Maastricht University

Neural coding of speech and language : fMRI and EEG studies

Citation for published version (APA):

Correia, J. M. (2015). *Neural coding of speech and language : fMRI and EEG studies*. [Doctoral Thesis, Maastricht University]. Maastricht University. https://doi.org/10.26481/dis.20150327jc

Document status and date:
Published: 01/01/2015

DOI:
10.26481/dis.20150327jc

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 26 Apr. 2024

Neural Coding of Speech and Language:

fMRI and EEG studies

João M. Correia

© João M. Correia, Maastricht 2015.

All rights reserved.

The work presented in this thesis was funded by the European Union Marie Curie Initial Training Network and was conducted at Maastricht University.

Cover	Bruno Maltez
Production	CPI Wöhrmann Print Services, B.V.
ISBN	978-94-6203-811-0

Neural Coding of Speech and Language:

fMRI and EEG studies

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public on Friday 27th March 2015 at 12.00 hours

by

João M. Correia

Promotor

Prof. Dr. Bernadette M. Jansma

Co-promotor

Dr. Milene Bonte

Assessment Committee

Prof. Dr. Rainer Goebel (Chair)

Prof. Dr. Riitta Salmelin (Aalto University, Helsinki, Finland)

Dr. Matthew Davis (Medical Research Council, Cambridge, UK)

Dr. Federico De Martino

CONTENTS

1	General Introduction	7
2	Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe	35
3	EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations	61
4	Brain-based decoding of spoken syllables isolates articulatory representations during passive listening	93
5	Information-based connectivity – IBC: a functional connectivity approach for fMRI using multivariate classification	123
6	Summary and future directions	151
7	Knowledge Valorization	161
	Acknowledgements	166
	Publications	172
	Curriculum Vitae	174

CHAPTER 1

General Introduction

Humans are social animals. When deprived of normal communication, our quality of life, independence and social integration become significantly reduced. Furthermore, in human societies, individuals benefit from the knowledge and position of others regarding resources. In this respect, language is not merely a communication tool, but a tool for survival. Language is also a key aspect of our personal identity, and one that has been used throughout history to impose invisible borders, separating cultures and societies. Overall, the ambit of language outreaches aspects related with living in community. It is a unique asset to human life that provides the ability to organize and express thoughts, ideas and emotions, both within ourselves and with others, regarding the past, the present and importantly, the future.

Verbal communication is innate to humans across all countries and cultures. Currently, 7106 languages are spoken worldwide (Lewis et al, 2014). A healthy newborn child exposed to any of these languages would effortlessly acquire the necessary ability to comprehend and express itself in that language via speech (Cutler, 2012). This ability is specific to humans, but independent of language specific characteristics. Hence, something in humans is specialized for speech and something in all languages facilitates its acquisition. The neurobiological mechanisms of speech and language remain however unknown and constitute a crucial scientific milestone to address speech and communication disorders that affect the quality of life of a significant part of the world population, and to further our understanding of human cognition.

This thesis investigates neural mechanisms involved in speech perception, and processes at the interface to higher-order linguistic networks. Specifically, we researched the interface of speech input with language independent neural representations underlying the meaning of spoken words, and underlying articulatory properties of spoken syllables. Based on brain activity obtained while subjects listen to speech in combination with multivariate classification methods, we studied speech processing at the informational level, as individual words and syllables are presented to healthy adults. Chapter 2 and 3 are spoken word comprehension studies that exploit the particular capacity of bilingual listeners to comprehend speech from two different languages in combination

with the higher sensitivity of multivariate classification to investigate the neural representation of spoken words invariant to language specific form. As bilinguals are able to access the same concepts from two equivalent concepts, but physically distinct words across two languages, this ability serves as a proxy to research the neural representation of semantic concepts independent of word specific characteristics. In chapter 2, brain responses to spoken words were obtained using fMRI and in chapter 3 using EEG technology, thus exploiting their advantageous spatial and temporal resolutions, respectively. In chapter 4 we investigated sensorimotor integration in speech perception, a neural process proposed to transform auditory signals to motor programs, thus crucial in a variety of speech related functions, from speech acquisition and development, monitoring of self-produced speech and to aspects of speech perception. Here, fMRI response patterns to spoken syllables were analyzed based on articulatory properties using a generalization approach that allowed us disentangling acoustic and articulatory properties inherent to speech. In chapter 5, we developed and validated a novel brain connectivity method for neuroimaging data that exploits information-based decoding analysis as generated in multivariate classification across independent regions of interest.

Throughout this introduction, we first provide relevant background on speech processing in the brain. Next, we describe methodological background on imaging techniques employed throughout this thesis and their ability to investigate speech perception. Thereafter, we introduce the principles of multivariate classification techniques for neuroimaging research and, in particular, focus on a cross-validation strategy adopted in this thesis that assured the generalization of experimental information across variation of distinct conditions. Finally, we discuss unresolved challenges of speech processing research addressed in this thesis, including the access and representation of conceptual knowledge and sensorimotor integration in speech perception, as well as, describe and propose how the neuroimaging technologies and analysis methods employed in this thesis can advance the research on these topics in particular, and on the research of speech and language in general.

Investigating speech perception in the brain

Speech sounds arriving to our ears are automatically transformed by the cochlea (inside the inner ears) into electric signals encoding spectro-temporal acoustic properties of speech and are subsequently transmitted via the auditory pathways to the primary auditory cortex (PAC) bilaterally. PAC is approximately localized in the medial part of the Heschl's gyrus (HG) within the temporal sulcus, and preserves a similar tonotopic (spectral based) organization as initially generated by the cochlea (Lauter et al., 1985; Formisano et al., 2003). Further auditory processing of speech sounds involves the left and right superior temporal gyrus (STG) and superior temporal sulcus (STS). Higher-order aspects of speech processing, enabling access to semantic conceptual information and access to articulatory gestures required for speech production remain largely unknown and is investigated in this thesis.

Early investigations of speech processing in the brain followed localism views of brain function (Gall, 1835) and relied on the study of behavior in patients with brain lesions. Following the findings from Carl Wernicke (1874) and Paul Broca (1861, 1865), the classical view of speech processing (Wernicke, 1874; Eggert, 1977; Lichtheim, 1885; Geschwind, 1965) postulated a framework for speech relying on the existence of two fundamental modules, the auditory and the motor modules. The auditory module (or Wernicke's area) included areas within the posterior superior temporal lobe and was proposed to be crucial for speech comprehension. The motor module (or Broca's areas) included areas in the inferior frontal lobe, specifically the pars opercularis and pars triangularis within the left inferior frontal gyrus (IFG) and was proposed to be crucial for speech production. Accordingly, lesions to the auditory and motor modules were observed to result in receptive and expressive aphasia, respectively. Additionally, the classical view included evidence from a type of conduction aphasia that although not directly affecting receptive or expressive speech, had an impact on normal speech functions. Conduction aphasia was defined as affecting the connectivity between the auditory and motor modules of speech and was observed following lesions along the white-matter fibers denominated

arcuate fasciculus (Wernicke, 1974). Lesions to this auditory-motor connection resulted in more subtle speech deficits, for example related to the ability to repeat speech input or to the presence of paraphasic errors in speech production.

At present, and in light of the classical view of speech processing, speech researchers still attribute fundamental speech functions to the three modules, auditory, motor and auditory-motor connection. However, new insights indicate that speech processing requires a far more inter-connected system in which no module fulfills an independent function. In particular, aphasias of speech are now believed to encompass a more variable cluster of symptoms than reported before, including speech comprehension deficits due to Broca's and conduction aphasia (Hickok and Poeppel, 2004), as well as, speech production deficits due to Wernicke's aphasia (Baker et al., 1981). Importantly, during developmental stages of speech acquisition, the network of speech and language is proposed to operate conjointly, sharing experience and creating functional links across multiple relevant speech and language areas (Pulvermuller and Fadiga, 2010). Furthermore, the left lateralization postulated by the classical view may not capture the bilateral complexity required in speech processing. Evidence from the wada test (pre-surgical hemispheric anesthesia, Boatman, 2004; McGlone, 1984) and PET (positron emission tomography) (Scott et al., 2004) earlier suggested that the right hemisphere alone is capable to assume many of the speech related functions previously thought to be exclusively controlled by the left hemisphere.

Today, important advancements in neuroimaging have made possible to investigate the healthy human brain at work, and in combination with efforts from brain lesion and behavioral observations are progressing the neuroscience of speech and language. Building on the original auditory-motor distinction, recent models of speech processing (Hickok and Poeppel, 2007; Rauschecker, 2009) propose a dual-stream, task modulated, network that includes a ventral stream involved in tasks related with semantic access, such as in speech comprehension, and a dorsal stream involved in tasks related with perception for action, such as in speech repetition. This cortical activity spread of speech

processing is proposed to rely on anatomical connections between posterior superior temporal regions engaged in speech input analysis and multiple distributed regions of the brain fundamental for speech, language and cognition. Furthermore, a temporal synchronization of neural responses within early auditory regions and across the network of speech and language is proposed to orchestrate neural processes tracking temporal signatures of speech signals that enable our capacity to perceive spoken language in such a fast and efficient manner (Giraud and Poeppel, 2012; Lu et a., 2007). Of particular relevance to the research pursued in this thesis, below we introduce neuroscientific findings that reveal fundamental aspects of our ability to develop and use verbal communication in everyday life. The first relates to the representation of semantic knowledge in the brain and its access from speech input. The second relates to the transformation of speech input onto motoric representations that guides and monitors speech production and that may be also essential in normal speech perception.

The representation of semantic knowledge

Most neuroscientists agree that there is no "grandmother cell" (Jerry Lettvin) (Barlow, 1995), and that semantic knowledge representation in abstract form emerges instead from connected assemblies of multiple cells. Detailed aspects of such neural representation, in particular related to local versus distributed and abstract versus modality-specific nature of these representations remain elusive. Localized brain lesions show the capacity to disrupt semantic knowledge selectively (Dronkers et al., 2007; Damasio et al., 1996), thus suggesting that its representation, access or retrieval includes local neural populations and is not the result of a 'universal' brain connectivity (Marie, 1897; Fodor, 1983). By assessing the location of brain lesions resulting in semantic deficits across a large number of patients, the anterior temporal lobe (ALT) has been pointed as a possible locus of semantic knowledge representation in abstract form (Damasio et al., 1996). Neuroimaging studies on healthy subjects have later agreed on the relevance of the ATL (left lateralized) in combining the representation of semantic knowledge across presentation modalities, such as visual and auditory (e.g., Chan et al., 2011a;

Visser et al., 2012). Another influential neuroscientific perspective (Patterson et al., 2007) suggests that semantic knowledge representation is distributed throughout the cortex involving, for example, neural systems required for perception and action, but that its integration relies on integration regions, also known as semantic hubs. In line with brain lesion evidence, a central semantic hub within the ATL (left lateralized) is proposed to host local semantic representations in abstract form. Beyond the localization of the nodes of a semantic representation network, the CDZ (convergence divergence zones) model (Damasio, 1989; Meyer and Damasio, 2009) proposes a temporal synchrony of distributed property-based representations within early sensorial and motoric activations. Semantic representations would thus be a product of an orchestrated network of higher-order integration zones and experience-based distributed activations on the cortex. Categorical divisions in semantic knowledge, such as different objects (e.g., tools, animals or dwellings) may in turn emerge from their intrinsic sensorimotor properties within sensory, motor and integration regions (Martin, 2007). Alternatively to distributed property-based views of semantic knowledge organization, evolutionary accounts (Caramazza and Mahon, 2006) propose a more focal organization shaped by survival constraints throughout human history. In the speech domain, the access to semantic knowledge in abstract form remains unresolved and is addressed in this thesis by investigating language independent representations of spoken words.

The importance of sensorimotor integration in speech perception

Sensorimotor integration, linking the neural systems involved in perceiving and producing speech is crucial for verbal communication. A system required for motor planning and execution covers a great part of the frontal and parietal lobes. This vast cortical extension provides humans the ability for motor control, such as that necessary for speech production. Sensorimotor integration in speech refers to the functional link between the neural systems involved in perceiving and producing speech. It is behaviorally related to speech production, for example during the development of speech acquisition or auditory monitoring of self-produced speech. The lack of exposure to speech

in early life significantly affects the ability to develop a normal speech production system and exposure to different accents can alter our own accent in a short time-frame (Hickok et al., 2011; Guenther and Vladusich, 2012). Moreover, the temporal-acoustic disturbance of self-produced speech sounds directly affects fluent speech production (Guenther and Vladusich, 2012). Whether sensorimotor integration also plays a role in speech perception remains an ongoing debate (Pulvermuller and Fadiga 2010; Hickok, 2010; Gallantuci, 2006) that links to models of speech perception relying on motor speech perception systems (Liberman, 1967). The motor-theory-of-speech-perception proposed by Liberman (1967) originates from hebbian principles (Hebb, 1949) related to the temporal correlation of brain activity during auditory and motor tasks, and to the difficulty of acoustic-based models alone to express co-articulatory aspects of consecutive speech sounds in words and sentences. Conversely, Hickok and Poeppel (2007) posit that speech perception is an auditory event, but one that is integrated with speech production systems in healthy persons. Such integration is proposed to rely on a brain region specialized in the translation of auditory to motor and motor to auditory representations of speech. This specialized sensorimotor integration region, at the border between the posterior sylvian fissure and the inferior parietal lobe (Spt – Sylvian parietal temporal) is involved in both speech perception and speech production, and most crucially, is involved in speech repetition and humming (Hickok et al., 2003). Furthermore, aphasias resulting from brain lesions (Geschwind, 1965) have improved our understanding of auditory to motor connectivity during speech. Specifically, conduction aphasia is described as a speech syndrome originating from a lesion disrupting the functional interface between the posterior temporal lobe and the pre-frontal and motor cortices. Conduction aphasics exhibit speech repetition deficits, however they also include spontaneous production and perception difficulties, suggesting that auditory and production speech areas are not independent and separable computational modules, but part of an integrated network (Hickok and Poepple, 2007; Pulvermuller and Fadiga, 2010). The locus of conduction aphasia is in turn not fully identified and both damage in anatomical connectivity (Jones et al., 2014; Geschwind, 1965) or in the neural

computations within sensorimotor integration regions (e.g., Spt area, Buchsbaum 2011; Guenther and Vladusich, 2012; Hickok et al., 2011). Overall, the underlying neural networks of sensorimotor integration remain unknown. Specifically, researching the representation of articulatory features of speech gestures during speech perception (Mesgarani et al., 2014) and production (Bouchard et al., 2013) may allow understanding translation mechanisms and networks responsible for sensorimotor integration.

Imaging techniques (EEG and fMRI) and speech perception

Neuroimaging refers to the measuring of brain activity that can be directly or indirectly mapped onto a functional-anatomical model of the brain. In this thesis, two neuroimaging techniques were applied in the investigation of the neural correlates of speech perception, specifically fMRI (functional magnetic resonance imaging) and EEG (electroencephalography). Both of these techniques are non-invasive and safe for subjects, thus providing the opportunity to measure brain activity of healthy humans while they perform cognitive psychological experiments, such as involving the presentation and processing of spoken stimuli.

Functional MRI reflects metabolic demands underlying neuronal activity. Magnetic properties of the brain tissue in respect to levels of oxygen consumption (Ogawa et al., 1990) are recorded within small volumetric units (voxels). These recordings involve strong magnetic fields (in this thesis, at 3 teslas) capable to align atoms in the brain along a certain magnetic direction in combination with radio frequency pulses that temporarily disturb this alignment. The recovery behavior of the atoms to the original magnetic alignment can in turn be measured non-invasively using radio frequency reception coils. Additional magnetic fields varying the absolute magnetization spatially allow linking the energy released during the recovery phase of the atoms to their position in the brain using inverse spatial-frequency Fourier transformations (Huettel et al., 2004). Crucially for fMRI, oxygenation

properties of the brain tissue associated to neural activity have different recovery behaviors and provide a proxy to brain function manipulated by psychological experiments. Functional MRI allows imaging neural activity with high spatial resolution (in this thesis, 3 and 2 squared millimeter isotropic voxels). However, due to sluggish hemodynamic characteristics of the tissues, the temporal resolution of fMRI is limited to the range of several seconds.

During the past two decades, speech researchers have used fMRI to map cortical responses to speech input by contrasting activations from different types of speech stimuli and/or of tasks performed by listeners. Listening to speech evokes a broad extension of the superior temporal lobe and is suggested to proceed from a sound-based (acoustic) analysis in the HG to a speech-based (phonetic-phonological) analysis in the STG/STS and a concept-based (lexical-semantic) analysis in antero-ventral regions (Binder, 2000). Speech input further evokes fMRI response beyond the temporal lobes, such as in the inferior parietal lobe (IPL), inferior frontal gyrus (IFG) and motor regions (Raizada and Poldrock, 2007; Hickok et al., 2003; Pulvermuller et al., 2006; Zatorre et al., 1992). Areas extending ventrally from the superior temporal lobe have been proposed to be recruited in speech comprehension tasks, linking speech sounds to meaning and semantic networks. Areas extending dorsally, such as involving posterior temporal to inferior parietal and further onto pre-frontal and frontal motor areas have been proposed to be recruited in tasks related to motor action (Hickok and Poeppel, 2007).

EEG reflects electric neural activity measured externally at the surface of the head (scalp). The majority of the neural connections in the brain do not involve a physically link, but instead emission (axons) and reception (dendrites) parts of neurons are separated by a small gap (synaptic cleft). Neural communication depends on the transfer of chemical messengers (neurotransmitters) across the synaptic cleft. When neurotransmitters bind to the dendrites of a target neuron, an electric signal (post-synaptic potential) is generated and subsequently forwarded to central, executive, parts of the cell (soma). Crucially for EEG, due to chemical/electrical unbalances between the inner and outer cell, this post-synaptic potential originates an electric current outside the neuron (Buzsaki et

al., 2012). The perpendicular orientation of pyramidal neurons in the cortex in reference to the scalp together with the additive contribution from multiple, synchronized, currents from several neurons generate a sufficiently large electric field that can be measured on the external surface of the scalp using EEG. Because EEG recordings are directly related to electric neural activity, their temporal resolution is considered high (in this thesis, 4 milliseconds). This high temporal resolution permits analyzing the temporal dynamics of brain responses, including oscillatory properties that underlie synchronization and de-synchronization aspects of neural activity that may encompass short and long distance communication (Makeig et al., 2002). However, due to volume conduction, restrictions associated with undesirable interactions between distant electric fields and inverse calculation problems associated with the dimensionality reduction from EEG generation (three-dimensional - brain) to EEG acquisition (two-dimensional - scalp), its spatial resolution is low (in the range of centimeters). EEG signals recorded at several sites on the scalp are highly correlated and any interpretation about spatial distribution of underlying neural sources should be avoided or carried out with great care and knowledge of source localization techniques.

Due to the fast temporal resolution of EEG recordings, EEG is a crucial tool to investigate temporal dynamics of speech perception. The averaged EEG responses (ERP, event-related potentials) across different experimental conditions allowed speech researchers identifying certain time-intervals in which fundamental speech processes may occur (Kutas and Schmitt, 2003; Indefrey and Levelt, 2004). ERP components at early time intervals, 100-200 ms after word onset, have been associated with phonetic/phonological processing (Uusvuori et al., 2008; Bonte & Blomert, 2004; Dumay et al., 2001; Sanders et al., 2003). Intermediate time intervals (200-300 ms) have been suggested to reflect early aspects of lexical access (Bonte et al., 2006, Salmelin, 2007; Hagoort et al., 2004, Van den Brink et al., 2001), followed by lexical/semantic processing in the 300-600 ms window, as indicated by ERP modulations dependent of semantic attributes of words, semantic priming and semantic context (Kutas & Hillyard, 1980; Hagoort, 2008). Complementary to

ERP modulations, oscillatory effects over time can be estimated using frequency analysis (e.g., Fourier transformations). Synchronization of EEG signals at different frequencies over time as well as oscillatory phase changes are informative of neural mechanisms that may subserve speech processing (Hagoort, 2008; Makeig et al., 2002; Luo & Poeppel, 2007). Specifically, a synchronization of EEG responses at the theta oscillation (4-8 Hz) is suggested to modulate neural activity during the perception of speech and may be related to the intrinsic rate of acoustic change of speech signals (Peele and Davis, 2012; Giraud and Poeppel, 2012).

EEG/fMRI analysis methods: multivariate classification

Apart from activation-based analysis of fMRI responses to speech, recent approaches investigate fMRI representations that relate to our capacity to discriminate individual speech sounds from the same stimuli type, such as individual vowels, syllables, words or concepts. These approaches are more challenging, since fMRI responses that discriminate between, for example, two different vowels, are small and likely to be neglected statistically. Multivariate statistics allow uncovering such small effects by looking at patterns of multiple voxels simultaneously (Pereira et al., 2009; Formisano et al., 2008a).

