true; "We got married in September": false). The true stories consisted of information stored in the autobiographical memory (AM) of the participants, whereas the details of the false stories did not belong to their AM.

*Image acquisition and experimental setup of the fMRI session.* Brain activity was measured using fMRI with a three-run event-related design (gradient echo echoplanar images, Philips Achieva 3T, TR 2.0 s, FA: 80°, TE 35 ms, 30 axial slices, 80 × 80 acquisition matrix, 3 × 3 × 4 mm voxel). High-resolution T1-weighted spoiled gradient recall (TR = 9.9 ms, TE = 4.6 ms, 170 sagittal slices, 1 mm isovoxel) images were obtained for each participant to provide detailed brain anatomy.

Behavioral responses were collected during the scanning sessions by means of a custom-made software developed in Visual Basic. The same software was used to present stimuli via IFIS-SA System (MRI Device Corporation, WI, USA) remote display.

During the scanning session, prior to the fMRI acquisition, subjects were asked to read both stories (i.e., the true and the false one) twice, in order to avoid the novelty effect of the incorrect information (Schomaker and Meeter, 2015). The order of presentation of the stories was counterbalanced between

subjects. The experimental stimuli were sentences representing a true or a false detail of the event described in the stories. The false and true item referring to the same AM detail differed only in one feature (i.e., He died in May vs. He died in April; My wedding dress was white vs. My wedding dress was ivory). During scanning, after a warning cue lasting 0.5 s, subjects were presented with a sentence (5.5 s). After a 12 s interval, subjects were asked to indicate whether the sentence belonged (true, T) or not (false, F) to their autobiographical memory by pressing one of two buttons on the keypad (2 s, Figure 5.1), followed by 10 s of inter-trial interval. Response times and accuracies were recorded. A total of 48 sentences (24 T and 24 F) were randomly presented to each subject in three runs. At the beginning and at the end of each run, a fixation cross was presented for 30 s to obtain a baseline measure of brain activity. Overall, each run lasted about 9 min. The true-false responses given during scanning were subsequently used for the behavioral and functional analyses.

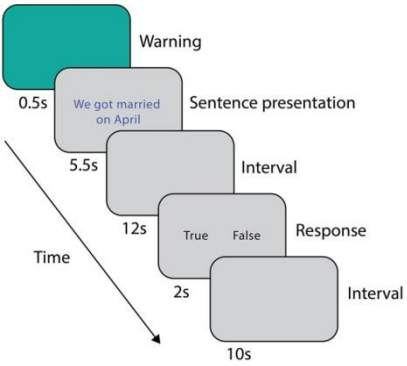


**Figure 5.1.** Experimental protocol for the fMRI scan session.

*Behavioral analysis.* A two-way ANOVA was performed on the response times with the following factors: group (two levels, wedding and funeral) and response (two levels, true and false).

Significance threshold was set at $p < 0.05$. Analyses were performed using SPSS 18 (SPSS Inc.).

*fMRI data preprocessing*. The AFNI software package was used to analyze functional imaging data (Cox, 1996). All volumes from the different runs were processed to remove spikes (3dDespike), temporally aligned (3dTshift), corrected for head movements (3dvolreg), spatially smoothed (3dmerge, Gaussian kernel 5 mm, FWHM) and scaled to voxel mean. Motion spikes were estimated through the evaluation of Framewise Displacement (FD) implemented in FSL (Jenkinson et al., 2012), with a cutoff of 0.6 mm (Power et al., 2012). Subsequently, a generalized least squares regression was performed (3dREMLfit) to model the motion spikes, movement parameters, signal trends and the temporal correlation structure with an ARMA(1,1) model, thus to remove nuisance signals from the data. Then, the residual signal for each voxel was normalized by subtracting the mean and dividing the result by its standard deviation. Afterwards, for each trial, the signal time points from the onset of the sentence to the motor response, were extracted and included in the multivariate analysis. A central moving average was computed ("temporal smoothing") (Friston et al., 1995; Strappini et al., 2017) by averaging the value of each point in time ("reference point") and the value of the two points on either side of the reference point. By this procedure, we generated seven overlapping windows, from 2 to 14 s after sentence onset. The duration of the explored window was decided following previous studies which showed that the retrieval of detailed autobiographical memories can spread over a long time (e.g., up to 20 s) (Svoboda et al., 2006), but also in order to avoid any overlap with the motor response.

Subsequently, single subject time series data were registered to the MNI152 standard space using the nonlinear registration implemented in AFNI (3dQWarp), and the acquisition matrix was resampled to a 3 mm iso-voxel. Finally, to reduce computational effort in the subsequent steps, a spatial mask was applied to select gray matter voxels only.

*Single-subject decoding analysis.* Since we were interested in selecting the subset of voxels with the highest discrimination ability in distinguishing between "true" and "false" responses, we used a modified version of the procedure originally adopted by Mitchell et al. (2008) and already validated on different datasets (Handjaras et al., 2016; Leo et al., 2016). Briefly, a machine-learning algorithm predicted the fMRI activation in the brain as a weighted sum of images, each one generated from a behavioral matrix (here, a binary vector which defined the "true" and "false" responses). In detail, a regression analysis, performed within a leave-two-stimuli-out cross-validation procedure, produced a learned scalar parameter that specifies the degree to which the dimension related to the truthfulness of the memories modulates the voxels activity. Hence, for each iteration of the cross-validation procedure, the model was first trained with 46 out of 48 stimuli (i.e., 23 "true" and 23 "false"), then only the 2,000 voxels that showed the highest coefficient of determination $R^2$ and with a cluster size larger than 20 voxels (to remove small isolated clusters) were considered. Once trained, the resulting algorithm was used to predict the fMRI activation within the selected 2,000 voxels of the two left-out stimuli (one related to a "true," one to a "false" response). Afterward, prediction accuracy was evaluated with a simple match between the predicted and the real fMRI activations of the two left-out

stimuli using cosine similarity. This leave-two-out procedure was iterated 576 times, training and testing all possible stimulus pairs between the true and false items. A bootstrapping procedure was used to measure the standard error of the accuracy (1,000 iterations) (Efron and Tibshirani, 1993). The algorithm for the single-subject decoding analysis was applied for each subject and time point (i.e., from 2 to 14 s after sentence onset), thus generating an accuracy value and a decoding map with the subset of brain voxels used during the procedure.

The single-subject accuracy was tested for significance against the null distribution of accuracies generated with a permutation test based on the same procedure defined above (Schreiber and Krekelberg, 2013; Handjaras et al., 2015). As the processing of false sentences does require the retrieval of information related to the true event counterpart, we adopted permutation tests: these are the most robust methods to assess statistical significance in conditions, such as our experiment, where the chance level is not necessarily centered on 50% and where the degrees of freedom are unknown, ranging between the number of the stimuli (i.e., 48) and the total number of comparisons (i.e., 576) (Schreiber and Krekelberg, 2013; Berry et al., 2019). Moreover, as the null distribution was always created upon individual brain activity in each subject, the significance threshold reflected any possible bias in the data. Briefly, in each subject and time point, a null distribution of accuracies was built by shuffling the behavioral matrix during the training phase. The procedure was repeated 100 times (Winkler et al., 2016) for each time point, leading to a null distribution of 700 accuracy values across the whole time window. Each single-subject accuracy was therefore tested against the null distribution of accuracy values

to identify a common significance threshold across the time window (one-sided rank test, $p < 0.05$; Table 5.1 and Figure 5.2).

*Group level map.* Subsequently, to measure the spatial consistency of the regions involved in autobiographical memory processing, a posterior probability map was built across the time windows by combining the single subject decoding maps at the time point with the highest accuracy value. This procedure therefore merged the most informative voxels involved in the "true" and "false" responses irrespectively of the time at which the voxels were maximally engaged. We arbitrarily selected a threshold ($p > 0.33$, minimum cluster size of 20 voxels) that represented the probability of a voxel to be informative in at least 5 subjects out of 14 (Figure 5.3; Leo et al., 2016).