Multivariate classification makes use of machine learning methods to learn associations between brain-states and underlying patterns of neural responses. Multivariate classification has unveiled new fields in brain sciences, extending from ambitious applications for brain-computer-interfaces (BCI, Sorger et al., 2012; Goebel et al., 2010) and brain-reading (Haynes et al., 2007), to applications of multivariate pattern analysis (MVPA) for fMRI and neuroimaging data that seek to answer advanced neuroscientific questions, as those explored in this thesis. Because multivariate classification exploits the interaction between multiple brain responses it encompasses higher sensitivity of data analysis in comparison to single response (univariate) methods (Pereira et al., 2009). This higher sensitivity of data analysis has enabled the investigation of neural processing and cognition with unprecedented success during the last decade (Haxby et al., 2001; Formisano et al., 2008b; Mitchell et

al., 2008). In MVPA, algorithms such as SVM (support vector machines, Cortes and Vapnik, 1995) are used to find discrimination between examples from different brain states associated to experimental conditions (classes). However, the increase of classification sensitivity is associated to an increase of model complexity, which directly introduces the risk of *overfitting* (learning single trial representations instead of their class-related commonalities of interest) (Lewis et al., 2013). To validate the success of multivariate classification and reduce the contamination of *overfitting*, the data is normally divided in a cross-validation fashion. Part of the data together the trial labels is used to train the classification model and the other part without the trial labels is used to test the generalization of the model.

Prevent *overfitting* in MVPA requires the reduction of the model complexity. Model complexity reduction can be accomplished by either increasing the number of observations (features) recorded (using longer or more experimental sessions) or reducing the number of variables (features) used in classification (using feature selection approaches, Lewis et al., 2013). A common feature selection approach used in fMRI-MVPA is the searchlight method (Kriegeskorte et al., 2006). The searchlight method creates multiple regions-of-interest based on a local selection of spatially proximal features. At each region-of-interest, a classifier is employed and classification performance indicates the presence of information content. Another method of feature selection for fMRI-MVPA is the RFE (recursive feature elimination, De Martino et al., 2008; Guyon et al., 2002). RFE exploits model parameters obtained in SVM training to select subsets of informative features. Furthermore, knock-out approaches (Carlson et al., 2003) assess the information within features by measuring their relative contribution to classification performance when these features are excluded from the analysis. Knock-out approaches are particularly relevant to test prior hypothesis.

Feature selection methods in EEG are important to investigate temporal patterns that relate to the evolution of information availability. An approach encompasses the selection of temporal intervals over time, or shifting time-windows (Chan et al., 2011; Hausfeld et al., 2012). Furthermore, highlighting

oscillatory properties of EEG signals is crucial to reveal fundamental neural mechanisms of speech processing. Given prior hypothesis related to the role of certain spectro-temporal aspects of neural oscillations during speech perception (Giraud and Poeppel 2012; Peelle and Davis, 2012), knock-out procedures may reveal suitable and were employed in chapter 2.

Cross-class generalization in multivariate classification

MVPA is often employed in subtraction logic to discriminate experimental classes. Beyond subtraction, MVPA permits identifying commonalities between neural representations shared across different classes. This approach is referred here as *cross-class generalization* and allows investigating whether a classifier trained to discriminate a set of classes is able to generalize this learning to another set of classes. It is a highly attractive approach since it finds commonalities – not differences – between experimental classes independently of their inherent variation. This strategy has permitted to disentangle the representation of spoken vowels independent of speaker (Formisano et al., 2008b), as well as, the representation of semantic-concepts from stimuli presentation modality (Shinkareva et al., 2011). In this thesis, cross-class generalization is employed to generalize the semantic meaning of spoken words across their bilingual equivalent words (chapter 2 and 3), and to generalize articulatory features of spoken syllables across acoustic variation of other articulatory dimensions (chapter 4). Overall, cross-class generalization is a type of cross-validation that statistically validates classification performance across multiple subsets of data originating from different experimental classes. Importantly however, cross-class generalization provides the possibility to find generalization of experimental classes across variation of certain properties reflected in natural divisions of the data and to find representation at a level of abstraction that is not possible with within-class generalization.

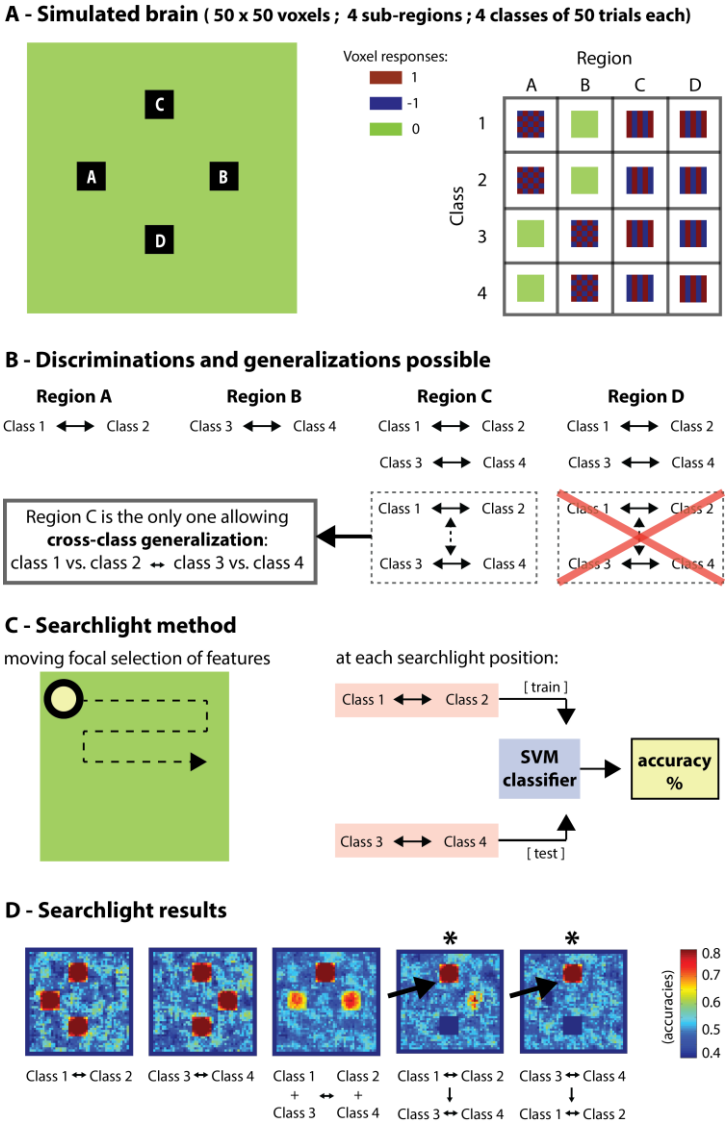
A possible strategy that accounts for the success of cross-class generalization in MPVA is to maximize the variability of the learning set, hence equipping the classification model for the ability to generalize unseen exemplars. In speech processing studies this variability is often introduced using multiple versions of

the same stimulus set that is independent of the dimension of interest, for example using different speech pronunciations when investigating speech meaning (Formisano et al., 2008b; Bonte et al., 2014). A minimum variability of two distinct stimulus versions may allow generalizing this learning to other stimulus versions across variation of other independent stimuli dimensions. However, in many experimental designs, such stimuli variability in the learning set is not attainable, for example when variability is limited to two versions, such as the variance of two languages in bilingual settings. In these cases, cross-class generalization cannot rely on variability within the training set and demands feature selection criteria capable to maximize the ratio between generalizable and non-generalizable features. The searchlight method selects features for classification within focal regions of interest, and may allow isolating generalizable response patterns for cross-class generalization. In EEG, the shifting time-windows approach (Hausfeld et al., 2012; Chan et al., 2011b; Simanova et al., 2010) may provide a similar feature reduction advantage in the temporal domain, and help investigate temporal dynamics of neural processes under cross-class generalization designs.

To illustrate the principles underlying cross-class generalization, we simulated fMRI responses containing discriminant information between classes and generalizable information across classes (Figure 1). Using a brain area of 50 by 50 voxels, we simulated four focal multivariate regions (region A, B, C and D) encoding the representation of four different experimental classes (Class 1, 2, 3 and 4). Informative voxels contained discriminative responses and noise (CNR = 2, contrast to noise ratio). Non-informative voxels did not contain information (CNR = 0). Noise followed a Gaussian distribution with mean = 0 and SD (standard deviation) = 1. Cross-class generalization is specifically simulated as the capacity of a multivariate classifier to generalize the discrimination of class 1 versus class 2 to the discrimination of class 3 versus class 4. In this example, cross-class generalization was simulated exclusively in region C, whereas region A and B encompassed the discrimination between class 1 and class 2, and between class 3 and class 4, respectively, but not their cross generalization. Region D was included to further challenge the methods

for cross-class generalization by including both discrimination patterns, but not allowing the generalization across each other. A focal feature selection method was employed to demonstrate the benefits of *searchlight* in isolating the commonalities between two classes from their inherent variation. We demonstrate how the *searchlight* method allows finding discrimination between class 1 and class 2, as well as, class 3 and class 4. Furthermore, *searchlight* was able to find generalization of these learning across each other in the target region C, but not in the region D. In addition, we also show how clustering classes introduces the risk of identifying the regions that are specific to discriminant differences, thus failing to isolate the generalization pattern of region C. Overall, this example aimed to provide a comprehensible description on the challenges of cross-class generalization and show that in cases where information is represented locally, the *searchlight* method is able to isolate commonalities between experimental conditions or brain-states using cross-class generalization strategies.

Figure 1. A) Simulation of cross-class generalization. Within a simulated brain of 50 by 50 voxels, we demonstrate the challenge to isolate a pattern of voxels containing a common information representation across the discrimination between two sets of classes. B) The discrimination between class 1 and class 2 is possible within three regions (i.e., region A, C and D). The discrimination between class 3 and class 4 is also possible within three regions (region B, C and D). However, only region C contains a representation that allows the generalization of the learning obtained from the discrimination between class 1 versus class 2 to the discrimination of class 3 versus class 4, and vice-versa. C) Description of the searchlight method. A focal selection of voxels moves through the entire brain. At each position, MVPA is performed and the performance of classification is mapped onto the center of the current searchlight position. D) Results of several classification procedures: discrimination of class 1 versus class 2; discrimination of class 3 versus class 4; discrimination of class 1 + class 3 versus class 2 + class 4; and finally, generalization across the learning of class 1 versus class 2 to class 3 versus class 4, and vice-versa. Black arrows point to the success of searchlight to isolate region C when performing cross-class generalization. The middle map shows the risk of clustering classes together when studying commonalities of representation.


A - Simulated brain (50 x 50 voxels ; 4 sub-regions ; 4 classes of 50 trials each)
C
A
B
D
Voxel responses:
1
-1
0
Region
A
B
C
D
Class
1
2
3
4
B - Discriminations and generalizations possible
Region A
Class 1 ↔ Class 2
Region B
Class 3 ↔ Class 4
Region C
Class 1 ↔ Class 2
Class 3 ↔ Class 4
Region D
Class 1 ↔ Class 2
Class 3 ↔ Class 4
Region C is the only one allowing
cross-class generalization:
class 1 vs. class 2 ↔ class 3 vs. class 4
C - Searchlight method
moving focal selection of features
at each searchlight position:
Class 1 ↔ Class 2
[train]
SVM classifier
accuracy %
Class 3 ↔ Class 4
[test]
D - Searchlight results
Class 1 ↔ Class 2
Class 3 ↔ Class 4
Class 1 + Class 3 ↔ Class 2 + Class 4
Class 1 ↔ Class 2
↓
Class 3 ↔ Class 4
Class 3 ↔ Class 4
↓
Class 1 ↔ Class 2
(accuracies)
0.8
0.7
0.6
0.5
0.4

In this thesis

Using the benefits of MVPA and cross-class generalization we investigated the representation of speech units, in particular words and syllables. This strategy allowed us tackle higher-order aspects of speech processing across variation of low-level acoustic properties. In **chapter 2**, we report research on spoken word comprehension in bilingual listeners using cross-class generalization to find language-independent representations of semantic-concepts. Comprehension of spoken words involves the processing and transformation of auditory signals to semantic concepts. It is a longstanding neuroscientific and neuropsychological debate whether semantic-conceptual knowledge emerges from the integration of distributed sensory specific representations and/or is constructed within higher level amodal regions (Patterson et al., 2007). We address this debate by combining fMRI, MVPA and an experimental design that exploits the unique capacities of bilingual listeners. Bilinguals, in fact, derive the same semantic concept from equivalent – but acoustically different - words in their first and second language, which provides a means to discover "purely" (i.e. input invariant) semantic representations in the brain. Specifically, using a searchlight approach (Kriegeskorte et al., 2006), we assessed the ability of brain-based multivariate classifiers trained with spoken words within one language to discriminate the translational equivalent words in the other language.

In **chapter 3**, we adapted the bilingual paradigm used in chapter 2 to EEG, and studied the temporal and temporal-oscillatory dynamics of spoken words and semantic-concepts representation. Using a shifting time-window feature selection approach, we aimed to identify EEG response patterns following spoken word presentation discriminative of words within the same languages and of semantic-concepts emerging from their across-language generalization. Furthermore, a moving filter band-out procedure (i.e., similar to knock-out feature selection, Carlson et al., 2003) in combination with the shifting time-windows assessed the contribution of specific temporal-oscillatory bins in the representation of words and semantic-concepts.

In **chapter 4**, the benefits of MVPA and cross-class generalization were employed to study the dorsal stream of speech processing involved in sensorimotor integration (Hickok and Poeppel, 2007). Sensorimotor integration links the neural systems involved in perceiving and producing speech and to is proposed to include the transformation between acoustic and articulatory features of speech. This link is suggested to play an important role in acquiring and monitoring speech production. However, its role in speech perception has been the object of debated over the past decades (Pulvermuller and Fadiga, 2010; Hickok et al., 2011; Galantucci et al., 2006). Methodological constraints related to experimental design and analysis methods have so far prevented the disentanglement of neural responses to acoustic versus articulatory features of speech. Here, using fMRI responses to individual spoken syllables, we investigated the articulatory representation of speech across acoustic variation of other articulatory dimensions.

In **chapter 5**, we investigate and validate a novel approach for functional connectivity in neuroimaging that exploits the benefits of MVPA in uncovering information patterns within cortical regions-of-interest (Kriegeskorte et al., 2006). The approach is entitled IBC (information-based connectivity) and assesses the consistency of single-trial MVPA predictions between regions. Brain areas whose decoding predictions co-vary are suggested to be modulated by common physiological states related to the availability of information representation across single-trials. Using fMRI simulations, we validated the principles underlying IBC and using the fMRI empirical results reported in chapter 4, we apply IBC to unravel unique information-based connections not attained otherwise using an activity-based fMRI time-series connectivity approach.

References

Baker E, Blumstein SE, Goodglass H (1981) Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia, 19*, 1–15.

Barlow HB (1995) The neuron in perception. Gazzaniga MS, editor. The cognitive neurosciences. Cambridge (MA): MIT Press. 415–34.

Binder JR, Frost Ja, Hammeke Ta, Bellgowan PS, Springer Ja, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex (New York, N.Y. : 1991)*, *10*(5), 512–28.

Boatman D (2004) Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92, 47–65.

Bonte M (2004) Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology*, *115*(2), 409–423.

Bonte ML, Blomert L (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Brain Research. Cognitive Brain Research*, *21*(3), 360–76.

Bonte M, Parviainen T, Hytönen K, Salmelin R (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(1), 115–23.

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(13), 4548–57.

Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–32.

Broca P (1865) Sur le sie`ge de la faculte´ du langage articule´. Bulletins de la Socie´te´ d'Anthropologie, 6, 337–393.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G (2011). Conduction aphasia, sensory-motor integration, and

phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–28.
Buzsáki G, Anastassiou Ca, Koch C (2012) The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nature Reviews. Neuroscience*, *13*(6), 407–20.
Caramazza A, Mahon BZ (2006) The organisation of conceptual knowledge in the brain: the future's past and some future directions. *Cognitive Neuropsychology, 23*, 13–38.

Carlson TA, Schrater P, He S (2003) Patterns of Activity in the Categorical Representation of Objects. *Journal of Cogntive Neoroscience, 15 (5)*, 704–717.

Chan AM, Baker JM, Eskandar E, Schomer D, Ulbert I, Marinkovic K, Cash SS (2011a) First-pass selectivity for semantic categories in human anteroventral temporal lobe. *The Journal of neuroscience*, *31*, 18119–29.

Chan AM, Halgren E, Marinkovic K, Cash SS (2011b). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*(4), 3028–39.
Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage*, *56*(2), 582–92.
Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning 20*:273–297.

Cutler A (2012) "*Native Listening: Language Experience and the Recognition of Spoken Words*", MIT Press 9780262017565.

Damasio AR (1989) Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition.* 33, 25-62.

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio A (1996) A neural basis for lexical retrieval. *Nature*, 380, 499-505.

Damasio H, Damasio AR (1980) The anatomical basis of conduction aphasia. Brain, 103, 337–350.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58.

Dronkers NF, Plaisant O, Iba-Zizen MT, Cabanis Ea (2007) Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain: A Journal of Neurology*, *130*(5), 1432–41.

Dumay N, Benraïss A, Barriol B, Colin C, Radeau M, Besson M (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, *13*(1), 121–43.

Fodor JA (1983) Modularity of Mind: An Essay on Faculty Psychology. *Cambridge, Mass.: MIT Press.*

Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859–69.

Formisano E, De Martino F, Valente G (2008a) Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging, 26*, 921–934.

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

Galantucci B, Fowler Ca, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–77.

Gall FJ (1835) "*On the Functions of the Brain and of Each of Its parts: With Observations on the Possibility of Determining the Instincts, Propensities, and Talents, Or the Moral and Intellectual Dispositions of Men and Animals, by the Configuration of the Brain and Head, Volume 1*." Marsh, Capen & Lyon.

Geschwind N (1965) Disconnexion syndromes in animals and man. Brain, 88, 237–294, 585–644.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–7.

Goebel R, Zilverstand A, Sorger B (2010) Real-time fMRI-based brain–computer interfacing for neurofeedback therapy and compensation of lost motor functions. *Imaging in Medicine, 2*, 407–415.

Guenther FH, Vladusich T (2012) A Neural Theory of Speech Acquisition and Production. *Journal of Neurolinguistics*, *25*(5), 408–422.

Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Machine Learning, 46*, 389–422.

Hagoort P (2008) The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1055–69.

Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004) Integration of word meaning and world knowledge in language comprehension. *Science (New York, N.Y.)*, *304*(5669), 438–41.

Hausfeld L, De Martino F, Bonte M, Formisano E (2012) Pattern analysis of EEG responses to speech and voice: influence of feature grouping. *NeuroImage*, *59*(4), 3641–51.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*, 2425–2430.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham D (2007) Reading hidden intentions in the human brain. *Current Biology, 17*, 323-328.

Hebb DO (1949) Organization of behavior. *New York: Wiley*, 335

Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–82.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67–99.

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

Hickok G (2010) The role of mirror neurons in speech and language processing. *Brain and Language*, *112*(1), 1–2.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Huettel, Scott A & Song, Allen W., 1971- & McCarthy, Gregory, 1952- (2004). *Functional magnetic resonance imaging* (1st ed). Sinauer Associates ; Basingstoke : Palgrave, Sunderland, Mass

Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–44.

Jones OP, Prejawa S, Hope TMH, Oberhuber M, Seghier ML, Leff AP, Price CJ (2014) Sensory-to-motor integration during auditory repetition: a combined fMRI and lesion study. *Frontiers in Human Neuroscience*, *8*, 24.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8.

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science, New Series, Vol. 207, 203-205.

Kutas M and Schmitt BM, Language in microvolts, In: Banich, M.T. and Mack, M. (Ed). *Mind, brain, and language: Multidisciplinary perspectives*, Lawrence Erlbaum Associates: NJ, 2003, pp. 171-209.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology, 94*, 1904–1911.

Lauter JL, Herscovitch P, Formby C, Raichle ME (1985) Tonotopic organization in human auditory cortex revealed by positron emission tomography. *Hearing Research*, *20*(3), 199–205.

Lewis MP, Gary FS, Charles DF (2014) *Ethnologue: Languages of the World, Seventeenth edition. Dallas, Texas: SIL International.* Online version: http://www.ethnologue.com.

Lewis-Peacock JA, Norman KA (2013) Multi-voxel pattern analysis of fMRI data. In M.S. Gazzaniga & G.R. Mangun (Eds.), *The Cognitive Neurosciences, 4th ed. Cambridge, Massachusetts: MIT Press*.

Liberman A M, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of speech code. *Psychological Review 74,* 431–461.