*Assessing the reliability of the group level map.* This group level map was the result of the aggregation of the single subject most discriminative voxels at different time points, in order to account for the possibility that individual subjects processed autobiographical memory content with different retrieval times. Therefore, we further tested the sparseness of the map obtained from this procedure, as we reasoned that the cognitive mechanisms underlying the discrimination of "true" and "false" responses would engage the same brain regions across subjects. Theoretically (e.g., assuming no variability across subjects), the ideal group map should include the same 2,000 voxels of the decoding procedure across all subjects and probability thresholds, albeit at different time points (Figure 5.4). On these assumptions, a permutation test was built by randomly combining the decoding maps at different time points across subjects and subsequently measuring the total number of voxels

at each probability threshold (10,000 iterations, $p < 0.05$) (Figure 5.4). We hypothesized that our group map should have the lower number of voxels, as compared to the null distribution, thus indicating that brain regions involved in the process remained significantly stable across subjects (i.e., no sparseness). In addition, to assess the spatial overlap of the decoding maps considering the same retrieval time for all the subjects, we included in the aforementioned test the seven group maps obtained by aggregating the decoding maps at a fixed time point (e.g., group map at the 2 s time point).

*Assessing the differences between negative and positive memories.* The group probability map was obtained by combining the subjects from the two groups, considering the discrimination between "true" and "false" responses irrespectively of the positive or negative emotional valence associated to the retrieved memory. Here we tested whether the different valence of the memories could affect when (i.e., the time point with the highest accuracy) or where (i.e., the brain regions involved in the process) the retrieval occurred. First, we compared the time points with the highest accuracy between the two groups (Mann-Whitney U test, two-tailed, $p < 0.05$). Second, we measured the spatial overlap within the two groups. To this aim, we first evaluated the spatial overlap of the decoding maps between the 14 subjects using the Sørensen-Dice (SD) coefficient (Dice, 1945; Kolasinski et al., 2016). Subsequently, the Ratio (R) between the averaged SD values within- and the averaged SD values between-groups was computed. R represents whether each group shows a higher within-group similarity ($R > 1$), a higher between-group similarity ($R < 1$), or a spatial overlap

between groups (R≈1). Confidence intervals of R were obtained through a permutation test (10,000 iterations, $p < 0.05$).
The multivariate pattern analyses were carried out using Matlab (Matworks Inc., Natick, MA, United States), while Connectome Workbench (Marcus et al., 2011) was used to render the brain meshes in Figure 5.3.

## Results

*Behavioral results.* Response times showed no significant effect for response [mean in s ± standard deviation; "True" trials: 1.15 ± 0.22; "False" trials: 1.19 ± 0.20; $F_{(1, 11)} = 0.12$, $p = 0.733$] or group [weddings: 1.21 ± 0.22; funerals: 1.09 ± 0.17; $F_{(1, 11)} = 1.06$, $p = 0.325$], nor for their interaction [$F_{(1, 11)} = 0.57$, $p = 0.466$]. Overall, this evidence indicated that at the button press (i.e., 17.5 s after sentence onset), the retrieval of the autobiographical information was already concluded regardless of the item truthfulness or valence. Response accuracy was at ceiling level (overall accuracy value across conditions: 98%).

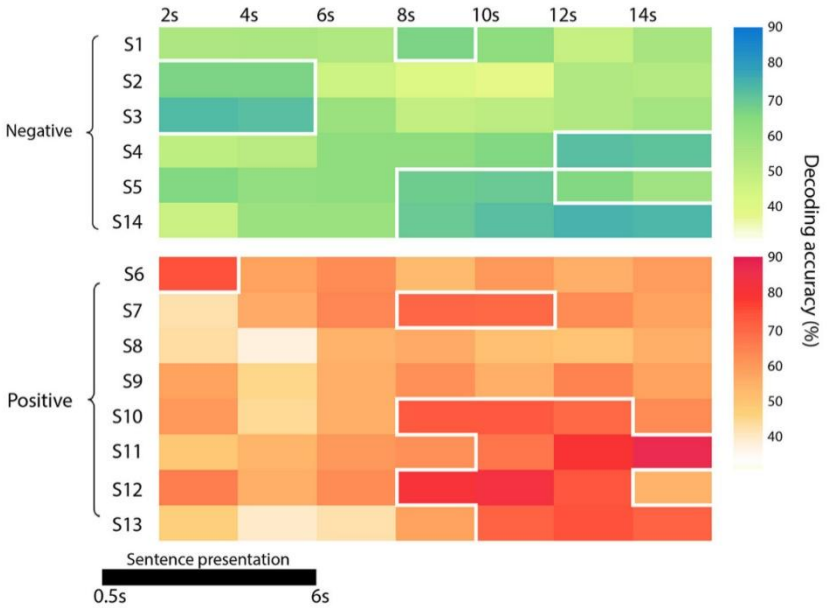


**Figure 5.2.** Diagram representing the accuracy of each subject and group (in green the negative one -the funeral of a loved one- and in red the positive event -wedding), at each time point. Significant time points ($p < 0.05$) are marked with a white border.

*Single-subject decoding results.* Since the time required for the retrieval of autobiographical memory may vary among subjects (Svoboda et al., 2006), we avoided a standard group level analysis, focusing only on the single subject decoding of "true" and "false" responses within a relative large time window, from 2 s after trial onset up to 14 s. As reported in detail in Table 5.1 and Figure 5.2, the decoding was successful in 12 out of 14 subjects ($p < 0.05$), ranging from 65.7% to 86.8%, although it occurred at different time points (mean ± standard deviation: 8 ± 4 s). Averaging the highest accuracies across time points and across all 14 subjects led to an overall mean accuracy of 71.4% with a standard error of 2.0%.

| Group | Subject | Time window | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 2 s | 4 s | 6 s | 8 s | 10 s | 12 s | 14 s |
| Funerals | *sub1* | 54.3%, *p* = n.s. | 54.7%, *p* = n.s. | 53.3%, *p* = n.s. | **66.2 ± 1.7%, *p* = 0.041** | 62.2%, *p* = n.s. | 47.0%, *p* = n.s. | 55.7%, *p* = n.s. |
| | *sub2* | **65.7 ± 1.8%, *p* = 0.044** | **65.7 ± 1.8%, *p* = 0.044** | 45.4%, *p* = n.s. | 39.7%, *p* = n.s. | 36.8%, *p* = n.s. | 53.0%, *p* = n.s. | 52.2%, *p* = n.s. |
| | *sub3* | **73.0 ± 1.7%, *p* = 0.009** | **72.0 ± 1.8%, *p* = 0.011** | 60.0%, *p* = n.s. | 48.2%, *p* = n.s. | 50.3%, *p* = n.s. | 53.0%, *p* = n.s. | 56.4%, *p* = n.s. |
| | *sub4* | 49.1%, *p* = n.s. | 51.0%, *p* = n.s. | 62.6%, *p* = n.s. | 62.2%, *p* = n.s. | 65.4%, *p* = n.s. | **72.0 ± 1.7%, *p* = 0.013** | **71.1 ± 1.7%, *p* = 0.019** |
| | *sub5* | 64.9%, *p* = n.s. | 61.7%, *p* = n.s. | 62.7%, *p* = n.s. | **68.3 ± 1.9%, *p* = 0.040** | **68.9 ± 1.7%, *p* = 0.030** | 65.3%, *p* = n.s. | 57.6%, *p* = n.s. |
| | *sub14* | 46.8%, *p* = n.s. | 59.9%, *p* = n.s. | 59.8%, *p* = n.s. | **69.0 ± 1.8%, *p* = 0.021** | **71.9 ± 1.7%, *p* = 0.009** | **74.1 ± 1.8%, *p* = 0.006** | **73.2 ± 1.8%, *p* = 0.007** |
| Weddings | *sub6* | **74.6 ± 1.5%, *p* = 0.017** | 57.4%, *p* = n.s. | 62.2%, *p* = n.s. | 52.0%, *p* = n.s. | 59.3%, *p* = n.s. | 55.3%, *p* = n.s. | 58.8%, *p* = n.s. |
| | *sub7* | 42.0%, *p* = n.s. | 56.0%, *p* = n.s. | 63.0%, *p* = n.s. | 69.7 ± 1.4%, *p* = 0.014 | 68.9 ± 1.6%, *p* = 0.021 | 62.3%, *p* = n.s. | 57.3%, *p* = n.s. |
| | *sub8* | 43.0%, *p* = n.s. | 37.0%, *p* = n.s. | 53.8%, *p* = n.s. | 56.2%, *p* = n.s. | 50.3%, *p* = n.s. | 49.4%, *p* = n.s. | 54.5%, *p* = n.s. |
| | *sub9* | 57.2%, *p* = n.s. | 44.5%, *p* = n.s. | 55.0%, *p* = n.s. | 61.6%, *p* = n.s. | 55.2%, *p* = n.s. | 64.3%, *p* = n.s. | 57.3%, *p* = n.s. |
| | *sub10* | 59.2%, *p* = n.s. | 43.2%, *p* = n.s. | 54.5%, *p* = n.s. | **72.6 ± 1.5%, *p* = 0.016** | **72.6 ± 1.7%, *p* = 0.016** | **68.8 ± 1.6%, *p* = 0.033** | 62.4%, *p* = n.s. |
| | *sub11* | 48.4%, *p* = n.s. | 53.2%, *p* = n.s. | 59.9%, *p* = n.s. | 61.0%, *p* = n.s. | **67.4 ± 1.9%, *p* = 0.049** | **78.9 ± 1.6%, *p* = 0.002** | **86.8 ± 1.4%, *p*<0.001** |
| | *sub12* | 64.8%, *p* = n.s. | 55.3%, *p* = n.s. | 61.9%, *p* = n.s. | **79.7 ± 2.3%, *p* = 0.004** | **81.8 ± 2.2%, *p* = 0.004** | **73.8 ± 2.5%, *p* = 0.020** | 53.5%, *p* = n.s. |
| | *sub13* | 45.9%, *p* = n.s. | 39.3%, *p* = n.s. | 41.9%, *p* = n.s. | 57.2%, *p* = n.s. | **71.0 ± 1.8%, *p* = 0.034** | **74.4 ± 1.8%, *p* = 0.011** | **70.9 ± 1.8%, *p* = 0.034** |