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

Luo H, Poeppel D (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–10.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. *Science (New York, N.Y.)*, *295*(5555), 690–4.

Marie P (1897) L'évolution du langage considéré au point de vue de l'étude de l'Aphasie.

Martin A (2007) The representation of object concepts in the brain. *Annual Reviews in Psychology, 58*, 25– 45.

McGlone J (1984) Speech comprehension after unilateral injection of sodium amytal. *Brain Lang.* 22, 150–157.

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science (New York, N.Y.)*, *343*(6174), 1006–10.

Meyer K, Damasio A (2009) Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neuroscience, 32*, 376 –382.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America, 87*, 9868–9872.

Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews. Neuroscience*, *8*, 976–87.

Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*(9), 320.
Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage, 45*, 199–209.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–70.
Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*(5), 351–60.
Raizada RDS, Poldrack Ra (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726–40.
Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–24.

Salmelin R (2007) Clinical neurophysiology of language: the MEG approach. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 118(2), 237–54.
Sanders LD, & Neville HJ (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Brain Research. Cognitive Brain Research*, 15(3), 228–40.

Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal Acoustic Society United States of America, 2 (115)*, 813-21.

Scott SK, Wise RJ (2004) The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. *Neuroimage, 54*, 2418 –2425.

Simanova I, van Gerven M, Oostenveld R, Hagoort P (2010) Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLoS ONE, 5*, e14465.

Uusvuori J, Parviainen T, Inkinen M, Salmelin R (2008) Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(2), 456–66.

Van den Brink D, Brown CM, Hagoort P (2001) Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–85.

Visser M, Jefferies E, Embleton KV, Lambon-Ralph MA (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience, 24*, 1766 –78.

Wernicke C (1874, 1977) Der aphasische symptomen complex: eine psychologische studie auf anatomischer basis.

Eggert GH (Ed.), Wernicke's works on aphasia: a sourcebook and review, 91–145. The Hague: Mouton.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*, 846-849.

CHAPTER 2

Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe

Corresponding publication

Correia JM, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M (2014)

Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe.

The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, *34*(1), 332–8.

Abstract

Bilinguals derive the same semantic concepts from equivalent – but acoustically different - words in their first and second language. The neural mechanisms underlying the representation of language independent concepts in the brain remain unclear. Here, we measured functional MRI in human bilingual listeners and reveal that response patterns to individual spoken nouns in one language (e.g. "horse" in English) accurately predict the response patterns to equivalent nouns in the other language (e.g. "paard" in Dutch). Stimuli were four monosyllabic words in both languages, all from the category of 'Animal' nouns. For each word, pronunciations from three different speakers were included, allowing the investigation of speaker independent representations of individual words. We employed multivariate classifiers and a searchlight method to map the informative fMRI response patterns that enable decoding spoken words within languages (within language discrimination) and across languages (across language generalization). Response patterns discriminative of spoken words *within-language* were distributed in multiple cortical regions, reflecting the complexity of the neural networks recruited during speech and language processing. Response patterns discriminative of spoken words *across-language* were limited to localized clusters in the left anterior temporal lobe, the left angular gyrus and the posterior bank of the left post-central gyrus, the right posterior STS/STG, the right medial ATL, the right anterior insula, and bilateral occipital cortex. These results corroborate the existence of 'hub' regions organizing semantic-conceptual knowledge in abstract form at the fine-grained level of within semantic category discriminations.

Introduction

In multilingual environments, verbal communication relies on extraction of unified semantic concepts from speech signals in native as well as non-native languages. How and where our brain performs these language-invariant conceptual transformations remains largely unknown. Spoken word comprehension proceeds from a sound-based (acoustic-phonological) analysis in posterior superior temporal areas towards meaning-based processing in a brain's semantic-conceptual network (Hickok & Poeppel, 2007; Binder et al., 2000). This network - including anterior and middle temporal cortex, posterior parietal cortex and inferior frontal gyrus - is involved in the neural representation of semantic memories, and connects to modality specific brain regions subserving perception and action (Martin, 2007; Meyer & Damasio, 2009; Meyer et al., 2010). Evidence from brain lesions in patients with selective deficits of semantic knowledge (semantic dementia) further suggests a crucial role of the anterior temporal lobe (ATL) as a 'semantic hub', where abstract, amodal representations of words and concepts are constructed (Damasio et al., 1996; Patterson et al., 2007). In neuroimaging studies it has been more difficult to find reliable ATL activation due to susceptibility artifacts or limited field-of-view (Visser and Lambon-Ralph., 2011). However, recent studies using distortion corrected fMRI (Binney et al., 2010; Visser et al., 2012), fMRI decoding (Kriegeskorte et al., 2007), MEG and EEG decoding (Chan et al., 2010; Marinkovic et al, 2003) or intracranial EEG (Chan et al., 2011) support the important role of the ATL in semantic processing.

Here, we investigate the semantic representation of spoken words at the fine-grained level of within-category distinction (animal nouns) and across-language generalization. We do so by combining fMRI, state-of-the-art multivariate pattern analysis and an experimental design that exploits the unique capacities of bilingual listeners. In separate Dutch and English blocks, we asked bilingual participants to listen to individual animal nouns and to detect occasional non-animal target nouns. Speech comprehension in skilled bilinguals provides a means to discover neural representations of concepts while keeping the input

modality constant. Furthermore, our focus on within-category distinctions avoids global effects resulting from across-category differences (Caramazza & Mahon, 2006). Following supervised machine learning approaches, we train multivoxel classifiers (linear-SVM, Cortes & Vapnik, 1995) to predict the identity of the animal noun a participant was listening to from new (untrained) samples of brain activity. In a first analysis we aim to identify the network of brain regions involved in within-language word discrimination. To this end we train classifiers to discriminate brain responses to English (e.g. 'horse' vs. 'duck') and Dutch (e.g. 'paard' vs. 'eend') nouns. In a second analysis we aim to isolate brain regions involved in language-independent decoding of the animal nouns (See figure 1A). Here we train classifiers to discriminate brain responses to words in one language (e.g. 'horse' vs. 'duck') and test whether this training generalizes and allows discrimination of brain responses to the corresponding nouns in the other language (e.g. the Dutch equivalents, 'paard' vs. 'eend'). Importantly, each word was spoken by three female speakers allowing for speaker-invariant word discrimination. Moreover, all words are acoustically-phonetically distinct both within and across languages.

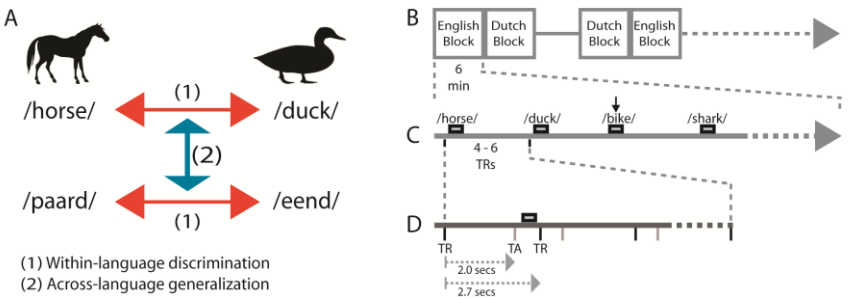


Figure 1. Experimental design. (A) Word discrimination (1) was performed for all pair-wise comparisons in English and in Dutch. Across-language generalization (2) was performed across equivalent words in the other language. Both generalization directions were performed, from English to Dutch and from Dutch to English. (B) Runs/blocks organization along the fMRI session. Only 2 runs are depicted for illustration (from the total of 4 runs). (C) Within each block, an interval of 4 to 6 scanning repetition times (TRs) separates the presentation of the nouns. The black vertical arrow represents a response from the participant to detect a non-animal word (e.g. bike). (D) In order to assure noise-free perception of the words, they were presented in a silent gap of 700 ms between the TR and the acquisition of the whole-brain volume (TA).

Methods

Participants

Ten native Dutch (L1) participants proficient in English (L2) took part in the study (3 males, 1 left-handed, mean age = 25.4 SD= 2.84). All participants were undergraduate or postgraduate students of Maastricht University studying and/or working in an English speaking environment. All participants reported normal hearing abilities and were neurologically healthy. English proficiency was assessed with the LexTALE test, a vocabulary test including 40 frequent English words and 20 non-words (Lemhöfer & Broersma, 2011). The mean test score was 85.56% correct (SD= 12.27%). This score is well above the average score (70.7%) of a large group of Dutch and Korean advanced learners of English performing the same test (Lemhöfer & Broersma, 2011). For comparison reasons, participants also conducted the Dutch version of the vocabulary test. The mean Dutch proficiency score was 91.11% (SD= 6.57). The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht, The Netherlands.

Stimuli

Stimuli consisted of Dutch and English spoken words representing four different animals (English: 'Bull', 'Duck', 'Horse' and 'Shark', and the Dutch equivalents: 'Stier', 'Eend', 'Paard' and 'Haai') and three inanimate object words (English: 'Bike', 'Dress' and 'Suit' and the Dutch equivalents 'Fiets', 'Jurk' and 'Pak') (Figure 1A). All stimuli were monosyllabic animal nouns acoustically/phonetically distinct from each other both within and across languages. Phonetic distance between word pairs was quantified using the Levenshtein distance, which gives the number of phoneme insertions, deletions and/or substitutions required to change one word into the other, divided by the number of phonemes of the longest word (Levenshtein, 1966). On a scale from 0 (no changes) to 1 (maximum number of changes), the mean (SD) Levenshtein distances corresponded to 0.83 (0.15) for Dutch word pairs, 0.93

(0.12) for English word pairs and 1.00 (0.00) for English-Dutch word pairs. Furthermore, all animal nouns had an early age of acquisition in Dutch (mean = 5.28 years SD = 0.98; De Moor et al. 2000) and a medium-high frequency of use expressed on a logarithmic scale in counts per million tokens in Dutch (mean = 1.29 SD = 0.71) and in English (mean = 1.50 SD = 0.42; Celex database Baayen et al., 1995). To add acoustic variability and allow for speaker-invariant MVPA analysis, the words were spoken by three female native Dutch speakers with good English pronunciation. Stimuli were recorded in a sound proof chamber at a sampling rate of 44.1 kHz (16 bit resolution). Post-processing of the recorded stimuli was performed in PRAAT software (Boersma & Weenink, 2011) and included band-pass filtering (80-10500 Hz), manual removal of acoustic transients (clicks), length equalization, removal of sharp onsets and offsets using 30 ms ramp envelopes, and amplitude equalization (average RMS). Stimulus length was equated to 600 ms (original range: 560–640 ms) using PSOLA (75–400 Hz as extrema of the F0 contour). We carefully checked our stimuli for possible alterations in F0 after length equation and did not find any detectable changes. We assured that the produced stimuli were unambiguously comprehended by the participants during the stimuli familiarization phase prior to the experiment.

Experimental procedures

The experimental session was organized in 4 runs, each run containing 2 blocks (one Dutch and one English). Each block included 27 nouns: 24 animal nouns and 3 (11.1%) non-animal nouns. Run 1 and 3 started with an English block followed by a Dutch block; Run 2 and 4 started with a Dutch block followed by an English block (see Figure 1B). Participants were instructed to actively listen to the stimuli and to press a button (with the left index finger) whenever they heard a non-animal word. The goal of the task was to help maintaining a constant attention level throughout the experiment and promote speech comprehension at every word presentation. All participants paid attention to the words as indicated by a mean (SD) detection accuracy of 97.5 (5.68) %. Non-animal trials were excluded from further analysis. The 24 relevant animal nouns in each block corresponded to 6 repetitions of each of the 4 animal

nouns. Because nouns were pronounced by 3 different speakers, each physical stimulus was repeated twice in each block. Stimulus presentation was pseudo-randomized within each block, avoiding consecutive presentations of the same words or sequences of words. Throughout the experiment, each animal noun was presented 24 times per language.

Stimuli were presented binaurally at a comfortable intensity level using MR compatible in-ear headphones. Stimulus presentation occurred in a silent interval of 700 ms between two volume acquisitions (Figure 1C-D). According to a slow event-related design, the average inter-trial-interval between two stimuli was 13.5 seconds (jittered between 10.8 sec and 16.2 sec, corresponding to 4, 5 or 6 fMRI scanning repetitions). Each block had 6 minutes duration; each run took 12 minutes, resulting in a full functional scanning time of 48 minutes. A gray fixation cross against a black background was used to keep the visual stimulation constant during the entire duration of a block. Block and run transitions were marked with written instructions.

Functional MRI acquisition

Functional and anatomical image acquisition was performed on a Siemens Allegra 3 Tesla scanner (head setup) at the Maastricht Brain Imaging Center. Functional runs were collected per subject with a spatial resolution of 3 mm isotropic using a standard echoplanar-imaging (EPI) sequence (repetition time [TR] = 2.7 s; acquisition time [TA] = 2.0 s; field of view [FOV] = 192 mm x 192 mm; matrix size = 64 x 64; echo time [TE] = 30 ms). Each volume consisted of 33 slices (10% voxel size gap between slices) that covered most of the brain of the participants (in 5 subjects the superior posterior parietal tip of the cortex had to be left out). The difference between the acquisition time and the repetition time introduced a silent gap used for the presentation of the auditory stimuli. High resolution (voxel size, $1 \times 1 \times 1$ mm^3) anatomical images covering the whole brain were acquired after the second functional run using a T1-weighted three-dimensional ADNI sequence (TR, 2050 ms; TE, 2.6 ms; 192 sagittal slices).

Functional MRI data pre-processing

FMRI data were pre-processed and analyzed using Brain Voyager QX Version 2.4 (Brain Innovation B.V. Maastricht, The Netherlands) and custom-made Matlab routines. Functional data were 3D motion corrected (trilinear interpolation), corrected for slice scan time differences, and were temporally filtered by removing frequency components of five or less cycles per time course (Goebel et al., 2006). Anatomical data were corrected for intensity inhomogeneity and transformed into Talairach space (Talairach & Tournoux, 1988). Functional data were then aligned with the anatomical data and transformed into the same space, to create 4D volume time-courses (VTCs). Individual cortical surfaces were reconstructed from gray–white matter segmentations of the anatomical data and aligned using a moving target-group average approach based on curvature information (cortex-based alignment) to obtain an anatomically aligned group-averaged 3D surface representation (Goebel et al., 2006; Frost & Goebel, 2012).

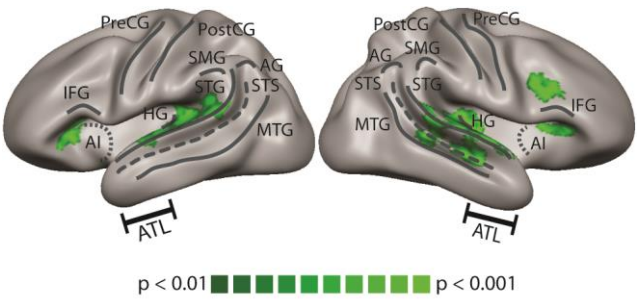


Figure 2. Overall-activation to the spoken words in the left and right hemispheres. Surface-based functional contrast maps of all words versus baseline (t-statistics). Statistical values were threshold for p<0.01 and subsequently corrected for multiple comparisons using a cluster size threshold. Significant activation clusters are depicted in green colors (minimum cluster sizes are 135 and 115 sq-mm for the left and right hemisphere, respectively). Landmark labels: AG (angular gyrus); AI (anterior insula); ATL (anterior temporal lobe); HG (Heschl's gyrus); IFG (inferior frontal gyrus); MTG (middle temporal gyrus); PostCG (posterior central gyrus); PreCG (precentral gyrus); SMG (supramarginal gyrus); STS (superior temporal sulcus).

FMRI data analysis: univariate statistics

Univariate effects were analyzed using a random-effects general linear model (GLM). A single predictor per stimulus condition was convoluted with a double gamma hemodynamic response function (HRF). In order to identify cortical regions generally involved in the processing of spoken words, we constructed functional contrast maps (t-statistics) by (1) comparing activation to all animal words versus baseline across subjects, (2) combining all possible binary contrasts within nouns of the same language, and (3) grouping all equivalent nouns into single concepts and contrasting all possible binary combinations of concepts. All GLM analyses were performed at the cortical surface after macro-anatomical variability had been minimized by applying cortex based alignment (Goebel et al., 2006; Frost & Goebel, 2012). Each univariate statistical map was corrected for multiple comparisons by applying a surface-based cluster size threshold with a false-positive rate (alpha) of 1% (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006). The cluster-size thresholds were obtained after setting an initial vertex-level threshold ($p<0.01$, uncorrected) and submitting the maps to a whole-brain correction criterion based on the estimate of the map's spatial smoothness (Goebel et al., 2006).

FMRI data analysis: multivariate pattern analysis (MVPA)

To investigate speech content information in the fMRI responses we used a supervised machine learning algorithm (linear support vector machine, linear-SVM; Saitta & Vapnik, 1995) combined with single-trial multivoxel classification (e.g. Cox & Savoy, 2003; Formisano et al., 2008). Classifications were performed to evaluate whether patterns of voxels conveyed information on the identity of spoken words (within-language word discrimination) and their language invariant representations (across-language word generalization). Within-language word discrimination entailed training a classifier to discriminate between two words of the same language (e.g. horse vs. duck) and testing in the same words spoken by a speaker not included in the training phase. Across-language word generalization was performed by training a classifier to discriminate between two words within one language (e.g. horse vs. duck) and testing in the translational equivalent words within the other language (e.g.

paard vs. eend), thus relying on language independent information of spoken words representing equivalent concepts in Dutch and English (figure 1A). Additional methodological steps encompassing the construction of the fMRI feature space (feature extraction and feature selection) and the computational strategy to validate (cross-validation) and display (discriminant maps) the classification results are described below in detail.

Classification procedure

Feature extraction. Initially, the pre-processed fMRI time series was split into single trials, time locked to the presentation of each stimulus. Then, for each trial, the time series of the voxels were normalized using the percentage of BOLD (Blood Oxygenation Level Dependency) signal change in reference to a baseline extracted from the averaged interval between 2.0 seconds preceding stimulus onset and 0.7 seconds following stimulus onset (2 TRs). Finally, one feature was calculated per voxel from the averaged interval between 3.4 and 6.1 seconds succeeding stimulus onset (2 TRs). Considering the timing of stimulus presentation, this ensured a strict separation between the feature estimates for two subsequent trials.

Feature selection. In order to avoid degraded performances of the classification algorithm due to the high dimensionality of the feature space (model *overfitting*, for a description see Norman et al., 2006), a reduction of the number of features (voxels) is usually performed. In the present work, feature selection was accomplished using the 'searchlight method' (Kriegeskorte et al., 2006) that restricts features to a spherical selection of voxels repeated for all possible locations of the gray matter. Here searchlight used a linear-SVM classifier and a sphere radius of 2.5 voxels.

Cross-validation. Cross-validation for both types of analysis exploited the divisions of stimuli naturally resulting from our design. Cross-validation for word discrimination was performed by assessing the accuracy of *across-speaker generalizations*. At each cross-validation fold, the stimulus examples recorded from one speaker were left out for testing, while the examples recorded from the other two speakers were used for training. Hence, during word discrimination, classifiers were assessed to discriminate the words participants

listened to independently of the speaker pronouncing the words. Cross-validation for language invariant semantic representations was performed by assessing the accuracy of *across-language generalizations*. This resulted in two folds: training in Dutch words and testing in their English translational equivalents; and training in English words and testing in their Dutch translational equivalents. Cross-validation approaches based on the generalization of other stimuli dimensions have been successfully used before (e.g., Formisano et al., 2008; Buchweitz et al., 2011) and enable detecting activation patterns that convey information on a stimulus dimension, such as word identity or semantic information, regardless of variations in other stimulus dimensions, such as speaker or language input. Hence, the within-language word discrimination and the across-language generalization analyses relied on a different number of training and testing trials (within-language: 18 trials for training and 6 trials for testing per class; across-language: 24 trials for training and 24 trials for testing per class).

Discriminative maps

At the end of the searchlight computation, accuracy maps of within-language word discrimination and across-language word generalization were constructed. Accuracy maps were averaged within each subject across binary comparisons and cross-validation folds. Thereafter, individual averaged accuracy maps were projected onto the grouped averaged cortical surface and anatomically aligned using cortex-based alignment (Goebel et al., 2006). To assess the statistical significance of the individual accuracy maps (chance level is 50%) across subjects a two-tailed non-parametric Wilcoxon test was employed. In the within-language discrimination analysis, one map was produced per language and subsequently combined into a single map by means of conjunction analysis (Nichols et al., 2005). The within-language word discrimination map, thus depicts regions with consistent sensitivity in English and Dutch. For the across-language word generalization, one map was produced from all possible binary language generalizations. The resulting statistical maps were corrected for multiple comparisons by applying a cluster-size threshold with a false-positive rate (alpha) of 5% (within language discrimination) or 1 % (across-language

generalization) after setting an initial vertex-level threshold (within language: $p<0.05$, uncorrected; language generalization: $p<0.01$, uncorrected) and submitting the maps to a whole-brain correction criterion based on the estimate of the map's spatial smoothness (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006).

Results

Univariate Analysis

Figure 2 (and Table 1) illustrates the fMRI responses elicited by all animal words across subjects, as computed by a functional contrast map (t-statistics) comparing all animal words versus baseline. The words elicited extensive bilateral activation in the superior temporal cortices as well as in the right inferior frontal lobe and the bilateral anterior insulas. To assess univariate differences between the responses evoked by individual words, we conducted all possible word-to-word contrasts within the same language (e.g. horse versus duck), as well as all possible concept-to-concept contrasts (e.g. horse + paard versus duck + eend). None of the possible contrasts yielded significant differences within or across participants.

Multivariate Analysis

Figure 3 depicts the statistical maps of searchlight selections for which the word discrimination and the language generalization analyses yielded accuracies significantly above chance level (50%). Significance was tested with a Wilcoxon test across participants and corrected for multiple comparisons using a surface-based cluster size threshold.

Word discrimination identified a broad set of cortical areas in both hemispheres (Figure 3A and Table 2). Discriminative voxels surviving cluster-size multiple comparisons correction clustered in the dorsal STG, posterior STG/supramarginal gyrus (SMG), angular gyrus (AG), and middle and anterior temporal lobe (ATL) in the right hemisphere, as well as bilaterally in the inferior

and superior frontal cortex, superior parietal cortex, occipital cortex and the anterior insulas.

Figure 3B (and Table 3) shows the statistical maps with consistent across-language generalization performances. Here, a region within the left ATL, in the anterior portion of the superior temporal sulcus (STS) was highlighted together with other clusters in the left and right hemispheres. Specifically, the left inferior parietal lobe/AG, the posterior bank of the left post-central gyrus, the right posterior STS/STG, the anterior portion of the right insula, the medial part of right ATL and areas within the occipital cortices bilaterally seem sensitive to generalization. In the across-language generalization analysis, classifications relied on language invariant properties of the words across subjects. There was no reportable difference for the directionality of the generalization. Both, Dutch to English and English to Dutch yielded similar results.

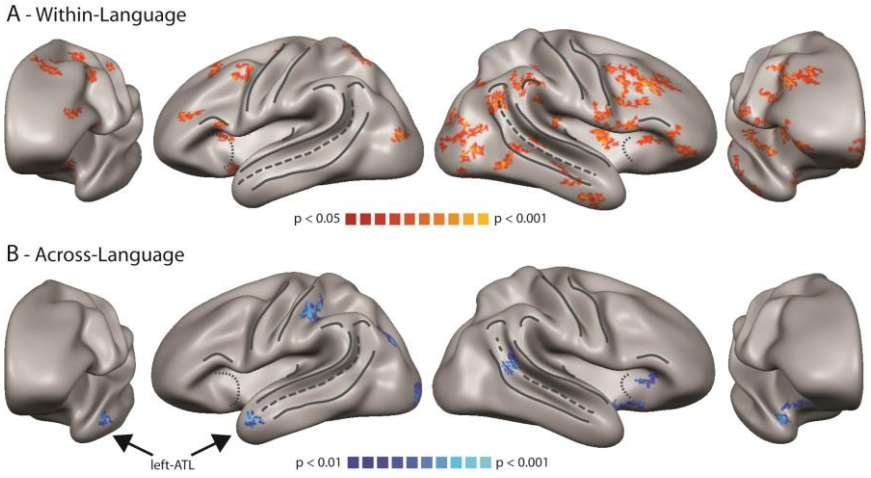


Figure 3. Statistical maps indicating the cortical areas that produced classification accuracies above chance level (50%) using a two-tailed Wilcoxon test and corrected for multiple comparisons using a cluster size threshold. (A) Significant word discrimination clusters are depicted in orange colors (minimum cluster sizes are 16 and 17 sq-mm for the left and right hemisphere, respectively). (B) Significant language generalization clusters are depicted in blue colors (minimum cluster sizes are 31 and 36 sq-mm for the

left and right hemisphere, respectively). The black arrows point the cluster found within the left anterior temporal lobe (ATL).

Discussion

By combining fMRI multivariate pattern analysis and an experimental design that exploits the unique capacities of bilingual listeners we identified a distributed brain network enabling within-category discrimination of individual spoken words. Most crucially, we were able to isolate focal regions including the left anterior temporal lobe that play a relevant role in the generalization of the meaning of equivalent words across languages. These findings provide direct evidence that in bilingual listeners, semantic-conceptual knowledge is organized in language independent form in focal regions of the cortex.

The univariate analysis showed distributed brain activation associated with the processing of spoken words (Binder et al., 2000; Price, 2010; Pulvermuller et al., 2009), but did not allow the discrimination of individual words or of language-independent concepts. In a first multivariate analysis, we identified the network of brain regions informative of within-language word discrimination. Each area identified in the word discrimination map (Figure 3A, Table 2) was able to consistently predict single words in both languages based on a local interaction of neighboring voxels. Since the discrimination was generalized across speakers, the maps reflect neural representations of the words that are robust to variations of low-level auditory features and speaker related phonetic-phonological features. Within-language word discrimination relied on acoustic-phonetic and semantic-conceptual differences between the nouns, as well as possible other differences reflecting their individual properties. Accordingly, our results yielded a distributed cortical network associated with multiple aspects of speech processing (Hickok & Poeppel, 2007), including regions associated with spectro-temporal analysis in the dorsal STG, a sensorimotor interface at the posterior STG/inferior parietal lobe, and an articulatory network involving (pre-)motor areas and the inferior frontal gyrus. Moreover, the involvement of

sensori-motor areas during speech perception is a highly debated topic (Hickok et al., 2011; Galantucci et al., 2006; Pulvermuller et al., 2006), for which multivoxel speech decoding may provide important contributions, as to describe the neural code holding phonological and articulatory features of speech.

The second multivariate analysis - across-language generalization - relied uniquely on language invariant semantic-conceptual properties of the nouns. This analysis revealed neural representations of the animal words independently of the language in which they were presented in the left ATL (STS), but also in other clusters, namely the left angular gyrus, the posterior bank of the left post-central gyrus, the right posterior STS/STG, the right anterior insula, the medial part of right ATL and areas within visual cortex (Figure 3B, Table 3). This finding is consistent with the observation that the failure to recognize words in bilingual aphasics typically occurs simultaneously in first and second language (e.g. in Alzheimer's disease, Hernàndez et al. 2007; primary progressive aphasia, Hernàndez et al., 2008; or anomic aphasia, Kambanaros & van Steenbrugge, 2006). Furthermore, our observation of language-independent representations of spoken words in the left ATL converges with previous findings showing fMR adaptation effects in this region for across-language semantic priming of visual words in bilinguals (Crinion et al., 2006). It corroborates neuropsychological evidence of semantic dementia (Damasio et al., 1996) and indicates a role of the left ATL as a central-hub in abstract semantic processing (Patterson et al., 2007). However, the location of our left ATL cluster appears more dorsal, comprising the anterior STS, in comparison to more ventral regions in the anterior MTG/ITS commonly reported in studies using visual stimulation (Binder et al., 2009). Such a dorsal shift for semantic-conceptual processing of auditory words within the ATL has not been attained from lesion studies given the commonly large coverage of the lesions, but has been reported in recent fMRI studies (Visser & Lambon Ralph, 2011; Scoot et al., 2000).

We also observed significant language-independent decoding in other clusters of the left and right hemisphere. Namely, the left angular gyrus and the

posterior bank of the left post-central gyrus, the right posterior STS/STG, the right anterior insula, medial portions of the right ATL, and regions of the occipital cortices bilaterally. These regions have been associated with cross-modal (Shinkareva et al., 2010; Visser et al., 2012) and bilingual (Buchweitz et al., 2011) processing of nouns and with the interface between semantic representations and articulatory networks during speech processing (Linda, 2009; Hickok & Poeppel, 2007). Furthermore, the left angular gyrus has been associated with both semantic representation (Binder et al., 2009) and semantic control (Visser & Lambon-Ralph, 2011). Our results suggest a similar role of these regions when bilinguals process within category semantic distinctions of spoken words. At a more general level, our results indicate the involvement of the right hemisphere in speech processing and lexical-semantic access, consistent with previous findings in patients with left hemispheric damage (Stefanatos et al., 2005), subjects undergoing selective hemispheric anesthesia (wada procedure, McGlone et al., 1984) and brain imaging studies (Binder et al., 2010; Hickok and Poeppel, 2007).

Conclusions

Brain-based decoding of individual spoken words at the fine-grained level of within semantic category is possible within and across the first and second language of bilingual adults. In particular, our observation of language invariant representations of spoken words in the left ATL concur with the role attributed to this region as a central 'semantic hub' emerged by the integration of distributed sensory and property-based specific representations (Patterson et al., 2007). Our results show benefits of multivariate pattern analysis based on the generalization of the pattern information across specific stimulus dimensions. This approach enabled examining the representation of spoken words independently of the speaker and the representation of semantic-conceptual information independently of the input language. Further studies employing similar decoding techniques for bilingual perception of a large variety of individual words or combinations of words in sentences promise to extend these findings to a brain-based account of unified semantic representations (Hagoort, 2005) as they are constructed in real-life multilingual environments.

Acknowledgements

This study was supported by the EU Marie Curie Initial Training Network Grant PITN-GA-2009-238593. Additionally, the authors would like to thank Marieke Mur, Job van den Hurk, Francesco Gentile and Inge Timmers for the valuable discussions throughout the design and implementation of the study and Sanne Kikkert for her assistance during the stimuli recording. The authors declare no conflict of interests.

References

Baayen, R. H., Piepenbrock, R. & Gulikers, L., (1995). The CELEX Lexical Database (CD-ROM). *Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA*.

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*, 2767–96.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A, Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex*, *10*, 512–28.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Ralph, M. a L. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral cortex*, *20*, 2728–38.

Boersma P. & Weenink D. (2001). Praat: doing phonetics by computer. *Glot International 5*, 341-345.

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and language*, *120*, 282–9.

Caramazza, A., & Mahon, B. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology*, *23*, 13–38.

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Cash, S. S., et al. (2011). First-pass selectivity for semantic categories in human anteroventral temporal lobe. *The Journal of neuroscience*, *31*, 18119–29.

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*, 3028–39.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*, 261–270.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., Aso, T. (2006). Language control in the bilingual brain. *Science*, *312*, 1537–40.

Damasio, H., Grabowski T. J., Tranel D., Hichwa R. D., & Damasio A. (1996). A neural basis for lexical retrieval. *Nature*, 380, 499-505.

De Moor, W., Ghyselinck, M., & Brysbaert, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, & Brysbaert. *Psychologica Belgica*, *40*(2), 99–114.

Forman, S. D., Cohen, J. D., Fiztgerald, M., Eddy, W. F., Mintun, M. A. & Noll D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of cluster-size threshold. *Magnetic resonance in medicine, 33*, 636-647.

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

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*, 1369–1381.

Galantucci B., Fowler C. A., Turvey M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review, 15*, 361-77.

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human brain mapping*, *27*, 392–401.

Hagler, D. J., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, *33*, 1093–103.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*, 416–23.

Hernàndez, M., Caño, A., Costa, A., Sebastián-Gallés, N., Juncadella, M., & Gascón-Bayarri, J. (2008). Grammatical category-specific deficits in bilingual aphasia. *Brain and language*, *107*, 68–80.

Hernàndez, M., Costa, A., Sebastián-Gallés, N., Juncadella, M., & Reñé, R. (2007). The organisation of nouns and verbs in bilingual speakers: A case of bilingual grammatical category-specific deficit. *Journal of Neurolinguistics*, *20*, 285–305.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*, *8*, 393–402.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*, 407–22.

Kambanaros, M., & van Steenbrugge, W. (2006). Noun and verb processing in Greek-English bilingual individuals with anomic aphasia and the effect of instrumentality and verb-noun name relation. *Brain and language*, *97*, 162–77.

Kriegeskorte, N., & Bandettini, P. (2007). Combining the tools: activation- and information-based fMRI analysis. *NeuroImage*, *38*, 666–8.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3863–8.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior research methods*, *44*, 325–43.

Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. (1965). *Doklady Akademii Nauk SSSR*, 163 (4): 845–848.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*, 487–97.

Martin, A. (2007). The representation of object concepts in the brain. *Annual review of psychology*, *58*, 25–45.

McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain and language*, *22*, 150–7.

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in neurosciences*, *32*(7), 376–82.

Meyer, K., Kaplan, J. T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature neuroscience*, *13*, 667–8.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*, 653–60.

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*, 424–30.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews. Neuroscience*, *8*, 976–87.

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

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7865–70.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and language*, *110*(2), 81–94.

Saitta, L., Vapnik, V. (1995). Support-Vector Networks. *Machine Learning, 20*, 273–297.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–6.

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*, 2418–25.

Shuster, L. I. (2009). The effect of sublexical and lexical frequency on speech production: An fMRI investigation. *Brain and language*, *111*, 66–72.

Stefanatos, G. A., Gershkoff, A., & Madigan, S. (2005). On pure word deafness, temporal processing, and the left hemisphere. *Journal of the International Neuropsychological Society, 11*, 456-70.

Talairach, J., & Tournoux, P. (1988). Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging. *Thieme Medical Publishers*.

Visser, M, & Lambon Ralph, M. a. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of cognitive neuroscience*, *23*, 3121–31.

Visser, M., Jefferies, E., Embleton, K. V., & Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of cognitive neuroscience*, *24*, 1766–78.

Supplementary tables

Table 1. Overall-activation to the spoken words. Talairach coordinates (center of gravity) for activation clusters in the left and right hemispheres, peak p-value and size of each cluster.

Brain Region	Talairach Coordinates			P value	Cluster size (sqmm)
	x	y	z		
Left Hemisphere					
Superior Temporal Gyrus	-45	-27	8	0.002	767
Superior Anterior Insula	-27	24	8	0.004	98
Right Hemisphere					
Superior Temporal Gyrus	48	-20	7	1×10^{-6}	1172
Superior Anterior Insula	35	20	13	7.8×10^{-5}	184
Inferior Frontal Gyrus	41	11	25	3×10^{-5}	140

Table 2. Within-language discrimination. Talairach coordinates (center of gravity) for clusters in the left and right hemisphere contributing to within language discrimination of words, peak p-value and size of each cluster. STG (superior temporal gyrus).

Brain Region	Talairach Coordinates			P value	Cluster size (sqmm)
	x	y	z		
Left Hemisphere					
Inferior Frontal	-50	12	15	0.004	16
	-39	31	20	0.037	19
Superior Frontal	-27	-14	49	0.027	34
	-22	9	49	0.002	17
	-39	1	41	0.037	30
	-20	20	43	0.004	59
Central Gyrus	-41	-25	49	0.006	16
Superior Parietal	-26	-59	52	0.006	16
	-21	-52	59	0.027	16
Anterior Insula	-29	10	12	0.037	24
	-26	6	-9	0.009	52
	-34	8	14	0.014	16
Occipital	-37	-79	10	0.002	24
Medial Frontal	-3	5	29	0.019	44
	-2	18	21	0.004	42
	-5	34	7	0.027	26
	-7	45	25	0.006	16
Medial Superior Frontal	-7	-1	51	0.009	17
	-7	18	45	0.006	63
Medial Middle Cortex	-3	-21	26	0.037	139
	-1	-26	11	0.014	47
Medial Superior Parietal	-10	-64	42	0.009	48
	-9	-49	41	0.009	19
	-9	-57	24	0.002	42
Medial Occipital	-3	-81	10	0.009	17

Brain Region	Talairach Coordinates			P value	Cluster size (sqmm)
	x	y	z		
Right Hemisphere					
Inferior Frontal	42	12	8	0.004	19
	14	42	-9	0.006	74
	45	35	1	0.037	61
Superior Frontal	38	20	35	0.019	134
	42	9	25	0.027	36
	35	-2	40	0.004	17
Precentral Gyrus	57	-4	18	0.006	20
Superior Parietal	33	-63	32	0.006	33
	37	-37	38	0.037	38
Supramarginal Gyrus	48	-36	27	0.006	18
	53	-42	34	0.006	46
Angular Gyrus	45	-55	23	0.006	45
Middle Temporal	49	-36	8	0.013	21
	41	-28	-19	0.027	143
	48	-50	-5	0.002	30
	54	-45	4	0.027	17
Anterior Temporal	41	-26	19	0.013	20
	48	-10	-23	0.037	46
	58	-20	-8	0.037	21
Dorsal STG	34	-30	14	0.037	17
Posterior Insula	45	-12	4	0.019	86
	35	1	16	0.027	43
	49	7	14	0.013	23
	49	-19	18	0.037	18
	35	-4	8	0.013	40
Anterior Insula	28	-3	-18	0.009	49
	37	19	8	0.013	19
Occipital	39	-79	8	0.037	54
	40	-65	1	0.009	24

	24	-80	18	0.009	17
Medial Frontal	10	49	11	0.004	46
Medial Superior Frontal	8	22	48	0.027	21
	7	3	56	0.009	24
Medial Middle Cortex	2	-30	11	0.004	85
	1	1	10	0.004	32
	1	-9	9	0.004	36
	3	6	-4	0.002	39
	4	-14	36	0.006	38
Medial Parietal	10	-52	56	0.019	18
Medial Occipital	4	-71	2	0.037	18
	1	-82	7	0.019	26
Medial Inferior Temporal	28	-24	-18	0.037	24
	29	-41	-16	0.006	25

Table 3. Across-language generalization. Talairach coordinates (center of gravity) for generalization clusters in the left and right hemispheres, peak p-value and size of each cluster.

Brain Region	Talairach Coordinates			P value	Cluster size (sqmm)
	x	y	z		
Left Hemisphere					
Anterior Temporal Lobe	-47	0	-15	0.004	52
Posterior Central Gyrus	-43	-31	40	0.002	95
Inferior Parietal/Angular Gyrus	-36	-69	30	0.004	62
Lateral Occipital Lobe	-27	-88	-1	0.004	68
Medial Occipital Lobe	-10	-68	5	0.002	54
Right Hemisphere					
Superior Anterior Insula	32	23	10	0.004	61
Inferior Anterior Insula	30	7	-8	0.002	44
Medial Anterior Temporal Lobe	38	3	-15	0.002	38
Posterior Superior Temporal Sulcus	46	-45	15	0.002	92
Medial Occipital Lobe	1	-78	7	0.002	52

CHAPTER 3

EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations

Corresponding manuscript

Correia JM, Jasma BM, Hausfeld L, Kikkert S, Bonte M (2015)

EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations.

Frontiers in Psychology, 6(2), 1-10.

Abstract

Spoken word recognition and production require fast transformations between acoustic, phonological and conceptual neural representations. Bilinguals perform these transformations in native and non-native languages, deriving unified semantic concepts from equivalent, but acoustically different words. Here we exploit this capacity of bilinguals to investigate input invariant semantic representations in the brain. We acquired EEG data while Dutch subjects, highly proficient in English listened to four monosyllabic and acoustically distinct animal words in both languages (e.g. 'paard'-'horse'). Multivariate pattern analysis (MVPA) was applied to identify EEG response patterns that discriminate between individual words within one language (within-language discrimination) and generalize meaning across two languages (across-language generalization). Furthermore, employing two EEG feature selection approaches, we assessed the contribution of temporal and oscillatory EEG features to our classification results. MVPA revealed that within-language discrimination was possible in a broad time-window (~50-620 ms) after word onset probably reflecting acoustic-phonetic and semantic-conceptual differences between the words. Most interestingly, significant across-language generalization was possible around 550-600 ms, suggesting the activation of common semantic-conceptual representations from the Dutch and English nouns. Both types of classification, showed a strong contribution of oscillations below 12 Hz, indicating the importance of low frequency oscillations in the neural representation of individual words and concepts. This study demonstrates the feasibility of MVPA to decode individual spoken words from EEG responses and to assess the spectro-temporal dynamics of their language invariant semantic-conceptual representations. We discuss how this method and results could be relevant to track the neural mechanisms underlying conceptual encoding in comprehension and production.

Introduction

Speech processing is a surprisingly flexible and accurate cognitive ability that allows humans to comprehend spoken language in real-time. At the individual word level, speech processing requires a continuous mapping of complex and variable auditory input signals to words and their semantic-conceptual representations. In turn, when we speak, we start from ideas and concepts and convert these into articulatory motor programs. In multilingual environments, these transformations involve the extraction of unified semantic concepts from variable acoustic/phonological word forms in native and non-native languages. When and how the bilingual brain performs these language-invariant conceptual transformations remains essentially unknown and is a focus of the present electroencephalography (EEG) study.