**Table 5.1.** Table representing the raw accuracy value, its standard error and p-value of each subject and group at each time point. Significant time points ($p < 0.05$) are marked in bold.

*Group level map.* To highlight brain regions involved in the discrimination of "true" and "false" responses, a posterior probability map was built across the whole time window, by combining the single subject decoding maps at the time point

with the highest accuracy. The regions involved in the process are depicted in Figure 5.3.

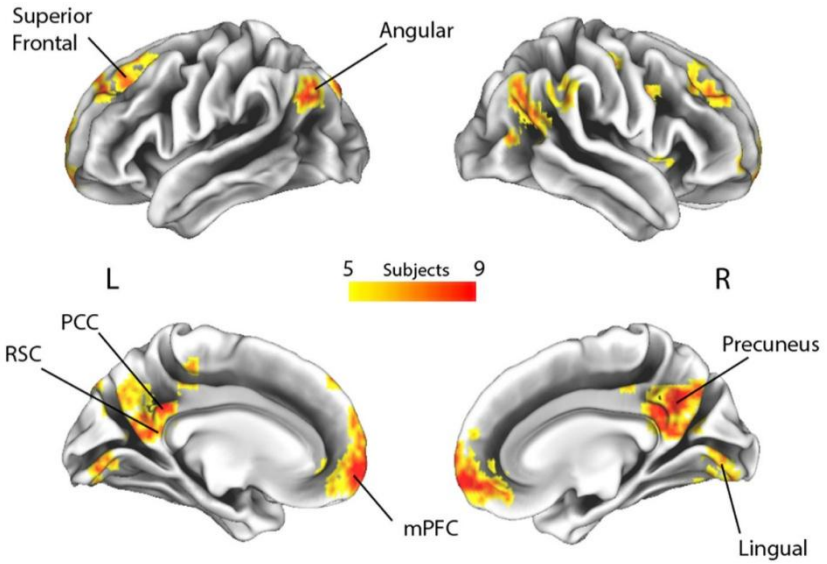


**Figure 5.3.** Spatial overlap of the decoding maps of all subjects across all time points ($p > 0.33$, which represents the probability of a voxel to be informative in at least 5 out of 14 subjects, irrespective of timing). L, Left; R, Right; RSC, retrosplenial cortex; PCC, posterior cingulate cortex; mPFC, medial prefrontal cortex.

By applying a probabilistic threshold of $p > 0.33$ (i.e., the probability of a voxel to be informative in at least 5 out of 14 subjects), irrespectively of timing, a broad set of cortical areas was identified, which comprised several bilateral nodes of the Default Mode Network (DMN), including medial prefrontal, superior frontal and angular regions, retrosplenial cortex, posterior cingulate and precuneus. Precuneus showed the highest overlap among subjects (i.e., nine). In addition, a large cluster was identified bilaterally in early visual cortical areas. Interestingly, in our experiment, other medial temporal lobe key

regions, such as the hippocampal and parahippocampal cortex and the amygdala, did not reveal enough discrimination capacity to detect true from false items.

*Reliability of the group level map.* Individuals processed the autobiographical memory content with different retrieval times (Svoboda et al., 2006). Therefore, to test whether the cognitive mechanism underlying the discrimination of true and false contents is based on the engagement of the same brain regions across our subjects, we combined single subject decoding maps at different time points showing the lowest sparseness (i.e., highest spatial overlap), to built the best group probability map across subjects. The results, represented in Figure 5.4, suggest that the best map includes the lowest number of voxels, irrespective of the chosen probability threshold, as compared to a null distribution built by combining different single subject decoding maps at random time points ($p < 0.05$). Moreover, the seven group maps obtained by aggregating the single subjects decoding maps at each time point fell within the confidence intervals of the null distribution, thus indicating that a standard group level analysis would have led to a non-optimal result.

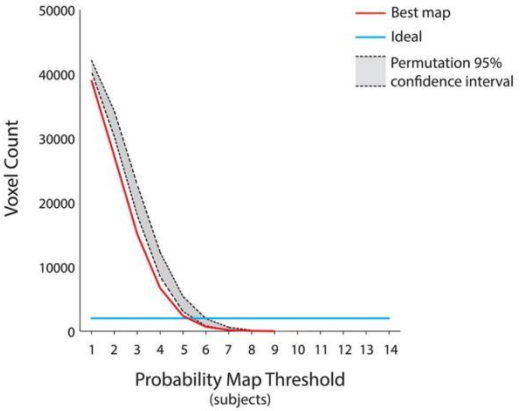


**Figure 5.4.** Assessment for the group level map. Since the group level map of Figure 5.3 was the result of the aggregation of the individual subject decoding maps at different time points, we further tested its sparseness using a permutation test by randomly combining the decoding maps at different time points across subjects and subsequently measuring the total number of voxels at each probability threshold ($p < 0.05$). The ideal group map (e.g., no variability across subjects) is represented by the light blue line, the group level map is represented by the red curve, whereas the 95% confidence interval of the null distribution is outlined in gray. The group level map has a number of voxels lower than the null distribution, irrespective of the chosen probability threshold. Moreover, all the group maps obtained by aggregating the subjects' decoding maps at each of the seven fixed time points fell within the null distribution area ($p < 0.05$).

*Differences between negative and positive memories.* First, we examined whether the discrimination between true and false events occurred using brain activity extracted at different time points in the two groups. No temporal differences were found between subjects who retrieved memories from their wedding and subjects who recalled events from the funeral of a loved person. Moreover, we tested whether there was a significant spatial overlap of the decoding maps between the two groups. To this aim, we developed an ad hoc measure R, based on the SD coefficient (Dice, 1945; Kolasinski et al., 2016), as detailed in the Methods section (see above). We were not able to demonstrate that the two groups had a specific decoding map, since the R

index fell within the confidence interval (R = 1.01, 95% confidence intervals: 0.91–1.16).

## Discussion

The present fMRI study was designed to determine whether neural activity can discriminate true from false memories of real autobiographical events, to investigate individual differences in AM processing, and to isolate specific effects of the emotional valence (i.e., positive or negative) on AMs. Given the subjective nature of autobiographical memories, a multivariate technique (Mitchell et al., 2008) was used to evaluate the retrieval process in each subject independently. Results showed that neural activity discriminated AMs in 12 out of 14 participants (mean accuracy ~71%) across a retrieval time of up to 14 s, although discrimination occurred at different time points across subjects. In addition, to overcome single subject differences, we examined the recognition of real AMs also at a group level by combining the individual decoding maps, and highlighted a set of brain regions which mainly overlaps with the AM core network (i.e., medial prefrontal, superior frontal and angular regions, retrosplenial cortex, posterior cingulate, precuneus and early visual areas) described by Cabeza and colleagues (Cabeza and St Jacques, 2007). Finally, we found no specific effects of either positive or negative emotional valence on AMs.