EEG allows studying non-invasively and with high temporal resolution the neural dynamics of speech processing. The temporal dynamics of EEG signals are informative of temporal order effects during speech processing. ERP (event-related potential) components at early time intervals, 100-200 ms after word onset, have been associated with phonetic/phonological processing (Uusvuori et al., 2008; Bonte & Blomert, 2004; Dumay et al., 2001; Sanders et al., 2003). Intermediate time intervals (200-300 ms) have been suggested to reflect early aspects of lexical access (Bonte et al., 2009, Salmelin, 2007; Hagoort et al., 2004, Van den Brink et al., 2001), followed by lexical/semantic processing in the 300-600 ms window, as indicated by ERP modulations dependent on semantic attributes of words, semantic priming and semantic context (Kutas & Hillyard, 1980; Hagoort, 2008). Spatially, this temporal signature of speech processing may reflect a spread of information from primary auditory areas to anterior temporal and frontal regions, mid-inferior and posterior temporal regions (Marinkovic et al., 2003) corresponding to the network of brain areas observed in functional magnetic resonance imaging (fMRI) studies of speech processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Binder et al., 2000). Complementary to ERP modulations, the oscillatory dynamics of EEG signals measured extracranially (Doelling et al.,

2014; Shahin et al., 2009; Hagoort et al., 2004; Strauß et al., 2014) and intracranially (Luo & Poeppel, 2007; Giraud & Poeppel 2012) have provided important insights regarding the function of underlying neural oscillations. Namely, an entrainment of theta band oscillations to the phoneme/syllable rate of speech signals, and the entrainment of gamma band oscillations to the phase of such theta band oscillations are suggested to reflect synchronization mechanisms that optimize the parsing of the speech signal into its relevant units (Lakatos et al., 2005; Giraud & Poeppel 2012; Peelle & Davis, 2012; Obleser et al., 2012).

A challenge is to investigate how these temporal and oscillatory EEG dynamics encode the representation of specific speech units, such as individual words and concepts. Recently, methods based on machine learning comprising multivariate statistics (MVPA, multivariate pattern analysis, Haxby et al., 2001, Formisano et al., 2008a) have shown their potential to solve this challenge. MVPA of EEG signals extends traditional univariate methods by exploiting the interaction between multiple signal features (e.g. spectro-temporal features across multiple electrodes and/or time points) using classification algorithms (Chan et al., 2011b; Hausfeld et al., 2012; Herrmann et al., 2012; Brandmeyer et al., 2013). The higher sensitivity of MVPA to find information content within brain imaging signals has significantly contributed to our understanding of the brain's responses to speech and language. In fMRI studies, multi-voxel patterns across early and higher-order auditory cortex have been shown to successfully predict the (perceptual) identity of individual speech sounds and speaker's voices (Formisano et al., 2008b; Kilian-Huetten et al., 2011; Bonte et al., 2014). Furthermore, fMRI responses in inferior parietal areas have been shown to differentiate words across different semantic categories (e.g., tools and dwellings, Shinkareva et al., 2011). At a more fine-grained within-category level, MVPA was recently shown to accurately predict which spoken noun a bilingual listener was listening to in one language (e.g., "horse" in English) based on the fMRI response patterns to equivalent nouns in the other language (e.g., "paard" in Dutch; Correia et al., 2014). This generalization of the meaning of words across languages specifically relied on focal regions, including the left anterior temporal lobe (left-ATL), suggesting the existence of 'hub' regions organizing

semantic-conceptual knowledge in abstract form (Patterson et al., 2007; Damasio et al., 1996; Visser & Lambon Ralph, 2011; Scott et al., 2000; Correia et al., 2014). Although more challenging in terms of the robustness of single trial estimates, also spatially/temporally distributed EEG/MEG patterns have been observed to discriminate individual speech sounds (Hausfeld et al., 2012), and words from different perceptual and semantic categories (Chan et al., 2011b; Simanova et al., 2010; Sudre et. al., 2012). Classification performances in EEG-MVPA studies on speech processing are typically low (e.g., below 0.55 in binary classification of spoken vowels, Hausfeld et al., 2012; or below 0.60 in binary classification of spoken words, Simanova et al., 2010). Besides the low signal-to-noise ratio of single trial EEG signals, EEG-based classification of individual words may be limited by the continuous and temporally variable processing of their phonological and semantic features (Van Petten et al., 1999). Importantly, however, multivariate approaches in EEG allow unravelling subtle differences in the neural processing of individual speech sounds that remain obscured in univariate approaches relying on average activation differences between experimental conditions.

Here, we employ MVPA to investigate spectro-temporal EEG response patterns capable of discriminating semantic-conceptual representations of words at the fine-grained level of within-category distinctions (animal nouns). To this end, we exploit the unique capacity of bilingual subjects to access semantic-conceptual information of spoken words from two languages. In separate Dutch and English blocks, we asked bilingual participants to listen to individual animal nouns (66.6% trials) and to detect non-animal target nouns (33.3% trials). The non-animal target nouns were presented as control task to ensure speech comprehension at every word presentation, but were not included in the analysis. Following supervised machine learning approaches, we trained multivariate classifiers (linear-SVM) to predict the identity of the perceived animal noun from new (untrained) samples of EEG activity (Figure 1A). In a first analysis we aimed to identify the EEG correlates involved in within-language word discrimination. To this end we trained classifiers to discriminate EEG responses to English (e.g. 'horse' vs. 'duck') and Dutch (e.g. 'paard' vs. 'eend') nouns. Importantly, stimuli included three exemplars of each

noun, pronounced by three different female speakers, allowing for speaker-invariant word discrimination ('*within-language*'). In a second analysis we aimed to assess the EEG correlates involved in language-independent decoding of the animal nouns ('*across-language*'). Here we trained classifiers to discriminate EEG responses to words in one language (e.g. in English, 'horse' vs. 'duck') and tested whether this training generalizes and allows discrimination of EEG responses to the corresponding nouns in the other language (e.g. in Dutch, 'paard' vs. 'eend'). Importantly, all words were acoustically-phonetically distinct both within and across languages. Based on this approach, we aimed to investigate whether language-independent representations are detectable in the EEG responses to individual spoken words. In particular, this approach allowed us to extract critical time windows and frequency ranges within the EEG relevant to semantic-conceptual encoding.

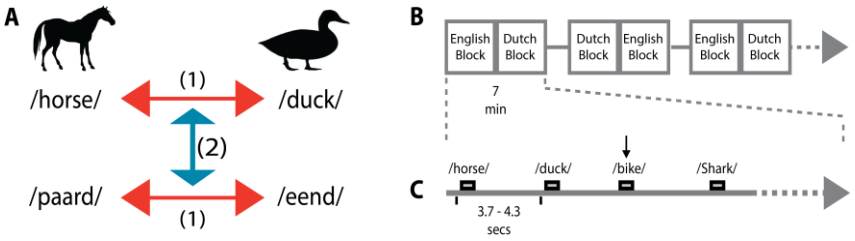


Figure 1. Experimental design. A) Within-language discrimination (1) was performed for all pairwise comparisons in English and Dutch. Across-language generalization (2) was performed across translational equivalent words in the other language. Both generalization directions were performed, from English to Dutch and from Dutch to English. B) Runs, blocks organization along the EEG session. Only 3 runs out of 8 runs are depicted for illustration. Each run (7 min) was composed by two blocks (English and Dutch). C) Within each block, a jittered interval (ITI) of 3.7-4.3 secs separates the presentation of the words. The black vertical arrow represents a response from the subjects to detect a non-animal word (e.g., bike).

Methods

Participants

Sixteen native Dutch (L1) participants proficient in English (L2) took part in the study (8 males and 8 females, right-handed, mean age=28.9 SD=3.4). The participants were undergraduate or postgraduate students of Maastricht University studying or working in an English speaking environment. All participants reported normal hearing abilities and were neurologically healthy. English proficiency was assessed with the LexTALE test, a vocabulary test including 40 frequent English words and 20 non-words (Lemhöfer & Broersma, 2012). The mean test score was 89.6% correct (SD=11.2%). This score is well above the average score (70.7%) of a large group of Dutch and Korean advanced learners of English performing the same test (Lemhöfer & Broersma, 2012). For comparison reasons, participants also conducted the Dutch version of the vocabulary test. The mean Dutch proficiency score was 94.1% (SD=3.3). The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht, The Netherlands.

Stimuli

Stimuli consisted of Dutch and English spoken words representing four different animals (English: 'Bull', 'Duck', 'Horse' and 'Shark', and the Dutch equivalents: 'Stier', 'Eend', 'Paard' and 'Haai') and six inanimate object words (English: 'Bike', 'Coat', 'Dress', 'Road', 'Suit' and 'Town'; and the Dutch equivalents: 'Fiets', 'Jas', 'Jurk', 'Weg', 'Pak' and 'Stad'). All animal nouns were monosyllabic and acoustically/phonetically distinct from each other both within and across languages. Phonetic distance between word pairs was quantified using the Levenshtein distance, which gives the number of phoneme insertions, deletions and/or substitutions required to change one word into the other, divided by the number of phonemes of the longest word (Levenshtein, 1966). On a scale from 0 (no changes) to 1 (maximum number of changes), the mean (SD) Levenshtein distances corresponded to 0.83 (0.15) for Dutch word pairs,

0.93 (0.12) for English word pairs and 1.00 (0.00) for English-Dutch word pairs. Furthermore, all animal nouns had an early age of acquisition in Dutch (mean=5.28 years SD=0.98; De Moor et al., 2000) and a medium-high frequency of use expressed on a logarithmic scale in counts per million tokens in Dutch (mean=1.29 SD=0.71) and in English (mean=1.50 SD=0.42; Celex database, Baayen et al., 1995). To add acoustic variability and allow for speaker-invariant MVPA analysis, the words were spoken by three female native Dutch speakers with good English pronunciation. Stimuli were recorded in a sound proof chamber at a sampling rate of 44.1 kHz (16 bit resolution). Post-processing of the recorded stimuli was performed in PRAAT software (Boersma & Weenink, 2011) and included band-pass filtering (80-10500 Hz), manual removal of acoustic transients (clicks), length equalization, removal of sharp onsets and offsets using 30 ms ramp envelopes, and amplitude equalization (average RMS). Stimulus length was equated to 600 ms (original range: 560–640 ms) using PSOLA (75–400 Hz as extrema of the F0 contour). We carefully checked the stimuli for possible alterations in F0 after length equation and did not find any detectable changes. We assured that the produced stimuli were unambiguously comprehended by the participants during the stimuli familiarization phase prior to the experiment.

Experimental procedures

The experimental session was organized in 8 runs, each run containing 2 blocks (one Dutch and one English). Each block included 36 nouns: 24 animal nouns and 12 (33.3%) non-animal nouns. The order of English and Dutch blocks was counterbalanced across runs: odd runs started with an English block followed by a Dutch block; even runs started with a Dutch block followed by an English block (Figure 1B). Participants were instructed to actively listen to the stimuli and to press a button (with the left index finger) whenever they heard a non-animal word. The goal of the task was to help maintaining a constant attention level throughout the experiment and to promote speech comprehension at every word presentation. All participants paid attention to the words as indicated by a mean (SD) detection accuracy of 98.3 (1.4) %. Data from non-animal trials were excluded from further analysis. The 24 animal nouns in each

block corresponded to 6 repetitions of each of the 4 animal nouns. Because nouns were pronounced by 3 different speakers, each physical stimulus was repeated twice in each block. Stimulus presentation was pseudo-randomized within each block, avoiding consecutive presentations of the same words or sequences of words. Throughout the experiment, each animal noun was presented 48 times per language.

EEG acquisition and preprocessing

Data were recorded with a sampling rate of 250 Hz in an electrically shielded and sound-proof room from 62 electrode positions (Easycap, Montage Number 10, 10-20 system) relative to a left mastoid reference signal. The ground electrode was placed on the Fz electrode. Impedance levels were kept below 5 kΩ. During the EEG measurement, stimuli were presented binaurally at a comfortable intensity level. According to an event-related design (figure 1C-D), the averaged inter-trial-interval between two stimuli was 4 seconds (jittered randomly between 3.7 sec and 4.3 sec). Each run took 7 minutes, resulting in a total EEG measurement time of 56 minutes. A gray fixation cross against a black background was used to keep the visual stimulation constant during the whole duration of a block. Block and run transitions were marked with written instructions. Participants were instructed to minimize eye-movements during the auditory presentation and fixate on the fixation cross.

Data preprocessing was performed using EEGlab (Delorme & Makeig, 2004) and included band-pass filtering (0.1 to 100 Hz) followed by epoch extraction locked to the onset of the animal nouns (-1000 to 1000 ms) and baseline correction (-1000 to 0 ms).

Removal of signal artifacts was performed in two steps. First, the data were visually inspected and epochs containing nonstereotypical artifacts including high-amplitude, high-frequency muscle noise, swallowing, and electrode cable movements, were rejected (mean 4.31 trials per subject, SD 2.36). Second, stereotypical artifacts related to eye movements, eye-blinks and heart beat artifacts were corrected with extended INFOMAX ICA (Lee et al., 1999) as implemented in EEGLAB. Because data were recorded at 62 channels, runica decomposed the data in 62 component activations per subject. These

component activations were categorized as EEG activity or nonbrain artifacts by visual inspection of their scalp topographies, time courses, and frequency spectra. Criteria for categorizing component activations as EEG activity included (1) a scalp topography consistent with an underlying dipolar source, (2) spectral peak(s) at typical EEG frequencies, and (3) regular responses across single trials, i.e., an EEG response should not occur in only a few trials (Delorme et al., 2004). Based on these criteria, component activations representing nonbrain artifacts were removed, and EEG data were reconstructed from the remaining component activations representing brain activity. The resulting ICA-pruned data sets were baseline corrected (-1000 to 0 ms) and used for further analysis.

ERP and ERSP analysis

First, in order to validate typical EEG responses to spoken words reported in the literature, we performed univariate analyses. These were conducted in EEGlab (Delorme & Makeig, 2004) and included: (1) an ERP analysis based on the average amplitude of signal change over time with respect to baseline (-1000 to 0 ms) and (2) an ERSP (event-related spectral perturbation) analysis based on averaged power changes of all words over frequency and time with respect to baseline (-1000 to 0 ms). For the ERSP analysis we employed a Hanning taper fast fourier transform (FFT) filter from 1 to 60 Hz on a linear frequency scale with steps of 2 Hz, producing 30 filtered signals. Group statistics for the ERP and ERSP were performed at random-effects using two-sided Wilcoxon tests for each time-point versus zero baseline and corrected for multiple comparisons using FDR (alpha=5%).

Multivaritate classification analysis

Multivariate classification was employed to investigate whether specific temporal or spectrotemporal EEG signatures enable the discrimination of words within and across languages. To this end we used a supervised machine learning algorithm (linear support vector machines, linear-SVM; Cortes & Vapnik, 1995) as implemented by the Bioinformatics Matlab toolbox (maximum number of learning iterations = 15000). Classifications were performed to evaluate whether patterns of EEG data pertained relevant

information encoding the representations of spoken words (within-language discrimination) as well as their language invariant semantic-conceptual representations (across-language generalization). All classifications were binary (i.e., chance-level is 0.5) and involved discrimination and generalization between two words. The results of these binary predictions were then averaged across all possible pair-wise classifications. Additional methodological steps encompassing the computational strategy to validate the classification results (cross-validation) and to select the EEG features used for classification (feature selection) are described below.

Cross-validation approaches

Cross-validation of the multivariate classification analysis served two purposes: 1) to obtain robust estimates of the discrimination accuracies; 2) to allow generalization of classes by using distinct class groupings during the training and testing phases of classification. Cross-validation for within-language word discrimination relied on speaker identity. Here, we trained a classifier to discriminate words based on samples recorded from two out of the three speakers that pronounced the words (32 trials per word) and tested whether this training was able to generalize the left-out speaker pronouncing the same words (16 trials per word). This cross-validation procedure assured word discrimination invariant to neural activations specific to acoustic-phonetic characteristics of the speakers. Cross-validation for across-language generalization of semantic concepts relied on language independent information of the words. Here, we trained a classifier to discriminate words within one language (48 trials per word) and tested whether this training generalized to the other language (48 trials per word). Hence, in across-language generalization, we aimed to isolate semantic conceptual properties of the words that were language invariant.

Feature selection approaches

Temporal-windows approach (shifting-windows + all channels):

To investigate the temporal evolution of spoken word decoding, we selected EEG response features using shifting-windows (width=40 ms - 10 time points)

across all channels (Figure 2B). Restricting the EEG signal features to specific time windows permits the calculation of changes in classification accuracies over time informative of spoken word processing. Because the temporal-windows approach reduces the number of EEG features used for classification, it increases the temporal sensitivity of the classifiers to speaker and language invariant information of the spoken words due to a potentially better match between the training and testing patterns (Hausfeld et al., 2012). Additionally, it reduces the high dimensionality of the feature space, thus avoiding degraded classification performances (model overfitting; for a description, see Norman et al., 2006). The empirical null distribution was computed per subject using 200 label permutations. Individual statistical significance ($p<0.05$) was calculated based on deviance from permuted accuracies. Group level statistics were calculated based on the overlap of significant subjects across time intervals using a binomial test with $n=16$ (number of subjects) and $p=0.05$ (Darlington & Hayes, 2000; Hausfeld et al., 2012) and corrected for multiple comparisons (time windows) using FDR correction (alpha=5%).

Time-frequency approach (filtered-band-out + shifting-windows + all channels):

To assess the importance of brain oscillations in specific frequency bands to the performance of the classifiers we employed a feature selection approach combining temporal shifting windows and filter-band-out (Figure 2C). The original epoched EEG responses (-1000 to 1000 ms) were filtered prior to classification using an FIR (finite impulse response) filter as implemented in EEGlab (Delorme & Makeig, 2004). The width of the filtered-out frequency band was set to 4 Hz, centered on frequencies from 2 up to 60 Hz and in frequency steps of 2 Hz, producing 30 filtered signals. For each of the filtered signal versions, we subsequently performed the temporal-windows approach to assess the importance of each frequency band over time. The importance of the left-out frequency band was quantified in terms of a change in classification performance with respect to the non-filtered signal. To prevent a modulation of time-frequency importance due to differences in the original classification accuracy, a normalization of the importance of each time-frequency bin with respect to the accuracy limits (0-1) was performed using 'odds-ratio'

normalization (Szumilas, 2010). Odds-ratio values above 1 indicate a reduction of classification accuracy after a specific frequency band is filtered out. This approach allowed us to investigate the contribution of each frequency band over time without disrupting EEG spectral interactions that may be crucial in many cognitive processes, including speech processing (Giraud & Poeppel, 2012; Peelle & Davis, 2012; Henry & Obleser, 2012). Group statistics were performed in random-effects (two-sided Wilcoxon's test) and corrected for multiple comparisons using FDR correction (alpha=5%).

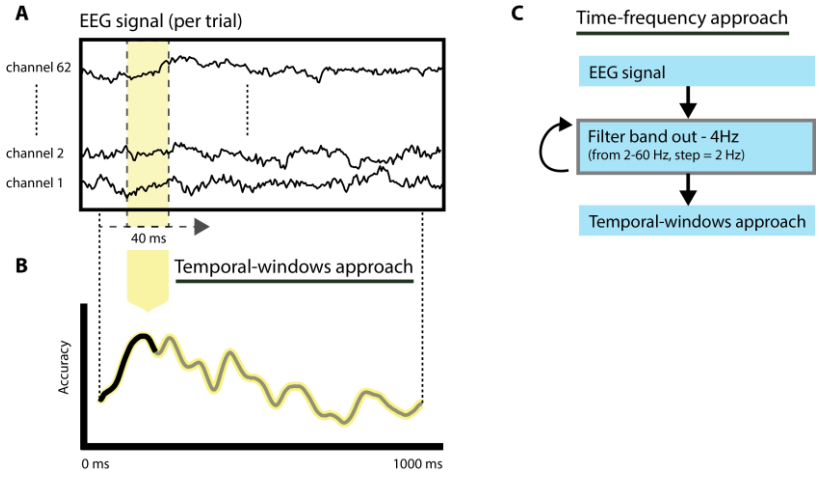


Figure 2. Illustration of the feature selection approaches. A) The original epoched EEG response per word corresponds to the signal of all EEG channels and the interval from 0-1000 ms after word onset. B) Temporal-windows approach. Classification relies on temporal windows of 40 ms (10 time-points) and all channels, resulting in classification accuracies over time. C) Time-frequency approach. Thirty leave-band-out filtered versions of the signal are created (from 2 to 60 Hz, band-width = 4Hz) following classification based on the same procedure employed in the temporal-windows approach.