Our experimental approach attempted to investigate individual differences in AM processing using a functional task. Indeed, neuroimaging studies have focused on behavioral scores or trait measures that can account for modulation effects in commonly activated brain areas (Miller and Van Horn, 2007). Usually, these studies included intra-scanner behavioral performance measures, such as accuracy (Callicott et al., 1999; Gray et al., 2003) or reaction time (Rypma et al., 2002; Wager et al., 2005; Schaefer et al., 2006). A small number of studies related

brain activation to tasks or measures administered outside of the scanner, including measures of working memory span or fluid intelligence (Gray et al., 2003; Geake and Hansen, 2005; Lee et al., 2006) and measures of personality traits (Gray and Braver, 2002; Kumari et al., 2004). In particular, authors correlated the successful retrieval from episodic (Horn and Miller, 2008; King et al., 2015) or working memory (Rypma and D'Esposito, 2000) with neural activity in specific brain regions. However, only a few studies considered individual variability across the whole brain (McGonigle et al., 2000; Feredoes and Postle, 2007; Seghier et al., 2008).

Several studies showed individual variability in performance and neural activity depending on age (Maillet and Rajah, 2014) and gender (Hill et al., 2014). With respect to AM studies, Piefke and Fink concluded that both factors influence the performance in AM tasks and its underlying neural mechanisms. In particular, aging and gender appear to affect the functional hemispheric lateralization of AM recollection and the degree of involvement of prefrontal, hippocampal, and parahippocampal brain areas (Piefke and Fink, 2005).

As recently demonstrated, individual variability in cognitive strategies during AM retrieval, and particularly the tendency to recollect autobiographical memories from an egocentric perspective, exerted a significant effect on a pivotal region within the AM network, the precuneus, in line with the established role for this region in self-centered representations (Hebscher et al., 2018). Indeed, this recent voxel-based morphometry study showed that larger precuneus volumes were associated with the tendency to recollect autobiographical memories from an egocentric perspective. In addition, Sheldon and colleagues evaluated the impact of individual differences during

autobiographical retrieval. Their results showed that self-reported individual differences related to how the subject recalls past events were associated to the intrinsic connectivity between the medial temporal lobe structures and the other nodes of the AM network (Sheldon et al., 2016).

The role of commonalities and differences between subjects, particularly in the time point at which recollection of AMs occurs, needs to be further investigated in order to uncover the association between brain activity and cognitive strategies used to retrieve AMs, as well as with personality traits. Our data showed that the retrieval of AMs relies on the same neural network across subjects, although with individual differences in the time course.

At group level, we evaluated whether neural activity can discriminate true from false autobiographical events, finding a widespread set of brain regions which mainly overlaps with the previously identified AM network (Cabeza and St Jacques, 2007).

The successful recollection from AM is still not fully understood. Rather, several studies investigated the issue of the "feeling of rightness" phenomenon and suggested that the ventromedial PFC could be crucial. Indeed, the activation of this area is commonly observed in tasks requiring self-referential processing (Craik et al., 1999; Gusnard et al., 2001; Kelley et al., 2002) and in decision making tasks under uncertainty, in control processes providing a "feeling of rightness" and in the processing of self-referential information that monitor the veracity of autobiographical memories (Gilboa, 2004).

Other studies have examined the functional networks that subserve the subjective perceptions of familiarity and unfamiliarity in autobiographical recollection. A complex of fronto-parietal regions (lateral PFC and PPC) is involved in

cognitive and attentional control processes that guide the recovery of information from memory, as well as in the evaluative processes that monitor retrieval outcomes and guide mnemonic decisions (Tailby et al., 2017).

Interestingly, key medial temporal regions, such as the hippocampal and parahippocampal cortical areas, did not retain enough ability to discriminate between true and false sentences in our experiment. This presumably depends on the adopted task: subjects were presented with sentences that could belong, or not, to their AM, but differed in one detail only. We speculate that, to monitor the veracity of autobiographical memories, subjects should access their AMs for processing both true and false sentences. Indeed, since the hippocampus is the structure engaged in the initial access to AMs (Daselaar et al., 2008), both types of trial may have recruited it to the same extent.

Since our aim was to investigate which regions of the AM circuit can discriminate true from false AMs, we did not evaluate the recollection of other memories. Thus, we could not exclude that the same neural network could discriminate the truthfulness of other kind of memories.

We also examined whether retrieval of positive and negative emotional events from AM would exert distinctive effects on brain response. First, we assessed whether the discrimination between true and false events in the two groups occurred using brain activity extracted at different time points. No temporal differences were found between subjects who retrieved memories from their wedding and subjects who recalled events from the funeral of a loved one. Moreover, we did not find any significant difference in the spatial overlap of the decoding maps of the two subgroups, thus suggesting that emotional valence did not affected neither the temporal nor the spatial pattern of

activity during the retrieval. Indeed, decoding negative and positive autobiographical episodes was a challenging task with fMRI data and in a previous attempt Nawa and colleagues reported accuracies at chance level using an across-participants approach, whereas only half of the sample yielded a significant decoding with a within-participant approach (Nawa and Ando, 2014).

The choice of evaluating the two events (i.e., weddings and funerals) was based on the extensive evidence that emotionally arousing experiences are well-remembered (Brown and Kulik, 2003). Memories of unpleasant occasions, such as an automobile accident, a mugging, or the death of a loved one, are retrieved better than memories of routine days (Pillemer, 1984; Bohannon, 1988; Conway, 1995; Neisser et al., 1996; Sharot et al., 2007). Memories of pleasant occasions, such as birthdays, holidays, and weddings, are also well-retained (Buchanan, 2007). Thus, the strength of the memories of events varies with the emotional significance of the events.

The potential modulatory effect of the valence (either positive or negative) has been previously investigated, but with somehow conflicting results. In some cases, positive events were recalled more easily and directly with respect to negative ones, and led to an increased recovery of peripheral sensory and contextual details (Berntsen, 2002; Schaefer and Philippot, 2005; Kensinger and Schacter, 2006; Ford et al., 2009). The advantage for positive memories seems to be particularly evident when information is self-relevant (Holland and Kensinger, 2010) and some researchers have ascribed it to an overall bias toward accessing positive life experiences (Walker et al., 2003; Berntsen et al., 2011). On the other hand, some studies suggested that positive autobiographical memories are remembered less specifically

than negative events (Walker et al., 2003), and that "tunnel memories"—enhanced memory for the central details of an event—are limited to emotionally negative memories. Finally, negative past experiences are remembered with greater emotional intensity than positive memories (Berntsen, 2002).

Our data suggest that monitoring the veracity of highly emotional autobiographical memories requires a unique network of brain regions, irrespectively of the positive or negative valence of the event. In line with previous neuropsychological and neuroimaging evidence, we found that this memory system is mostly right-lateralized. This could reflect the emotional re-experiencing occurring during retrieval and is consistent with findings across different domains that suggest preferential right-hemisphere involvement in emotional and in social cognitive processes (see Svoboda et al., 2006 for a review).

In conclusion, we demonstrated that the entire AM network, with the exception of the medial temporal lobe regions, is engaged in monitoring the veracity of autobiographical memories. This process is mainly influenced by individual differences, rather than by the emotional valence of the experience. In line with previous neuroimaging studies (Miller and Van Horn, 2007), our data confirm that the patterns of brain activity during retrieval of AMs are consistent across subjects, though at different time points. This may be related to the unique manner in which subjects re-experience an autobiographical memory and to the different cognitive strategies used to process information. For this reason, a better understanding of the relationship between AM retrieval and the neural system that underlies this process should rely on the conjoint use of single-subject and group-level data analyses.

# 6. Conclusions

In this dissertation, I described four MVPA algorithms successfully applied in three different fMRI studies.

In the first experiment described in Chapter 2, a rank-based multi-class decoding algorithm was combined with a searchlight procedure to identify the regions in the left temporal and frontal cortex able to discriminate the seven Italian vowels during their listening, imagery and production. Furthermore, the BOLD activity of these regions was used to test the reconstruction of two possible alternative models, one based on motor, articulatory features and one comprising acoustic frequency-based descriptions. This process was performed using canonical correlation analysis, as detailed in Chapter 3.

In the second experiment reported in Chapter 4, we were able to predict brain activity of the left parietal areas elicited by thirty concrete nouns employing a representational similarity encoding algorithm. In this study, four different alternative models were tested: two semantic models built using language-based features, and two visual models, which provided a description of the shape of the objects and of their low-level spatial frequencies.

Finally, in the third fMRI experiment described in Chapter 5, we used a multivariate technique proposed by Mitchell and colleagues (2008) to recognize memories of real autobiographical events in each subject independently, highlighting both the time frame at which the successful recollection occurred and the brain networks involved in the process.

Overall, all these studies highlight the increased sensitivity of the MVPA approach, while the statistical robustness of all the procedures was achieved by means of permutation tests (Schreiber and Krekelberg, 2013).

# References


Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain and language*, 122(1), 42-54.

Akama, H. (2018). Individual typological differences in a neurally distributed semantic processing system: Revisiting the Science article by Mitchell et al. on computational neurolinguistics. *F1000Research*, 7.

Allen, E. J., Moerel, M., Lage-Castellanos, A., De Martino, F., Formisano, E., & Oxenham, A. J. (2018). Encoding of natural timbre dimensions in human auditory cortex. *Neuroimage*, 166, 60-70.

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., ... & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature neuroscience*, 10(6), 687.

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

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS biology*, 8(9), e1000489.

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., ... & Raizada, R. D. (2016a). Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cerebral Cortex*, 27(9), 4379-4395.

Anderson, A. J., Zinszer, B. D., & Raizada, R. D. (2016b). Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities. *NeuroImage*, 128, 44-53.

Andersson, J. L., Jenkinson, M., & Smith, S. (2007). Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2. FMRIB Analysis Group of the University of Oxford.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., & Knösche, T. R. (2006). Connectivity-based parcellation of Broca's area. *Cerebral cortex*, 17(4), 816-825.

Archila-Meléndez, M. E., Valente, G., Correia, J. M., Rouhl, R. P., van Kranen-Mastenbroek, V. H., & Jansma, B. M. (2018). Sensorimotor Representation of Speech Perception. Cross-Decoding of Place of Articulation Features during Selective Attention to Syllables in 7T fMRI. *eNeuro*, 5(2).

Ardila, A., Bernal, B., & Rosselli, M. (2016). Why Broca's area damage does not result in classical Broca's aphasia. *Frontiers in human neuroscience*, 10, 249.

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

Asaridou, S. S., Takashima, A., Dediu, D., Hagoort, P., & McQueen, J. M. (2015). Repetition suppression in the left inferior frontal gyrus predicts tone learning performance. *Cerebral cortex*, 26(6), 2728-2742.

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *The Journal of the Acoustical Society of America*, 63(5), 1535-1555.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS computational biology*, 9(8), e1003167.

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic bulletin & review*, 23(4), 1122-1142.

Barsalou, L. W. (2017). What does semantic tiling of the cortex tell us about semantics?. *Neuropsychologia* 105, 18-38.

Basilakos, A., Rorden, C., Bonilha, L., Moser, D., & Fridriksson, J. (2015). Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke*, 46(6), 1561-1566.

Beautemps, D., Badin, P., & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio-and labio-film data and articulatory-acoustic modeling. *The Journal of the Acoustical Society of America*, 109(5), 2165-2180.

Benuzzi, Francesca, et al. "Eight Weddings and Six Funerals: An fMRI Study on Autobiographical Memories." *Frontiers in behavioral neuroscience 12 (2018)*.

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & cognition*, 30(7), 1010-1020.

Berntsen, D., Rubin, D. C., & Siegler, I. C. (2011). Two versions of life: Emotionally negative and positive life events have different roles in the organization of life story and identity. *Emotion*, 11(5), 1190.

Berry, K. J., Johnston, J. E., & Mielke Jr, P. W. (2019). A Primer of Permutation Statistical Methods. Springer, 8(490), 978-3.

Bilenko, N. Y., & Gallant, J. L. (2016). Pyrcca: regularized kernel canonical correlation analysis in python and its applications to neuroimaging. *Frontiers in neuroinformatics*, 10, 49.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic bulletin & review*, 23(4), 1096-1108.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767-2796.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of theoretical Biology*, 38(2), 205-287.

Boersma, P. (2006). Praat: doing phonetics by computer. http://www. praat. org/.

Bohannon III, J. N. (1988). Flashbulb memories for the space shuttle disaster: A tale of two theories. *Cognition*, 29(2), 179-196.

Boller, F. (1978). Comprehension disorders in aphasia: A historical review. *Brain and Language*, 5(2), 149-165.

Bona, S., Cattaneo, Z., & Silvanto, J. (2015). The causal role of the occipital face area (OFA) and lateral occipital (LO) cortex in symmetry perception. *Journal of Neuroscience*, 35(2), 731-738.

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *Journal of Neuroscience*, 36(20), 5462-5471.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience*, 34(13), 4548-4557.

Bouchard, K. E., Conant, D. F., Anumanchipalli, G. K., Dichter, B., Chaisanguanthum, K. S., Johnson, K., & Chang, E. F. (2016). High-resolution, non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. *PLoS One*, 11(3), e0151327.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441), 327.

Bracci, S., & de Beeck, H. O. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, 36(2), 432-444.

Brown, R., Kulik, J. (2003). Flashbulb memories, in Memory and Emotion: The Making of Lasting Memories, ed. McGaugh J. L., editor. (New York, NY: Columbia University Press; ), 73–99

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological bulletin*, 133(5), 761.

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

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neuroscience*, 12(1), 1-47.

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in cognitive sciences*, 11(5), 219-227.

Cabeza, R., Prince, S. E., Daselaar, S. M., Greenberg, D. L., Budde, M., Dolcos, F., ... & Rubin, D. C. (2004). Brain activity during episodic retrieval of autobiographical and laboratory events: an fMRI study using a novel photo paradigm. *Journal of cognitive neuroscience*, 16(9), 1583-1594.

Cahill, L. (2010). Sex influences on brain and emotional memory: the burden of proof has shifted. In Progress in brain research (Vol. 186, pp. 29-40). Elsevier.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., ... & Weinberger, D. R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral cortex*, 9(1), 20-26.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4), 572-582.

Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The emergence of semantic meaning in the ventral temporal pathway. *Journal of cognitive neuroscience*, 26(1), 120-131.

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 57(1), 8-16.

Chakrabarti, S., Sandberg, H. M., Brumberg, J. S., & Krusienski, D. J. (2015). Progress in speech decoding from the electrocorticogram. *Biomedical Engineering Letters*, 5(1), 10-21.

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

Chang, K. M. K., Mitchell, T., & Just, M. A. (2011). Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation. NeuroImage, 56(2), 716-727.

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, 111(40), 14565-14570.

Chen, H. Y., Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2017). Are there multiple kinds of episodic memory? An fMRI investigation comparing autobiographical and recognition memory tasks. *Journal of Neuroscience*, 37(10), 2764-2775.

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

Chouinard, P. A., Meena, D. K., Whitwell, R. L., Hilchey, M. D., & Goodale, M. A. (2017). A tms investigation on the role of lateral occipital complex and caudal intraparietal sulcus in the perception of object form and orientation. *Journal of cognitive neuroscience*, 29(5), 881-895.

Cipolotti, L., & Maguire, E. A. (2003). A combined neuropsychological and neuroimaging study of topographical and non-verbal memory in semantic dementia. *Neuropsychologia*, 41(9), 1148-1159.

Clarke, A., & Tyler, L. K. (2015). Understanding what we see: how we derive meaning from vision. *Trends in cognitive sciences*, 19(11), 677-687.

Conant, D. F., Bouchard, K. E., Leonard, M. K., & Chang, E. F. (2018). Human sensorimotor cortex control of directly measured vocal tract movements during vowel production. *Journal of Neuroscience*, 38(12), 2955-2966.