Results

ERPs and time-frequency analysis

We first conducted univariate analyses of ERP and time-frequency changes relatively to stimulus baseline in order to assess the overall spectro-temporal characteristics of EEG responses evoked by the animal words. Figure 3 illustrates the averaged ERP responses elicited by the different animal words, including the expected ERP peaks (channel Fcz, Figure 3A) and their corresponding topographies (Figure 3B), in the N1 window (120-160 ms), the P2 window (230-390 ms) and the N400 window (550-800 ms). To assess univariate differences between the ERP responses we conducted all possible word-to-word contrasts within the same language (e.g., horse vs duck), as well as all possible concept-to-concept contrasts (e.g., horse + paard vs duck + eend). None of the possible contrasts yielded significant differences within or across participants.

The analysis of averaged power changes in different frequency bands (Figure 3C) shows an average power increase (ERS, event-related synchronization) of slow oscillations (1 to 10 Hz) starting 100 ms after stimulus onset, followed by a steep reduction in alpha power (ERD, event-related desynchronization) between 400 and 500 ms. At later time intervals, the ERS of slow oscillations (1 to 8 Hz) was maintained. These differences did not allow the systematic discrimination of individual words nor of language-independent concepts.

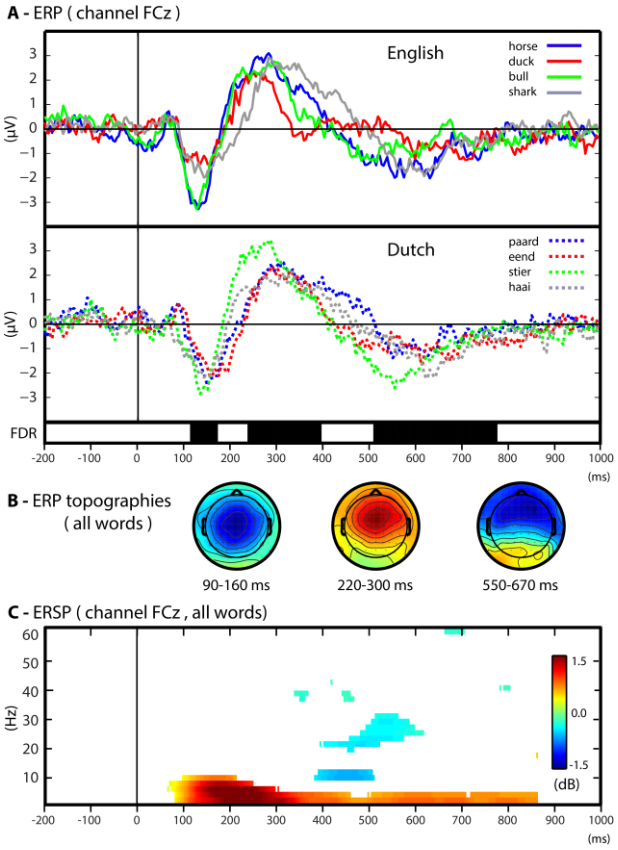


Figure 3. Univariate results. A) ERP in respect to baseline of each word over the channel FCz. The ERPs for English and Dutch words are plotted separately. Group level statistics of all words with respect to baseline (Wilcoxon's test, FDR corrected < 0.05) is depicted in black bars during the time course of the ERP responses. B) ERP scalp maps for time-intervals characteristic of the ERP components (N1: 90-160; P2: 220-300; N400: 550-670). C) ERSP (dB) with respect to baseline for all words. The ERSP time-frequency plot includes a statistical threshold for group level significance (Wilcoxon's test in respect to baseline period, FDR correction, alpha=0.05).

Multivariate analysis (MVPA)

The multivariate analysis consisted of assessing the ability of multivariate classifiers to discriminate words within the same language and across first and second language in bilingual subjects. To assess the contribution of specific

EEG features used for classification we used two feature selection approaches: a *temporal-windows approach*, relying on restricted time intervals (40 ms) shifted over time and all EEG channels; and a *time-frequency approach*, relying on a combined selection of features using the *temporal-windows approach* and a moving filter-band-out procedure (4 Hz bands with an step of 2 Hz).

The *temporal-windows* feature selection approach enabled identifying specific time-intervals related to word decoding. Within-language discrimination (Figure 4A) was significantly possible throughout most of the time-course from ~50 until 620 ms after word onset. Within this broad time window, salient local maxima of accuracies were identified for the temporal windows (40 ms) around 160 ms (accuracy = 0.535), 225 ms (accuracy = 0.537), 390 ms (accuracy = 0.533), 570 ms (accuracy = 0.513) and 820 ms (accuracy = 0.512). Interestingly, across-language generalization (Figure 4B) led to significant classification in more restricted temporal windows with significant results between 550-600 ms (maximum accuracy = 0.511) and 850-900 ms (maximum accuracy = 0.508). A further time-interval showing a trend (uncorrected $p<0.05$) for across-language generalization capacity was observed around 400 ms (maximum accuracy = 0.507).

The *time-frequency* feature selection approach assessed the contribution of oscillatory activity in specific frequency bands to word decoding across the different time windows. For this purpose, 'odds-ratio' values were computed, group averaged and thresholded for statistical significance (random-effects, FDR=5%). Overall, the temporal profiles of the *time-frequency approach* match consistently with that of the *temporal-windows approach,* confirming that reductions in classification accuracy due to the omission of specific frequency bands occurred in time windows relevant for word decoding (Figure 4C). For within-language discrimination of words, reductions in classification accuracy especially occurred when omitting slow oscillations (below 12 Hz, delta, theta and alpha). For across-language generalization (Figure 4D), the period around 600 ms that showed significant generalization capacity, was characterized by accuracy reductions when filtering out frequencies up to 10 Hz (delta-theta-alpha). In other time windows a contribution of slow oscillations was also

observed for this analysis, although involving slower oscillations (delta/ low theta, below 6 Hz). Visual inspection of Figure 4C-D further suggested that besides the sensitivities for oscillations below 12 Hz, for both types of analysis smaller classification drops occurred across gamma band (above 30 Hz) as well as across broad-band oscillation profiles.

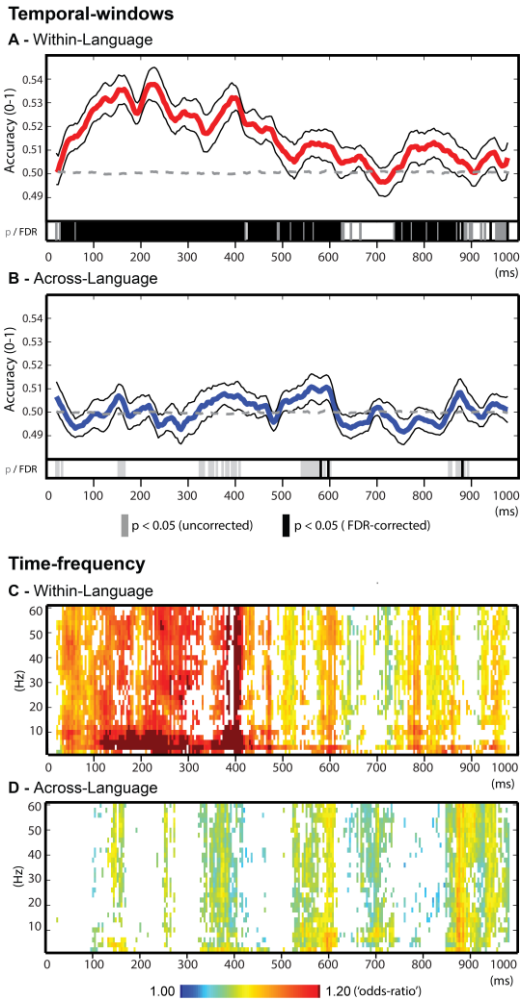


Figure 4. Decoding results for the temporal-windows and time-frequency feature selection approaches. A) Temporal-windows approach for within-language discrimination. Group average accuracy time-course depicted in red line, the black lines represent one standard error above and below the average accuracy. B) Temporal-windows approach for across-language generalizations. Group average accuracy time-course depicted in blue line, upper and lower standard errors in black lines. A-B) Statistical results are reported at the group level (binomial test, $p<0.05$) in gray bars and in black bars after FDR correction (alpha=5%). C) Time-frequency approach for within-language discrimination. D) Time-frequency approach for across-language generalization. C-D) Results are reported as averaged 'odds-ratio' values at the group level (scaled between 1 and 1.2) and threshold using Wilcoxon's test following FDR correction (alpha=5%).

Discussion

By combining EEG multivariate pattern analysis and an experimental design that exploits the unique capacities of bilingual listeners we identified specific time windows and oscillations enabling within-category discrimination of individual spoken words. We demonstrated within-language word decoding in a broad time-window from ~50 to 620 ms after word onset with a strong contribution of slow oscillations (below 12 Hz). Most importantly, we were able to isolate specific time windows, including the 550-600 ms window, in which EEG features enabled the generalization of the meaning of the words across their Dutch and English word forms. Our results demonstrate the feasibility of using MVPA to identify individual word representations based on speech evoked EEG signals. Furthermore, they indicate the advantage of feature selection approaches in assessing temporal and temporal-oscillatory EEG response features in classification.

The univariate analyses illustrate ERP and oscillatory responses typically elicited by individual spoken words (Strauß et al., 2014; Bastiaansen et al., 2008; Hagoort et al., 2004; Kutas & Federmeier, 2000; Bonte et al., 2009) indicating a progression from acoustic-phonetic to lexical-semantic processing. The ERPs to the individual words show variability as a consequence of acoustic-phonetic differences and other word-specific properties. However, these differences did not allow the systematic discrimination of individual words nor of language-independent concepts. The prevalence of slow oscillatory activity (below 12 Hz) while subjects listened to the words indicates the crucial role of these frequencies in the processing and comprehension of speech (Strauß et al., 2014; Hagoort et al., 2004; Giraud & Poeppel, 2012). The analysis also showed that the univariate frequency power changes were not suitable for distinguishing individual words or across-language generalization of semantic concepts.

Importantly, the multivariate analyses allowed finding neural time-course correlates of the individual words that were invariant to the acoustic-phonetic characteristics of the speakers (within-language discrimination) as well as to the language in which the meaning was presented (across-language generalization).

Within-language word discrimination relied on acoustic-phonetic and semantic-conceptual differences between the nouns, but also on possible other differences reflecting their individual properties. Accordingly, within-language discrimination was possible for both approaches of feature selections employed. In the *temporal-windows* approach (Figure 4A), investigating the temporal evolution of classification across consecutive short time-intervals of 40 ms, classification performance was significant from ~50 until 620 ms after word onset. In accordance with the ERP literature, decoding in this broad time window may be reflect a progression from phonetic-phonological processing (100-200 ms; Uusvuori et al., 2008; Bonte & Blomert, 2004; Dumay et al., 2001; Sanders et al., 2003) to initial lexical access (200-300 ms; Bonte et al., 2009, Salmelin, 2007; Hagoort et al., 2004, Van den Brink et al., 2001), and lexical semantic processing (300-600 ms; Kutas & Hillyard, 1980; Hagoort, 2008). These results are also consistent with previous single trial auditory word classification (Simanova et al., 2010) that showed initial prominent classification capability centered around 240 ms followed by a second less prominent capability around 480 ms after word onset.

The second multivariate analysis - across-language generalization - relied uniquely on language invariant semantic-conceptual properties of the nouns. This analysis, and especially the temporal-window approach (Figure 4B), revealed language invariant EEG features coding for the animal words in much more restricted time-windows including the 550-600 ms window and the 850-900 ms window at the end of the EEG epoch. ERP research has commonly associated similar time intervals with lexical-semantic processing of words across different task and sentence contexts (Kutas & Federmeier, 2000; Hagoort, 2008). Here, we indicate the potential of EEG signals to represent semantic-conceptual information of individual words independent of their acoustic-phonetic implementation or word-form. In order to isolate these input-invariant lexical-semantic representations we used animal nouns that were acoustically-phonetically distinct both within and across languages and were presented together with non-animal nouns that served as targets. In everyday speech processing, it is more difficult to disentangle input-driven versus input-independent processes as initial lexical-semantic access is influenced by both

acoustic-phonetic word form information (McClelland & Elman, 1986; Marslen-Wilson, 1987) and semantic or task context (Bonte, 2004; Çukur et al., 2013; Obleser et al., 2004), leading to early lexical and/or semantic ERP modulations around 200-300 ms (e.g., Bonte et al., 2006; Van den Brink et al., 2001; Travis et al., 2012; Strauß et al., 2014). Our approach presents a way to disentangle these aspects of comprehension. Importantly, by using words belonging to the same semantic category – animals – we reduced the influence of larger scale semantic category differences that can also drive the decoding of individual nouns (Simanova et al., 2010; Shinkareva et al., 2011; Chan et al., 2011b).

In later time-windows, significant classification for within-language discrimination (750-900 ms) and across-language generalization (850-900 ms) may reflect effects specific to our paradigm. That is, the slow presentation of words and/or the use of a target detection task, may have led to e.g. subvocal rehearsal in working memory (Baddeley, 2003; Kutas & Federmeier, 2000; Buchsbaum et al., 2011) and/or response monitoring towards the end of the trial (Wang et al., 2012).

In bilinguals, the active translation of written words during speech production tasks has been shown to elicit ERP differences for translation direction around 400 ms after word presentation (Christoffels et al., 2013). In the current study the effect of direct translations was minimized in several ways. First, we avoided active translations from second to native language and vice-versa by separately presenting words in Dutch and English blocks and using catch trials consisting of Dutch and English non-animal words, respectively. Furthermore, we used a selection of words with relatively early age of acquisition and of medium-high frequency of use in both languages.

To further understand the EEG temporal patterns allowing classification, we employed a time-frequency feature selection approach that assessed the relative contribution of oscillatory bands. We observed a significant contribution of slow EEG oscillations (below 12 Hz) for within-language and across-language classification, which links to the synchronization of oscillatory bands observed in the ERSP analysis. Furthermore, in the time windows during which the

slower oscillations most strongly influenced classification performance, results also indicated a contribution from higher, gamma band oscillations (above 30 Hz). It would be interesting to replicate this possible co-occurrence of slower and gamma band modulations in future studies with bilinguals, and, in particular to test how they relate to suggested processing of (phonemes, syllables and semantic information (Peña & Melloni, 2012; Giraud & Poeppel, 2012; Lakatos et al., 2005; Peelle & Davis, 2012).

We may hypothesize that the neural processing underlying the EEG-based translations of animal nouns occurs in a brain network that was recently identified in an fMRI study using a comparable bilingual paradigm (Correia et al., 2014). In particular, in this previous study, language-invariant classification of animal words was found to rely on focal brain regions, including the left anterior temporal lobe (left-ATL), corroborating the existence of "hub" regions organizing semantic-conceptual knowledge in abstract form. Correspondingly, recent models of conceptual knowledge (Pattersson et al., 2007), brain lesion studies (Damasio et al., 1996) and neuroimaging evidence (Visser et al., 2012; Correia et al., 2014) locate a possible semantic hub within the left-ATL, integrating distributed semantic-conceptual information throughout the cortex. Furthermore, distributed neural representations of semantic information may also connect to modality specific brain regions subserving perception and action (Martin, 2007; Meyer & Damasio, 2009). Interestingly, magnetoencephalography (MEG) studies have related time windows starting at 400 ms after spoken word onset to semantic processing in bilateral anterior temporal areas (Marinkovic et al., 2003; Chan et al., 2011a), suggesting a putative link between the present finding of language-independent word decoding in the 550-600 ms time window and processing in these brain regions. At present, this spatial-temporal association remains speculative, but similar classification paradigms using simultaneous fMRI and EEG recordings (de Martino et al., 2011) may allow investigating the joint spatio-temporal representation of spoken words. Furthermore, earlier indications of semantic/conceptual representations of our words are observed in a spread time window between 320-420 ms after word onset (uncorrected $p<0.05$). These and possibly even earlier semantic activations elicited by the individual

animal words may be more difficult to detect due to variability in the exact timing of these initial activations.

Overall, our results show the benefit of EEG-based MPVA to investigate the representation of semantic concepts independently of the input language and more generally of individual spoken words independently of the speaker. Although the obtained accuracies are relatively low, they demonstrate the sensitivity of multivariate classification to distinguish subtle representations extracted from single-trial EEG responses that may not be present in the averaged EEG signal across multiple trials (Hausfeld et al., 2012, Makeig et al., 2002). Furthermore, our results show the potential of feature selection approaches based on moving temporal windows to highlight time windows associated with the neural processing of specific characteristics of speech and language (e.g. language independent semantic processing, see also Hausfeld et al., 2012; Simanova et al., 2010; Chan et al. 2011b). Future studies including different sets of words, languages or feature selection approaches may help confirming the generalization of our results. Beyond decoding language-invariant semantic-concepts during listening, EEG-based MVPA may also be used to investigate whether semantic-concepts share a similar neural representation during reading and speaking (Pickering & Garrod, 2013; Hickok et al., 2011). When we speak, we start from ideas and concepts and convert these into articulatory motor programs. ERP studies on speech production (e.g. picture naming), relate early windows, 100-200 ms after stimulus onset to interactive processing of visual encoding and accessing concepts for language use (Rahman & Sommer, 2003; Redmann et al., 2014). Like in speech comprehension, this interaction between input-dependent and abstract semantic-conceptual representations in speech production, together with their strong context and task-dependency (e.g. Aristei et al., 2011, Jescheniak et al., 2002), makes it difficult to isolate abstract semantic conceptual representations using univariate analysis methods. Because our EEG-based MVPA approach may disentangle these processes, it would thus be of interest to employ this same approach in speech production studies (e.g. Koester & Schiller, 2008; and Schmitt et al., 2000). In particular, a similar bilingual paradigm involving word naming in bilingual speakers would allow investigating the timing of language-

independent semantic-conceptual representations. Furthermore, the classification of spoken words across and within languages in bilingual speakers and across and within speech modality (perception and production) may allow to investigate neural representations crucial for the initiation of speech production (Levelt, 1989, Indefrey & Levelt, 2004, 2011; Rahman & Sommer, 2003), as well as, for the monitoring of speech output (Hickok et al., 2011).

References

Aristei S, Melinger A, Rahman AR (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, *23*(7), 1567-1586.

Baayen RH, Piepenbrock R, Gulikers L (1995) The CELEX Lexical Database (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley A (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208.

Bastiaansen MCM, Oostenveld R, Jensen O, Hagoort P (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, *106*(1), 15–28.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex (New York, N.Y. : 1991)*, *10*(5), 512–28.

Boersma P, Weenink D (2011) Praat: doing phonetics by computer. Glot Int 5:341–345.

Bonte M (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology*, *115*(2), 409–423.

Bonte M, & Blomert L (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology*, *115*(2), 409–423.

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(13), 4548–57.

Bonte M, Parviainen T, Hytönen K, Salmelin R (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(1), 115–23.

Bonte M, Valente G, Formisano E (2009). Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(6), 1699–706.

Brandmeyer A, Farquhar JDR, McQueen JM, & Desain PWM (2013). Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. *PloS One*, *8*(7), e68261.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–28.

Chan AM, Baker JM, Eskandar E, Schomer D, Ulbert I, Marinkovic K, Halgren E (2011a). First-pass selectivity for semantic categories in human anteroventral temporal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(49), 18119–29.

Chan AM, Halgren E, Marinkovic K, Cash SS (2011b). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*(4), 3028–39.

Christoffels IK, Ganushchak L, & Koester D (2013). Language conflict in translation: An ERP study of translation production. *Journal of Cognitive Psychology*, *25*(5), 646–664.

Correia JM, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience*, *34*(1), 332–8.

Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20:273–297.

Cox DD, & Savoy RL (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Çukur T, Nishimoto S, Huth AG, Gallant JL (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–70.

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR (1996) A neural basis for lexical retrieval. *Nature*, 380:499 –505.

Darlington RB, Hayes AF (2000). Combining independent p values: extensions of the Stouffer and binomial methods. *Psychol. Methods*, 5 (4), 496–515.

De Martino F, de Borst AW, Valente G, Goebel R, Formisano E (2011). Predicting EEG single trial responses with simultaneous fMRI and relevance vector machine regression. *NeuroImage*, *56*(2), 826–36.

De Moor W, Ghyselinck M, Brysbaert M (2000) A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, and Brysbaert. Psychologica Belgica 40:99 –114.

Delorme A, Makeig S (2004). EEGLAB: An open source toolbox for analysis of single trial.

Doelling KB, Arnal LH, Ghitza O, Poeppel D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85 Pt 2*, 761–8.

Dumay N, Benraïss A, Barriol B, Colin C, Radeau M, Besson M (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, *13*(1), 121–43.

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

Formisano E, De Martino F, Valente G (2008a). Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging*, *26*(7), 921–34.

Giraud A-L, & Poeppel D (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–7.