Connolly, A. C., Gleitman, L. R., & Thompson-Schill, S. L. (2007). Effect of congenital blindness on the semantic representation of some everyday concepts. *Proceedings of the National Academy of Sciences*, 104(20), 8241-8246.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., ... & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, 32(8), 2608-2618.

Connolly, A. C., Sha, L., Guntupalli, J. S., Oosterhof, N., Halchenko, Y. O., Nastase, S. A., ... & Haxby, J. V. (2016). How the human brain represents perceived dangerousness or "predacity" of animals. *Journal of neuroscience*, 36(19), 5373-5384.

Conway, M. A. (1995). Flashbulb Memories. Brighton: Erlbaum.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3), 201.

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

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3), 162-173.

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences*, 114(17), E3370-E3371.

Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8), 1914-1928.

Craik, F. I., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological science*, 10(1), 26-34.

D'Ausilio, A., Craighero, L., & Fadiga, L. (2012b). The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics*, 25(5), 328-335.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012a). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882-887.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. Human brain mapping, 8(2-3), 109-114.

Damasio, A. R., & Geschwind, N. (1984). The neural basis of language. *Annual review of neuroscience*, 7(1), 127-147.

Dang, J., & Honda, K. (2002). Estimation of vocal tract shapes from speech sounds with a physiological articulatory model. *Journal of Phonetics*, 30(3), 511-532.

Daselaar, S. M., Rice, H. J., Greenberg, D. L., Cabeza, R., LaBar, K. S., & Rubin, D. C. (2007). The spatiotemporal dynamics of autobiographical memory: neural correlates of recall, emotional intensity, and reliving. *Cerebral cortex*, 18(1), 217-229.

Davis, C., Kleinman, J. T., Newhart, M., Gingis, L., Pawlak, M., & Hillis, A. E. (2008). Speech and language functions that require a functioning Broca's area. *Brain and language*, 105(1), 50-58.

De Angelis, V., De Martino, F., Moerel, M., Santoro, R., Hausfeld, L., & Formisano, E. (2018). Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds. *NeuroImage*, 180, 291-300.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., ... & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115(6), 1753-1768.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.

Downing, P. E., Wiggett, A. J., & Peelen, M. V. (2007). Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *Journal of Neuroscience*, 27(1), 226-233.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384(6605), 159.

Düzel, E., Habib, R., Guderian, S., & Heinze, H. J. (2004). Four types of novelty–familiarity responses in associative recognition memory of humans. *European Journal of Neuroscience*, 19(5), 1408-1416.

Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage*, 36(3), 511-521.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the national academy of sciences, 113(28), 7900-7905.

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences*, 97(11), 6150-6154.

Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cerebral cortex*, 25(12), 4772-4788.

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

Feng, G., Gan, Z., Wang, S., Wong, P. C., & Chandrasekaran, B. (2017). Task-General and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral cortex*, 28(9), 3241-3254.

Feredoes, E., & Postle, B. R. (2007). Localization of load sensitivity of working memory storage: quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis. *Neuroimage*, 35(2), 881-903.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., ... & Seidenberg, M. S. (2015). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, 26(5), 2018-2034.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences*, 112(9), 2871-2875.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47), S102.

Ford, J. H., Addis, D. R., & Giovanello, K. S. (2012). Differential effects of arousal in positive and negative autobiographical memories. *Memory*, 20(7), 771-778.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). " Who" is saying" what"? Brain-based decoding of human voice and speech. *Science*, 322(5903), 970-973.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, 443(7107), 85.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, 2(1), 45-53.

Fullerton, B. C., & Pandya, D. N. (2007). Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *Journal of Comparative Neurology*, 504(5), 470-498.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual review of psychology*, 49(1), 87-115.

Gadian, D. G., Aicardi, J., Watkins, K. E., Porter, D. A., Mishkin, M., & Vargha-Khadem, F. (2000). Developmental amnesia associated with early hypoxic–ischaemic injury. *Brain*, 123(3), 499-507.

Gainotti, G. (2010). The influence of anatomical locus of lesion and of gender-related familiarity factors in category-specific semantic disorders for animals, fruits and vegetables: a review of single-case studies. *Cortex*, 46(9), 1072-1087.

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

Geake, J. G., & Hansen, P. C. (2005). Neural correlates of intelligence as revealed by fMRI of fluid analogies. *NeuroImage*, 26(2), 555-564.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4), 870-878.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual review of psychology*, 54(1), 91-114.

Ghio, M., Vaghi, M. M. S., Perani, D., & Tettamanti, M. (2016). Decoding the neural representation of fine-grained conceptual categories. *Neuroimage*, 132, 93-103.

Gick, B., & Stavness, I. (2013). Modularizing speech. *Frontiers in Psychology*, 4, 977.

Gilboa, A. (2004). Autobiographical and episodic memory—one and the same?: Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, 42(10), 1336-1349.

Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J., & Moscovitch, M. (2004). Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex*, 14(11), 1214-1225.

Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: a functional segregation within Broca's Area. *Neuroimage*, 114, 294-302.

Grabski, K., Schwartz, J. L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., ... & Sato, M. (2013). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, 26(3), 384-408.

Graham, K. S., Lee, A. C., Brett, M., & Patterson, K. (2003). The neural basis of autobiographical and semantic memory: new evidence from three PET studies. *Cognitive, Affective, & Behavioral Neuroscience*, 3(3), 234-254.

Gray, J. R., & Braver, T. S. (2002). Personality predicts working-memory—related activation in the caudal anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, 2(1), 64-75.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature neuroscience*, 6(3), 316.

Greenberg, D. L., Rice, H. J., Cooper, J. J., Cabeza, R., Rubin, D. C., & LaBar, K. S. (2005). Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval. *Neuropsychologia*, 43(5), 659-674.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(7), 4259-4264.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7), e159.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning*, 45(2), 171-186.

Handjaras, G., Bernardi, G., Benuzzi, F., Nichelli, P. F., Pietrini, P., & Ricciardi, E. (2015). A topographical organization for action representation in the human brain. *Human brain mapping*, 36(10), 3832-3844.

Handjaras, G., Leo, A., Cecchetti, L., Papale, P., Lenci, A., Marotta, G., ... & Ricciardi, E. (2017). Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia*, 105, 39-49.

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

Hardcastle, W. J., Laver, J., & Gibbon, F. E.. (2010). The handbook of phonetic sciences (2nd Edition). Wiley-Blackwell.

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. Neuroscience & Biobehavioral Reviews, 94, 31-44.

Harris, S., Sheth, S. A., & Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Annals of neurology*, 63(2), 141-147.

Hausfeld, L., Riecke, L., & Formisano, E. (2018). Acoustic and higher-level representations of naturalistic auditory scenes in human auditory and frontal cortex. *NeuroImage*, 173, 472-483.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404-416.

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron*, 87(2), 257-270.

Hebscher, M., Levine, B., & Gilboa, A. (2018). The precuneus and hippocampus contribute to individual differences in the unfolding of spatial representations during episodic autobiographical memory. *Neuropsychologia*, 110, 123-133.

Heim, S., Eickhoff, S. B., & Amunts, K. (2008). Specialisation in Broca's region for semantic, phonological, and syntactic fluency?. *Neuroimage*, 40(3), 1362-1368.

Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain and language*, 119(3), 214-220.

Hill, A. C., Laird, A. R., & Robinson, J. L. (2014). Gender differences in working memory networks: a BrainMap meta-analysis. *Biological psychology*, 102, 18-29.

Hinke, R. M., Hu, X., Stillman, A. E., Kim, S. G., Merkle, H., Salmi, R., & Ugurbil, K. (1993). Functional magnetic resonance imaging of Broca's area during internal speech.

*Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*.

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of life reviews*, 7(1), 88-131.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* 28, 321–377

Huang, J., Carr, T. H., & Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human brain mapping,* 15(1), 39-53.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *Journal of Physiology-Paris*, 102(1-3), 31-34.

Ibos, G., & Freedman, D. J. (2016). Interaction between spatial and feature attention in posterior parietal cortex. *Neuron*, 91(4), 931-943.

Jackson, R. L., Hoffman, P., Pobric, G., & Ralph, M. A. L. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience*, 36(5), 1490-1501.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., & Woolrich, M. W. (2012). Smith SM. FSL. *Neuroimage*, 62, 782-90.

Josephs, K. A., Duffy, J. R., Strand, E. A., Whitwell, J. L., Layton, K. F., Parisi, J. E., ... & Dickson, D. W. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain*, 129(6), 1385-1398.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia*, 83, 201-226.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, 97(22), 11793-11799.

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of neurophysiology*, 115(4), 2246-2250.

Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., & Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *Journal of Neuroscience*, 30(4), 1314-1321.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of cognitive neuroscience*, 14(5), 785-794.

Kemmerer, D. (2017). Categories of object concepts across languages and brains: The relevance of nominal classification systems to cognitive neuroscience. *Language, Cognition and Neuroscience*, 32(4), 401-424.

Kensinger, E. A., & Schacter, D. L. (2006). When the Red Sox shocked the Yankees: Comparing negative and positive memories. *Psychonomic Bulletin & Review*, 13(5), 757-763.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, 10(11), e1003915.

King, D. R., de Chastelaine, M., Elward, R. L., Wang, T. H., & Rugg, M. D. (2015). Recollection-related increases in functional connectivity predict individual differences in memory accuracy. *Journal of Neuroscienc*e, 35(4), 1763-1772.

Kolasinski, J., Makin, T. R., Jbabdi, S., Clare, S., Stagg, C. J., & Johansen-Berg, H. (2016). Investigating the stability of fine-grain digit somatotopy in individual human participants. *Journal of Neuroscience*, 36(4), 1113-1127.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature neuroscience*, 11(2), 224.

Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, 43(1), 97-109.

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4), 649-662.

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

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008b). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 4.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126-1141.

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2014). A conceptual framework of computations in mid-level vision. *Frontiers in Computational Neuroscience*, 8, 158.

Kumar, M., Federmeier, K. D., Fei-Fei, L., & Beck, D. M. (2017). Evidence for similar patterns of neural activity elicited by picture-and word-based representations of natural scenes. *NeuroImage*, 155, 422-436.

Kumari, V., Williams, S. C., & Gray, J. A. (2004). Personality predicts brain responses to cognitive demands. *Journal of Neuroscience*, 24(47), 10636-10641.

Kwok, V. P., Dan, G., Yakpo, K., Matthews, S., & Tan, L. H. (2016). Neural systems for auditory perception of lexical tones. *Journal of Neurolinguistics*, 37, 34-40.

Ladefoged, P., & Disner, S. F. (2012). Vowels and consonants. John Wiley & Sons.

Laukkanen, A. M., Horáč ek, J., Krupa, P., & Švec, J. G. (2012). The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results. *Biomedical Signal Processing and Control*, 7(1), 50-57.

Laurent, R., Barnaud, M. L., Schwartz, J. L., Bessière, P., & Diard, J. (2017). The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception. *Psychological review*, 124(5), 572.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J. H., Lee, S., & Kim, K. (2006). Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage*, 29(2), 578-586.

Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience*, 32(11), 3942-3948.

Leeds, D. D., Seibert, D. A., Pyles, J. A., & Tarr, M. J. (2013). Comparing visual representations across human fMRI and computational vision. *Journal of vision*, 13(13), 25-25.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: A set of semantic feature norms from the congenitally blind. *Behavior research methods*, 45(4), 1218-1233.

Leo, A., Handjaras, G., Bianchi, M., Marino, H., Gabiccini, M., Guidi, A., ... & Ricciardi, E. (2016). A synergy-based hand control is encoded in human motor cortical areas. *Elife*, 5, e13420.

Leoni, F. A., & Maturi, P. (2002). Manuale di fonetica. Roma: Carocci.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74(6), 431.

Liu, T., Hospadaruk, L., Zhu, D. C., & Gardner, J. L. (2011). Feature-specific attentional priority signals in human cortex. *Journal of Neuroscience*, 31(12), 4484-4495.

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cerebral cortex*, 13(12), 1334-1343.

Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., ... & Greenlee, J. D. (2016). Functional segregation of cortical regions underlying speech timing and articulation. *Neuron*, 89(6), 1187-1193.

Luria, A. R. (1966). Higher cortical functions in man. New York: Consultants Bureau Enterprises.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual review of psychology*, 60, 27-51.

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain?. *Trends in cognitive sciences*, 15(3), 97-103.

Mahon, B. Z., Kumar, N., & Almeida, J. (2013). Spatial frequency tuning reveals interactions between the dorsal and ventral visual systems. *Journal of cognitive neuroscience*, 25(6), 862-871.

Maillet, D., & Rajah, M. N. (2014). Age-related differences in brain activity in the subsequent memory paradigm: a meta-analysis. *Neuroscience & Biobehavioral Reviews*, 45, 246-257.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., ... & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18), 8135-8139.

Marcus, D., Harwell, J., Olsen, T., Hodge, M., Glasser, M., Prior, F., ... & Van Essen, D. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in neuroinformatics*, 5, 4.

Marie, P., Cole, M. F., & Cole, M. (1971). Papers on Speech Disorders: Compiled and Transl. Hafner.

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

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S. J., & Holmes, A. P. (2000). Variability in fMRI: an examination of intersession differences. *Neuroimage*, 11(6), 708-734.

McKinnon, M. C., Black, S. E., Miller, B., Moscovitch, M., & Levine, B. (2006). Autobiographical memory in semantic dementia: Implications for theories of limbic-neocortical interaction in remote memory. *Neuropsychologia*, 44(12), 2421-2429.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, 37(4), 547-559.

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

Miller, M. B., & Van Horn, J. D. (2007). Individual variability in brain activations associated with episodic retrieval: a role for large-scale databases. *International journal of psychophysiology*, 63(2), 205-213.

Mitchell, T. M. (1997). Machine learning. McGraw Hill.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine learning*, 57(1-2), 145-175.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880), 1191-1195.

Moore, C. A. (1992). The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. *Journal of Speech, Language, and Hearing Research*, 35(5), 1009-1023.

Mruczek, R. E., von Loga, I. S., & Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *Journal of Neurophysiology*, 109(12), 2883-2896.

Murakami, T., Kell, C. A., Restle, J., Ugawa, Y., & Ziemann, U. (2015). Left dorsal speech stream components and their contribution to phonological processing. *Journal of Neuroscience*, 35(4), 1411-1422.

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

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., ... & Haxby, J. V. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex*, 27(8), 4277-4291.

Nawa, N. E., & Ando, H. (2014). Classification of self-driven mental tasks from whole-brain activity patterns. *PloS one*, 9(5), e97296.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S. A. S. A., ... & Boone, K. (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*, 51(6), 1546-1554.

Neisser, U. (1996). Remembering the earthquake: Direct experience vs. hearing the news. *Memory*, 4(4), 337-358.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1), 1-25.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology*, 10(4), e1003553.

Noppeney, U., Friston, K. J., & Price, C. J. (2003). Effects of visual deprivation on the organization of the semantic system. *Brain*, 126(7), 1620-1627.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, 10(9), 424-430.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., ... & Rauschecker, J. P. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human brain mapping*, 27(7), 562-571.

Obleser, J., Leaver, A., VanMeter, J., & Rauschecker, J. P. (2010). Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Frontiers in psychology*, 1, 232.

Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and language*, 98(1), 112-117.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3), 145-175.

Papale, P., Betta, M., Handjaras, G., Malfatti, G., Cecchetti, L., Rampinini, A., ... & Leo, A. (2019). Common spatiotemporal processing of visual features shapes object representation. *Scientific reports*, 9(1), 7601.

Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K. N., Pietrini, P., & Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eneuro*, 5(3).

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cerebral cortex*, 19(9), 2156-2165.

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. *Journal of Neuroscience*, 34(1), 163-170.

Penfield, W., & Roberts, L. (2014). Speech and brain mechanisms (Vol. 62). Princeton University Press.

Pereira, F., Botvinick, M., & Detre, G. (2013). Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments. *Artificial intelligence*, 194, 240-252.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1), S199-S209.

Piefke, M., & Fink, G. R. (2005). Recollections of one's own past: the effects of aging and gender on the neural mechanisms of episodic autobiographical memory. *Anatomy and embryology*, 210(5-6), 497-512.