Gross J, Pollok B, Dirks M, Timmermann L, Butz M, Schnitzler A (2005). Task-dependent oscillations during unimanual and bimanual movements in the human primary motor cortex and SMA studied with magnetoencephalography. *NeuroImage*, *26*(1), 91–8.

Hagoort P (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1055–69.

Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004). Integration of word meaning and world knowledge in language comprehension. *Science (New York, N.Y.)*, *304*(5669), 438–41.

Hausfeld L, De Martino F, Bonte M, Formisano E (2012). Pattern analysis of EEG responses to speech and voice: influence of feature grouping. *NeuroImage*, *59*(4), 3641–51.

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Ramadge, PJ (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–16.

Henry MJ, & Obleser J (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(49), 20095–100.

Herrmann B, Maess B, Kalberlah C, Haynes J-D, & Friederici AD (2012). Auditory perception and syntactic cognition: brain activity-based decoding within and across subjects. *The European Journal of Neuroscience*, *35*(9), 1488–96.

Hickok G, Houde J, Rong F (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Hickok G, Poeppel D (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8:393– 402.

Indefrey P, & Levelt WJM (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–44.

Jescheniak JD, Schriefers H, Garrett MF, Friederici AD (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience*, *14*(6), 951–64.

Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(5), 1715–20.

Koester D, & Schiller NO (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *NeuroImage*, *42*(4), 1622–30.

Kutas M, & Federmeier K (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Kutas M, & Hillyard SA (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science, New Series, Vol. 207, 203-205.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology* 94, 1904–1911.

Lee TW, Girolami M, Sejnowski TJ (1999) Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput 11:417– 441.

Lemhöfer K, & Broersma M (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–43.

Levelt WJM (1989). Speaking: from intention to articulation. *Cambridge, Mass.: MIT Press*.

Levenshtein VI (1965) Binary codes capable of correcting deletions, insertions, and reversals. Doklady Akademii Nauk SSSR 163:845– 848.

Luo H, & Poeppel D (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–10.

Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002). Dynamic brain sources of visual evoked responses. *Science (New York, N.Y.)*, *295*(5555), 690–4.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*(3), 487–97.

Marslen-Wilson W (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.

Martin A (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

Mcclelland JL, Elman JL (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, 18, 1-86.

Meyer K, & Damasio A (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, *32*(7), 376–82.

Meyer K, Kaplan JT, Essex R, Webber C, Damasio H, Damasio A (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience*, *13*(6), 667–8.

Norman KA, Polyn SM, Detre GJ, Haxby, JV (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30.

Obleser J, Elbert T, Eulitz C (2004). Attentional influences on functional mapping of speech sounds in human auditory cortex. *BMC Neuroscience*, *5*, 24.

Obleser J, Herrmann B, Henry MJ (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Frontiers in Human Neuroscience*, *6*(8), 250.

Patterson K, Nestor PJ, Rogers TT (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, *8*(12), 976–87.

Peelle JE, & Davis MH (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*(September), 320.

Peña M, & Melloni L (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience*, *24*(5), 1149–64.

Pickering MJ, & Garrod S (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–47.

Rahman RA, & Sommer W (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Brain Research. Cognitive Brain Research*, *16*(3), 372–82.

Rauschecker JP, & Scott SK (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–24.

Redmann A, Fitzpatrick I, Hellwig F, Indefrey P (2014). The use of conceptual components in language production: an ERP study. *Frontiers in Psychology*, *5*(4), 363.

Salmelin R (2007). Clinical neurophysiology of language: the MEG approach. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, 118(2), 237–54.

Sanders LD, & Neville HJ (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Brain Research. Cognitive Brain Research*, 15(3), 228–40.

Schmitt BM, Münte TF, & Kutas M (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*(4), 473–84.

Scott SK, Blank CC, Rosen S, Wise RJ (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain : A Journal of Neurology*, *123 Pt 12*, 2400–6.

Shahin AJ, Picton TW, Miller LM (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, *70*(3), 259–66.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011). Commonality of neural representations of words and pictures. *NeuroImage*, *54*(3), 2418–25.

Simanova I, van Gerven M, Oostenveld R, Hagoort P (2010). Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PloS One*, *5*(12), e14465.

Strauß A, Kotz SA, Scharinger M, Obleser J (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *NeuroImage*, *97C*, 387–395.

Strijkers K, Costa A, Thierry G (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(4), 912–28.

Sudre G, Pomerleau D, Palatucci M, Wehbe L, Fyshe A, Salmelin R, Mitchell T (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage*, *62*(1), 451–63.

Szumilas M (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie Canadienne de Psychiatrie de L'enfant et de L'adolescent*, *19*(3), 227–9.

Travis KE, Leonard MK, Chan AM, Torres C, Sizemore ML, Qu Z,Halgren E (2012). Independence of Early Speech Processing from Word Meaning. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–10.

Uusvuori J, Parviainen T, Inkinen M, Salmelin R (2008). Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(2), 456–66.

Van den Brink D, Brown CM, Hagoort P (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–85.

Van Petten C, Coulson S, Rubin S, Plante E, Parks M (1999) Time course of word identification and semantic integration in spoken language. J. Exp. Psychol. Learn. Mem. Cognit. 25, 394–417.

Visser M, & Lambon Ralph MA (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of Cognitive Neuroscience*, *23*(10), 3121–31.

Visser M, Jefferies E, Embleton KV, Lambon Ralph MA (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience*, *24*(8), 1766–78.

Wang L, Jensen O, van den Brink D, Weder N, Schoffelen J-M, Magyari L, Bastiaansen M (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping*, *33*(12), 2898–912.

CHAPTER 4

Brain-based decoding of spoken syllables isolates articulatory representations during passive listening

Manuscript submitted

Correia JM, Jansma BM, Bonte M

Brain-based decoding of spoken syllables isolates articulatory representations during passive listening

Abstract

The brain's circuitry for perceiving and producing speech may show a notable level of overlap that is crucial for normal development and behavior. The extent to which sensorimotor integration plays a role in speech perception remains highly controversial however. Methodological constraints related to experimental designs and analysis methods have so far prevented the disentanglement of neural responses to acoustic versus articulatory speech features. Using a passive listening paradigm and multivariate decoding of single-trial fMRI responses to spoken syllables, we investigated brain-based generalization of articulatory features (place and manner of articulation, and voicing) beyond their acoustic (surface) form. For example, we trained a classifier to discriminate labio-dental versus alveolar places of articulation within stop syllables (e.g., /pa/ vs. /ta/) and tested whether this training generalizes to fricatives (e.g., /fa/ vs. /sa/). This novel approach revealed generalization of place and manner of articulation at multiple cortical levels, including auditory, sensorimotor and motor regions, suggesting the abstract representation of sensorimotor information. Additionally, generalization of voicing and place of articulation included the right anterior superior temporal sulcus associated with the perception of human voices. Our findings highlight the close connection between brain systems for speech perception and production, and in particular, the availability of abstract articulatory codes during passive speech perception.

Introduction

Speech perception and production are closely linked during verbal communication in everyday life. Correspondingly, the neural processes responsible for both faculties are inherently connected (Glasser & Rilling, 2008; Hickok & Poeppel, 2007), with sensorimotor integration subserving transformations between acoustic (perceptual) and articulatory (motoric) representations (Hickok, Houde, & Rong, 2011). In speech production, the role of sensorimotor integration can be behaviorally observed in motor speech development and speech control. Specifically, exposure to speech input at early age stimulates the acquisition and development of fluent speech production, and listening to the altered auditory feedback of one's own voice modulates articulatory speech gestures in adult age (Guenther & Vladusich, 2012). Furthermore, online monitoring of speech production (Postma, 2000) has been suggested to rely on both auditory-to-motor and motor-to-auditory transformations, for example as part of a state feedback control system (SFC, Hickok et al., 2011; Houde & Nagarajan, 2011).

In speech perception, the role of sensorimotor integration is less established. The likelihood of observing speech evoked activity in sensorimotor and motor regions may be influenced by the type of experimental task employed. In tasks explicitly requiring sensory-to-motor control, such as speech repetition (Caplan & Waters, 1995; Hickok, Okada, & Serences, 2009), humming (Hickok, Buchsbaum, Humphries, & Muftuler, 2003), and verbal rehearsal in working memory (Baddeley, A., Gathercole, S., & Papagno, 1998; Buchsbaum et al., 2011; Jacquemot & Scott, 2006), brain regions at the border of the posterior temporal and parietal lobes, including the sylvian parieto-temporal (Spt) region, the temporo-parietal junction (TPJ) and supramarginal gyrus (SMG), are typically implicated (Hickok & Poeppel, 2007). In contrast, a recent fMRI study using a continuous gender discrimination task showed distributed phonological representations of speech in the auditory (superior temporal) but not in the motor cortex (Arsenault & Buchsbaum, 2015). On the other hand, also in

experimental paradigms that do not explicitly require sensorimotor integration, the perception of speech is suggested to involve a co-activation of motor speech regions (Wilson, Saygin, Sereno, & Iacoboni, 2004; Zatorre, Evans, Meyer, & Gjedde, 1992). These activations may follow a topographic organization, such as when listening to syllables involving the lips (e.g. /ba/) versus the tongue (e.g. /da/) (Pulvermüller et al., 2006), and may elicit peripheral muscle activation of the respective muscle articulators (Fadiga, Craighero, Buccino, & Rizzolatti, 2002). Moreover, transcranial magnetic stimulation (TMS) applied to the lip and tongue areas of the left primary motor cortex may induce perceptual biases when listening to ambiguous spoken syllables in noisy conditions, such as along the /b/-/d/ continuum (D'Ausilio et al., 2009). Whether this co-involvement of motor speech areas in speech perception reflects an epiphenomenal effect due to an interconnected network for speech and language (Hickok, 2010), a compensatory effect observed in noisy conditions (Du, Buchsbaum, Grady, & Alain, 2014), or neural computations used for an articulatory-based segmentation of speech input in everyday life situations remains unknown (Galantucci, Fowler, & Turvey, 2006; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Pulvermüller & Fadiga, 2010).

During speech perception, neural representations of syllable identity in a manner consistent with abstract articulatory or perceptual features have been observed in the auditory cortex. Intracranial electroencephalography (iEEG) in response to all phonemes of the English alphabet has recently been shown to reflect an organization based on articulatory features of speech within the superior temporal lobe (Mesgarani, Cheung, Johnson, & Chang, 2014). Moreover, fMRI in combination with multivoxel pattern analysis (MVPA) has allowed unraveling perceptually related representations of speech in this region (Arsenault & Buchsbaum, 2015; Bonte, Hausfeld, Scharke, Valente, & Formisano, 2014; Formisano, De Martino, Bonte, & Goebel, 2008; Kilian-Hütten, Valente, Vroomen, & Formisano, 2011). In particular, MVPA permits isolating neural representations of stimulus classes from variation across other stimulus dimensions, such as the representation of vowels independent of

acoustic variation across speakers' pronunciations and vice-versa (Formisano et al., 2008) or the representation of semantic concepts independent of acoustic variation across language specific word forms in bilingual listeners (J. Correia et al., 2014). This MVPA-based generalization approach is essential to investigate the representation of articulatory features during speech perception across distinct acoustic (surface) forms. Crucially, the large spatial coverage of fMRI provides the opportunity to investigate how articulatory representations of phonemes emerge in an integrated language network beyond the superior temporal cortex, including temporal-parietal regions involved in sensorimotor integration as well as motor speech regions.

In this high-resolution fMRI study, we examined the role of articulatory representations during passive listening using a balanced set of spoken syllables in combination with multivariate decoding and a generalization approach. Our approach allowed disentangling neural responses to articulatory versus acoustic features by discriminating spoken syllables based on their articulatory features within one class of syllables (e.g. stops) and generalizing this discrimination to another class of syllables (e.g. fricatives). The use of a passive listening paradigm minimized task-specific and/or response related activations. Stimuli consisted of 24 consonant-vowel syllables constructed from eight consonants (/b/, /d/, /f/, /p/, /s/, /t/, /v/, /z/) and three vowels (/a/, /i/, /u/) forming two features per articulatory dimension (place of articulation: bilabial/labio-dental and alveolar; manner of articulation: stop and fricative; voicing: voiceless and voiced) (Figure 1A). A slow event-related design with an inter-trial interval of 12 to 16 seconds assured a maximally independent single-trial fMRI acquisition. A scanning protocol that introduced a silent period of 1 second between consecutive volume acquisitions optimized stimulus intelligibility and corresponding hemodynamic responses. Our combined MVPA generalization approach consisted of training, for example, a classifier to discriminate between two places of articulation (bilabial/labio-dental vs. alveolar) for stop consonants and testing whether this training generalizes to fricatives, i.e. decoding place of articulation independent of manner of articulation. This classification procedure based on generalization principles

allowed investigating the representation of articulatory features beyond acoustic variation across other articulatory dimensions and was performed for (1) place of articulation across manner of articulation, (2) manner of articulation across place of articulation, and (3) voicing across manner of articulation. The generalization analysis used a moving surface-based searchlight procedure (Chen et al., 2011; Kriegeskorte, Goebel, & Bandettini, 2006) that selects cortical vertices based on their geodesic proximity and a multivariate decoding procedure using support vector machines (SVM, Vapnik, 1995).

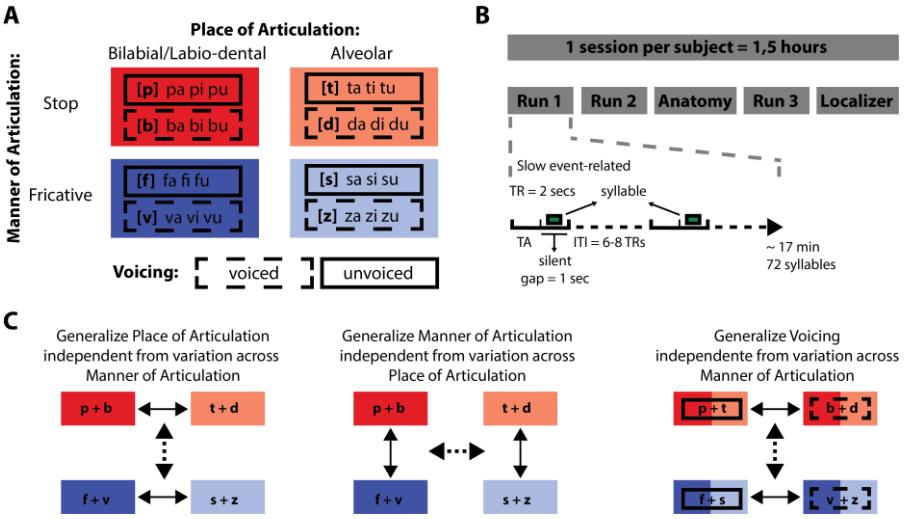


Figure 1. Experimental paradigm and analysis. A) The spoken syllable stimuli and their articulatory features. Stimuli were selected according to place of articulation (bilabial/labio-dental and alveolar), manner of articulation (stop and fricative) and voicing (voiced and unvoiced). B) The experimental session included three functional runs during which syllables were presented in a slow event-related fashion (ITI=12-16 s) during a silent gap (1 s) between consecutive volume acquisitions. C) fMRI decoding analysis based on three generalization strategies: (1) generalization of place of articulation across variation of manner of articulation; (2) generalization of manner of articulation across variation of place of articulation; (3) generalization of voicing across variation of manner of articulation, Both generalization directions were tested. In all generalization strategies all the stimuli were used.

Methods

Participants

Ten Dutch speaking participants (five males, five females; mean, SD age = 28.2, 2.35 years; 1 left handed) took part in the study. All participants were undergraduate or postgraduate students of Maastricht University, reported normal hearing abilities and were neurologically healthy. The study was approved by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht, The Netherlands.

Stimuli

Stimuli consisted of 24 consonant-vowel (CV) syllables pronounced by 3 female Dutch speakers, generating a total of 72 sounds. The syllables were constructed based on all possible CV combinations of 8 consonants (/b/, /d/, /f/, /p/, /s/, /t/, /v/ and /z/) and 3 vowels (/a/, /i/ and /u/). The eight consonants were selected to cover two articulatory features per articulatory dimension (Figure 1A): for place of articulation, bilabial/labio-dental (/b/, /p/, /f/, /v/) and alveolar (/t/, /d/, /s/, /z/); for manner of articulation, stop (/p/, /b/, /t/, /d/) and fricative (/f/, /v/, /s/, /z/); and for voicing, unvoiced (/p/, /t/, /f/, /s/) and voiced (/b/, /d/, /v/, /z/). The different vowels and the three speakers introduced acoustic variability. Stimuli were recorded in a soundproof chamber at a sampling rate of 44.1 kHz (16-bit resolution). Post-processing of the recorded stimuli was performed in PRAAT software (Boersma & Weenink, 2001) and included bandpass filtering (80-10500 Hz), manual removal of acoustic transients (clicks), length equalization, removal of sharp onsets and offsets using 30 ms ramp envelops, and amplitude equalization (average RMS). Stimulus length was equated to 340 ms using PSOLA (pitch synchronous overlap and add) with 75-400 Hz as extrema of the F0 contour. We carefully checked our stimuli for possible alterations in F0 after length equation and did not find detectable changes. The subjects reported that

the stimuli were unambiguously comprehended during a stimuli familiarization phase before the experiment.

Experimental procedures

The main experiment was divided in three slow event-related runs (Figure 1B). The runs consisted of randomly presenting each of the 72 sounds once, separated by an inter-trial-interval (ITI) of 12-16 s (corresponding to 6-8 TRs), while participants were asked to attend to the spoken syllables. Stimulus presentation was pseudorandomized such that consecutive presentations of the same syllables were avoided. Stimuli were presented binaurally at a comfortable intensity level during silent periods (1 s) between two acquisition volumes (TR=2 s, TA=1 s), using MR-compatible in-ear headphones. Participants were asked to fixate at a gray fixation cross against a black background to keep the visual stimulation constant during the entire duration of a run. Run transitions were marked with written instructions.

fMRI acquisition

Functional and anatomical image acquisition was performed on a Siemens TRIO 3 tesla scanner (Scannexus B.V.) at the Maastricht Brain Imaging Center. Functional runs used in the main experiment were collected per subject with a spatial resolution of 2 mm isotropic using a standard echo-planar imaging sequence [repetition time (TR) = 2.0 s; acquisition time (TA) = 1.0 s; field of view = 192 x 192 mm; matrix size = 64 x 64; echo time (TE) = 30 ms; multiband factor 2]. Each volume consisted of 25 slices aligned and centered along the Sylvian fissures of the participants. The duration difference between the TA and TR introduced a silent period used for the presentation of the auditory stimuli. High-resolution (voxel size 1 mm^3 isotropic) anatomical images covering the whole brain were acquired after the second functional run using a T1-weighted three-dimensional ADNI (Alzheimer's Disease Neuroimaging Initiative) sequence (TR = 2050 ms; TE = 2.6 ms; 192 sagittal slices).

Two additional localizer runs presented at the end of the experimental session were employed to identify fMRI activations related to *listening* and *repeating* the

spoken syllables and to guide the multivariate decoding analysis conducted in the main experiment. For the first localizer run, participants were instructed to attentively listen to the spoken syllables. For the second localizer run, participants were instructed to listen and repeat the spoken syllables. Both localizer runs consisted of nine blocks of eight syllables, with one syllable presented per TR. The blocks were separated by an ITI of 12.5-17.5 s (5-7 TRs). The scanning parameters were the same as used in the main experiment, with the exception of a longer TR (2.5 s) that assured that participants were able to listen and repeat the syllables in the absence of scanner noise (silent period = 1.5 s). Figure 2A-B shows the overall BOLD activation evoked during the localizer runs. Listening to the spoken syllables elicited activation in the superior temporal lobe in both hemispheres (Figure 2A), as well as in inferior frontal cortex/anterior insula in the left hemisphere. Repeating the spoken syllables additionally evoked activation in pre-motor (anterior inferior pre-central gyrus and posterior inferior frontal gyrus), motor (pre-central gyrus) and somatosensory (post-central gyrus) regions (Figure 2B). BOLD activation in these regions was statistically assessed using random-effects GLM (general linear model) statistics ($p<0.05$) and corrected for multiple comparisons using cluster-size threshold correction (alpha = 5%). Based on the activations evoked during both *listening* and *repeating* runs we defined one region of interest (ROI) per hemisphere (Figure 2C) that was used to guide the decoding analysis of the main experiment. The ROIs additionally included parts of the temporal lobes, inferior frontal cortices and parietal lobes that are typically activated in speech and language tasks (Binder, Desai, Graves, & Conant, 2009; Hickok & Poeppel, 2007; Price, 2010).