Pillemer, D. B. (1984). Flashbulb memories of the assassination attempt on President Reagan. *Cognition*, 16(1), 63-80.

Poeppel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, 92(1-2), 1-12.

Poldrack, R. A., Mumford, J. A., Schonberg, T., Kalar, D., Barman, B., & Yarkoni, T. (2012). Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS computational biology*, 8(10), e1002707.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, 59(3), 2142-2154.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience*, 35(7), 3276-3284.

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. *Neuroimage*, 62(2), 816-847.

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of cognitive neuroscience*, 28(5), 680-692.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences*, 17(9), 458-470.

Rampinini, A. C., & Ricciardi, E. (2017). In favor of the phonemic principle: a review of neurophysiological and neuroimaging explorations into the neural correlates of phonological competence. *Studi e Saggi Linguistici*, 55(1), 95-123.

Rampinini, A. C., Handjaras, G., Leo, A., Cecchetti, L., Betta, M., Ricciardi, E., ... & Pietrini, P. (2019). Formant space reconstruction from brain activity in frontal and temporal regions coding for heard vowels. *Frontiers in human neuroscience*, 13, 32.

Rampinini, A. C., Handjaras, G., Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G., & Pietrini, P. (2017). Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. *Scientific reports*, 7(1), 17029.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22), 11800-11806.

Reiterer, S., Erb, M., Grodd, W., & Wildgruber, D. (2008). Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging and Behavior*, 2(1), 1-10.

Ricciardi, E., & Pietrini, P. (2011). New light from the dark: what blindness can teach us about brain function. *Current opinion in neurology*, 24(4), 357-363.

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture?. *Neuroscience & Biobehavioral Reviews*, 41, 64-77.

Ricciardi, E., Handjaras, G., & Pietrini, P. (2014). The blind brain: How (lack of) vision shapes the morphological and functional architecture of the human brain. *Experimental Biology and Medicine*, 239(11), 1414-1420.

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *Journal of Neuroscience*, 34(26), 8837-8844.

Richmond, K., King, S., & Taylor, P. (2003). Modelling the uncertainty in recovering articulation from acoustics. *Computer Speech & Language*, 17(2-3), 153-172.

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*, 4(2), 93.

Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, 269(5225), 853-855.

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. *Journal of Cognitive Neuroscience*, 9(3), 295-317.

Romanski, L. M., & Averbeck, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annual review of neuroscience*, 32, 315-346.

Rypma, B., & D'Esposito, M. (2000). Isolating the neural mechanisms of age-related changes in human working memory. *Nature neuroscience*, 3(5), 509.

Rypma, B., Berger, J. S., & D'esposito, M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *Journal of cognitive neuroscience*, 14(5), 721-731.

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., & Formisano, E. (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences*, 114(18), 4799-4804.

Schaefer, A., & Philippot, P. (2005). Selective effects of emotion on the phenomenal characteristics of autobiographical memories. *Memory*, 13(2), 148-160.

Schaefer, A., Braver, T. S., Reynolds, J. R., Burgess, G. C., Yarkoni, T., & Gray, J. R. (2006). Individual differences in amygdala activity predict response speed during working memory. *Journal of Neuroscience*, 26(40), 10120-10128.

Schomaker, J., & Meeter, M. (2015). Short-and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience & Biobehavioral Reviews*, 55, 268-279.

Schomers, M. R., & Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in human neuroscien*ce, 10, 435.

Schreiber, K., & Krekelberg, B. (2013). The statistical analysis of multi-voxel patterns in functional imaging. *PLoS One*, 8(7), e69328.

Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336-354.

Scolari, M., Seidl-Rathkopf, K. N., & Kastner, S. (2015). Functions of the human frontoparietal attention network: Evidence from neuroimaging. *Current opinion in behavioral sciences*, 1, 32-39.

Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5), 550-571.

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist*, 19(1), 43-61.

Seghier, M. L., & Price, C. J. (2012). Functional heterogeneity within the default network during semantic processing and speech production. *Frontiers in psychology*, 3, 281.

Seghier, M. L., Fagan, E., & Price, C. J. (2010). Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *Journal of Neuroscience*, 30(50), 16809-16817.

Seghier, M. L., Lazeyras, F., Pegna, A. J., Annoni, J. M., & Khateb, A. (2008). Group analysis and the subject factor in functional magnetic resonance imaging: Analysis of fifty right-handed healthy subjects in a semantic language task. *Human brain mapping*, 29(4), 461-477.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences*, 99(16), 10917-10922.

Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Sciences*, 104(1), 389-394.

Sheldon, S., Farb, N., Palombo, D. J., & Levine, B. (2016). Intrinsic medial temporal lobe connectivity relates to individual differences in episodic autobiographical remembering. *Cortex*, 74, 206-216.

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

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. Neuroimage, 54(3), 2418-2425.

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

Skipper, J. I., Devlin, J. T., & 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 and language*, 164, 77-105.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, 25(1), 76-89.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23, S208-S219.

Snowden, J., Griffiths, H., & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive neuropsychology*, 11(3), 265-288.

Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage*, 20(4), 1944-1954.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America*, 27(3), 484-493.

Strappini, F., Gilboa, E., Pitzalis, S., Kay, K., McAvoy, M., Nehorai, A., & Snyder, A. Z. (2017). Adaptive smoothing based on Gaussian processes regression increases the sensitivity and specificity of fMRI data. *Human brain mapping*, 38(3), 1438-1459.

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, 44(12), 2189-2208.

Tailby, C., Rayner, G., Wilson, S., & Jackson, G. (2017). The spatiotemporal substrates of autobiographical recollection: using event-related ICA to study cognitive networks in action. *Neuroimage*, 152, 237-248.

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

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

Toda, T., Black, A. W., & Tokuda, K. (2008). Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture model. *Speech Communication*, 50(3), 215-227.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.

Tulving E. (1983). Elements of EpisodicMemory. Oxford: Clarendon Press

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, 53(1), 1-25.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, 8(3), 198-204.

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., ... & Taylor, K. I. (2013). Objects and categories: feature statistics and object processing in the ventral stream. *Journal of Cognitive Neuroscience*, 25(10), 1723-1735.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1), 273-289.

Van Eede, M., Macrini, D., Telea, A., Sminchisescu, C., & Dickinson, S. S. (2006, August). Canonical skeletons for shape matching. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 2, pp. 64-69). IEEE.

Van Horn, J. D., Grafton, S. T., & Miller, M. B. (2008). Individual variability in brain activity: a nuisance or an opportunity?. *Brain imaging and behavior*, 2(4), 327.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383(6597), 254.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, 277(5324), 376-380.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2013). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, 24(7), 1767-1777.

Wager, T. D., Sylvester, C. Y. C., Lacey, S. C., Nee, D. E., Franklin, M., & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *Neuroimage*, 27(2), 323-340.

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way!. *Review of General Psychology*, 7(2), 203-210.

Wang, X., Peelen, M. V., Han, Z., Caramazza, A., & Bi, Y. (2016). The role of vision in the neural representation of unique entities. *Neuropsychologia*, 87, 144-156.

Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *Journal of Neuroscience*, 35(36), 12545-12559.

Watson, D. M., Young, A. W., & Andrews, T. J. (2016). Spatial properties of objects predict patterns of neural response in the ventral visual pathway. *NeuroImage*, 126, 173-183.

Wiggs, C. L., Weisberg, J., & Martin, A. (1998). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia*, 37(1), 103-118.

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

Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W. F., & Heiss, W. D. (2005). Role of the contralateral inferior frontal gyrus in recovery of language function in poststroke aphasia: a combined repetitive transcranial magnetic stimulation and positron emission tomography study. *Stroke*, 36(8), 1759-1763.

Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E., & Smith, S. M. (2016). Faster permutation inference in brain imaging. *NeuroImage*, 141, 502-516.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta psychologica*, 132(2), 173-189.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, 8(8), 665.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3), 1116-1128.

Zhang, Q., Hu, X., Luo, H., Li, J., Zhang, X., & Zhang, B. (2016). Deciphering phonemes from syllables in blood oxygenation level-dependent signals in human superior temporal gyrus. *European Journal of Neuroscience*, 43(6), 773-781.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2), 248-248.