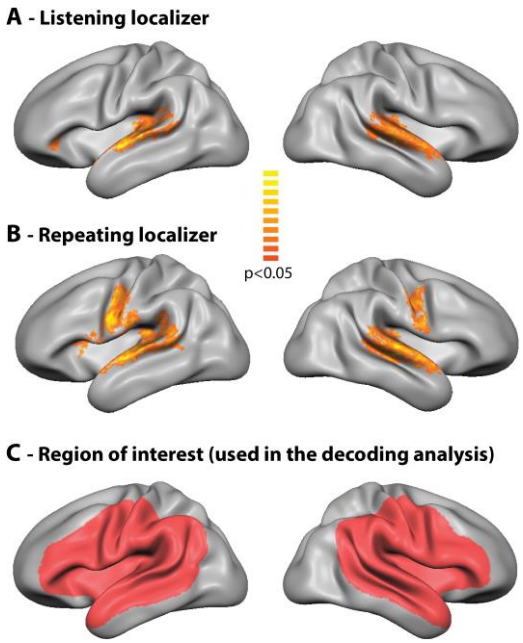


Figure 2. Group-level cortical activity maps and Regions of Interest (ROIs) based on the localizer runs. A) Passive listening. B) Verbal repetition. Group-level maps were statistically assessed using random-effects statistics (exact permutation tests = 1022) and corrected for multiple comparisons using cluster-size correction. C) A group region of interest was anatomically defined to include speech related regions that assured the inclusion of the activations elicited by the localizers.

fMRI data preprocessing

fMRI data were preprocessed and analyzed using Brain Voyager QX version 2.8 (Brain Innovation, Maastricht, The Netherlands) and custom-made MATLAB routines. Functional data were 3D motion corrected (trilinear sinc interpolation), corrected for slice scan time differences, and temporally filtered by removing frequency components of five or less cycles per time course (Goebel, Esposito, & Formisano, 2006). Anatomical data were corrected for intensity inhomogeneity and transformed into Talairach space (Talairach & Tournoux, 1988). Individual cortical surfaces were reconstructed from gray–white matter segmentations of the anatomical acquisitions and aligned using a

moving target-group average approach based on curvature information (cortex-based alignment, CBA) to obtain an anatomically aligned group-averaged 3D surface representation (Frost & Goebel, 2012; Goebel et al., 2006). Functional data were projected to the individual cortical surfaces, creating surface-based time courses. All statistical analyses were then conducted on the group averaged surface making use of CBA.

MVPA classification (generalization of articulatory features)

In order to investigate the local representation of spoken syllables based on their articulatory features we used multivariate classification in combination with a moving searchlight procedure that selected cortical vertices based on their spatial (geodesic) proximity. The particular aspect pursued by the MVPA analysis was to decode articulatory features of the syllables beyond their phoneme specific acoustic signatures (Figure 1C). Hence, we employed a classification strategy based on the generalization of articulatory features across different types of phonemes. For example, we trained a classifier to decode place of articulation features (bilabial/labio-dental vs. alveolar) from stop syllables (e.g., [/b/ and /p/] vs. [/t/ and /d/]) and tested whether this learning is transferable to fricative syllables (e.g., [/f/ and /v/] vs. [/s/ and /z/]), thus decoding place of articulation features across phonemes differing in manner of articulation. In total, we performed such a generalization strategy to investigate the neural representation of place of articulation (bilabial/labio-dental vs. alveolar) across manner of articulation (stop and fricatives); manner of articulation (stop vs. fricatives) across place of articulation (bilabial/labio-dental and alveolar); and voicing (voiced vs. unvoiced) across manner of articulation (stop and fricatives). Additional methodological steps encompassing the construction of the fMRI-feature space (fMRI-feature extraction and surface-based searchlight procedure) as well as the computational strategy to validate (cross-validation) and display (generalization maps) the classification results are described below in detail.

fMRI-feature extraction

Prior to classification, BOLD responses of each fMRI trial and each cortical vertex within the predefined left and right hemisphere ROIs were estimated by fitting a hemodynamic response function (HRF) using a general linear model (GLM). To account for the temporal variability of single-trial BOLD responses, multiple HRF fittings were produced by shifting their onset time (lag) with respect to the stimulus event time (number of lags = 21, interval between consecutive lags = 0.1 s, see also Ley et al., 2012; De Martino et al., 2008). At each trial, the GLM coefficient beta resulting from the best fitting HRF across lags in the whole brain was used to construct an fMRI-feature space composed by the number of trials by the number of cortical vertices, which was thereafter used in the multivariate decoding.

Surface-based searchlight procedure (selecting cortical vertices for classification)

To avoid degraded performances of the classification algorithm attributable to the high dimensionality of the feature space (model *overfitting*; for a description, see Norman et al., 2006), a reduction of the number of fMRI-features is usually performed. The moving *searchlight* approach (Kriegeskorte et al., 2006) restricts fMRI-features using focal selections centered at all voxels within spherical patches of the gray matter volume. Here, we used a searchlight procedure on the gray-white matter segmentation surface (Chen et al., 2011), which selected cortical vertices for decoding based on their spatial (geodesic) distance within circular surface patches with a radius of 10 mm (Figure 2). The surface-based searchlight procedure reduces the concurrent inclusion of voxels across different gyri that are geodesically distant but nearby in 3D volume space, and has been shown reliable for fMRI MVPA analyses (Chen et al., 2011; Oosterhof, Wiestler, Downing, & Diedrichsen, 2011). Crucially, the surface-based searchlight procedure assures an independent analysis of superior temporal and ventral frontal cortical regions that may be involved in the articulatory representation of speech.

Cross-validation

Cross-validation was based on the generalization of articulatory features of syllables independent of acoustic variation across other articulatory features. For each classification strategy, two cross-validation folds were created and included generalization in one and the opposite direction (e.g., generalization of place of articulation from stop to fricatives and from fricatives to stop syllables). Cross-validation based on generalization strategies is attractive since it enables detecting activation patterns resistant to variation across other stimuli dimensions (Buchweitz, Shinkareva, Mason, Mitchell, & Just, 2012; J. Correia et al., 2014; Formisano et al., 2008). In particular, generalization was performed across manner of articulation, both for place of articulation and voicing, and across place of articulation for manner of articulation.

Generalization maps

At the end of the searchlight decoding procedure, individual averaged accuracy maps for place of articulation, manner of articulation and voicing were constructed, projected onto the group-averaged cortical surface and anatomically aligned using CBA (Figure 3). To assess group-averaged statistical significance of cortical vertices (chance level is 50%), exact permutation tests were used (n = 1022). The resulting statistical maps were corrected for multiple comparisons by applying a cluster-size threshold with a false-positive rate (alpha = 5%) after setting an initial vertex-level threshold ($p < 0.05$, uncorrected) and submitting the maps to a correction criterion based on the estimate of the spatial smoothness of the map (Forman et al., 1995; Goebel et al., 2006; Hagler, Saygin, & Sereno, 2006).

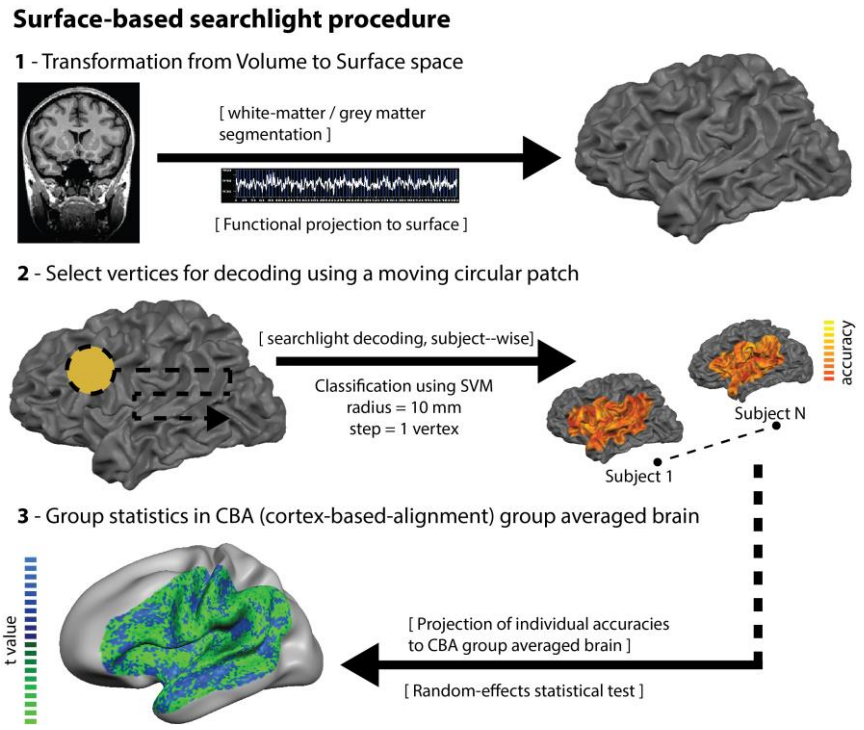


Figure 3. Surface-based searchlight procedure. 1) The first step encompassed creating individual cortical meshes on which functional responses were projected. 2) Using circular patches of 10 mm radius, the cortical mesh was analyzed multivariately using SVM classifiers. This analysis was performed within left and right hemisphere regions of interest (ROIs – see Figure 3). 3) Individual results were combined into a group-averaged cortical mesh using CBA and statistically assessed using random-effects statistics (exact permutation tests = 1022).

Results

To investigate the neural representation of articulatory features during speech perception, we employed a classification strategy that relied on generalizing the discriminability of three articulatory features across different syllable pairs. That is, using a moving searchlight procedure, we tested which cortical regions allow

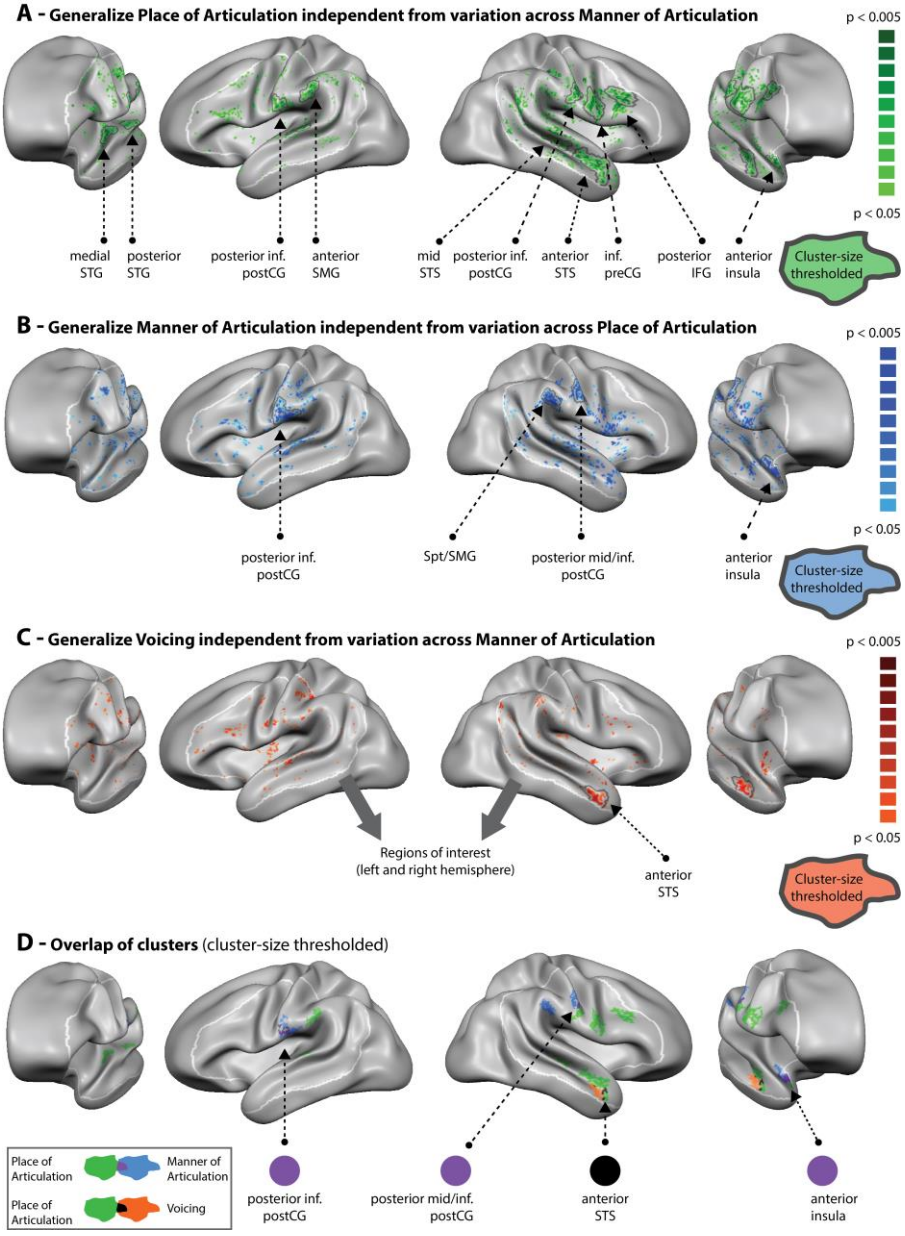
the generalization of (1) place of articulation across two distinct manners of articulation (Figure 4A), (2) manner of articulation across two distinct places of articulation (Figure 4B), and (3) voicing across two distinct manners of articulation (Figure 4C). For each type of generalization analysis, half of the available trials was used to learn to discriminate the neural responses evoked by the articulatory features of interest and the other half was used to test the generalization of this learning. The same stimuli were hence included in all generalization analyses.

Figure 4A-C depicts the resulting generalization maps for the three types of decoding with black circled clusters indicating regions that survived cluster-size multiple comparisons correction. The generalization maps reveal successful decoding of each of the three articulatory features within distinct but partially overlapping regions of the brain's speech perception network. We observed generalization foci within both the left and right hemispheres, suggesting the participation of bilateral language and speech regions in the representation of spoken syllables based on their articulatory/motoric properties.

Cortical regions enabling the generalization of place and manner of articulation were most widely distributed, including both superior temporal, pre-motor, motor and somatosensory regions bilaterally as well as sensorimotor areas at the border between the parietal and temporal lobes in the left hemisphere in case of place of articulation and in the right hemisphere in case of manner of articulation. Specific regions leading to significant generalization of place of articulation included localized clusters in the left medial and posterior superior temporal gyrus (STG), left posterior inferior post-central gyrus (CG) and left anterior SMG, as well as, right middle and anterior superior temporal sulcus (STS), right posterior inferior post-CG, inferior pre-CG and right posterior inferior frontal gyrus (IFG). Generalization of manner of articulation was significantly possible based on clusters within the left posterior inferior post-CG, right mid STS, right posterior mid/inferior post-CG, inferior anterior pre-CG/posterior IFG, right anterior SMG and right anterior insula.

In contrast to the contribution of multiple, distributed brain activity clusters to the representation of place and manner of articulation, generalization of voicing across different syllables was restricted to clusters in the right anterior STS. Finally, a comparison of the three types of generalization maps (Figure 4D) demonstrates spatially overlapping clusters of significant generalization for place and manner of articulation within the posterior inferior/mid post-CG bilaterally as well as in the right anterior insula, and for place of articulation and voicing within the right anterior STS.

Figure 4. Group-averaged generalization maps for A) place of articulation. B) manner of articulation and C) voicing as well as D) the spatial overlap of these three maps after cluster-size correction. All maps are projected onto the group-aligned cortical surface and statistically thresholded ($p < 0.05$). Regions surviving cluster-size threshold (> 18 sqmm) used for multiple comparison correction are depicted by a black line. The left and right hemisphere regions of interest are depicted by a white line. Purple colored vertices depict the overlap between regions generalizing place of articulation and manner of articulation. Black colored vertices depict the overlap between regions generalizing place of articulation and voicing. No overlap was found between manner of articulation and voicing.


A - Generalize Place of Articulation independent from variation across Manner of Articulation
p < 0.005
p < 0.05
medial STG
posterior STG
posterior inf. postCG
anterior SMG
mid STS
posterior inf. postCG
anterior STS
inf. preCG
posterior IFG
anterior insula
Cluster-size thresholded
B - Generalize Manner of Articulation independent from variation across Place of Articulation
p < 0.005
p < 0.05
posterior inf. postCG
Spt/SMG
posterior mid/inf. postCG
anterior insula
Cluster-size thresholded
C - Generalize Voicing independent from variation across Manner of Articulation
p < 0.005
p < 0.05
Regions of interest (left and right hemisphere)
anterior STS
Cluster-size thresholded
D - Overlap of clusters (cluster-size thresholded)
Place of Articulation
Manner of Articulation
Place of Articulation
Voicing
posterior inf. postCG
posterior mid/inf. postCG
anterior STS
anterior insula

Discussion

The present study aimed to investigate whether focal patterns of fMRI responses to speech input contain information regarding articulatory features when participants are attentively listening to spoken syllables in the absence of task demands that direct their attention to speech production or monitoring. Using high spatial resolution fMRI in combination with an MVPA generalization approach we were able to identify specific foci of brain activity that discriminate articulatory features of spoken syllables independent of their acoustic variation (surface form) across other articulatory dimensions. These results provide compelling evidence for interlinked brain circuitry of speech perception and production, and in particular, for the availability of abstract articulatory codes during on-line perception of spoken syllables within pre-motor and motor, somatosensory, auditory and/or sensorimotor integration areas.

Our generalization analysis suggests the involvement of pre-motor and motor areas in the neural representation of two important articulatory features during passive speech perception: manner and place of articulation. These findings are compelling as the role of pre-motor and motor areas during speech perception remains controversial (Liberman and Mattingly, 1985; Galantucci et al., 2006; Pulvermüller and Fadiga, 2010). Left hemispheric motor speech areas have been found to be involved during both the sub-vocal rehearsal and perception of the syllables /ba/ and /da/ (Pulvermüller et al., 2006) and of spoken words (Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2014). Furthermore, speech motor regions may bias the perception of ambiguous speech syllables under noisy conditions (D'Ausilio et al., 2009; Du et al., 2014) and have been suggested to be specifically important for the performance of tasks requiring subvocal rehearsal (Hickok & Poeppel, 2007; Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013). However, the involvement of (pre)motor cortex in speech perception may also reflect an epiphenomenal consequence of interconnected networks for speech perception and production (Hickok, 2010).

Importantly, the observed decoding and generalization capacity of activation patterns in (pre)motor regions for place across variation in manner and manner across variation in place of articulation indicate the existence of articulatory information representation that extends beyond mere activation spread also while passively listening to clear spoken syllables. Further investigations exploiting correlations of articulatory MVPA representations to behavioral measures of speech perception and their modulation by task difficulty (Raizada & Poldrack, 2007) may permit mapping aspects related to speech intelligibility and may lead to a further understanding of the functional relevance of such articulatory representations.

Our results also show articulatory representations of the spoken syllables in bilateral somatosensory areas. In particular, areas comprising the inferior posterior banks of the post-central gyri were sensitive to the generalization of place and manner of articulation. Somatosensory and motoric regions are intrinsically connected and proposed to have identical topographic-somatotopic representations for speech gestures (Bouchard, Mesgarani, Johnson, & Chang, 2013). Together with feedback from auditory cortex, somatosensory feedback during speech production is postulated to be part of a state feedback control system (SFC, Hickok et al., 2011; Houde and Nagarajan, 2011). The involvement of these somatosensory areas in the representation of articulatory features during passive speech perception extends recent findings showing the involvement of these regions in the neural decoding of place of articulation during an active perceptual task in English listeners (Arsenault & Buchsbaum, 2015). In particular, they may indicate automatic information transfer from auditory to somatosensory representations during speech perception similar to their integration as part of SFC systems for speech production.

Especially in auditory cortex, it is essential to disentangle brain activity indicative of articulatory versus acoustic features. So far, methodological constraints related to experimental designs and analysis methods have often prevented this differentiation. In a recent study, intracranical EEG (iEEG) responses to English phonemes, indicated a phonetic organization in the left superior temporal cortex, especially in terms of manner of articulation

(Mesgarani et al., 2014). Our cortical generalization maps not only show that it was possible to predict manner of articulation based on activation patterns in bilateral auditory areas associated with speech perception (STG), but also to predict place of articulation in these regions as well as voicing based on patterns within the right anterior superior temporal lobe (ATL/STS). Voicing representations within the right anterior STS may further relate to their involvement in the processing of human voices (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), and possibly to the proposed specialization of this area in perceiving vocal-tract properties of speakers (e.g., shape and characteristics of the vocal folds). A role of this phonological feature in the processing of human voices is also compatible with the previous finding that voicing was more robustly classified than either place or manner of articulation when subjects performed an active gender discrimination task (Arsenault & Buchsbaum, 2015). Because, in contrast to manner and place of articulation, voicing was not significantly classified outside the superior temporal cortex, in future studies using spoken words (Schomers et al., 2014) and/or sentences, it would be interesting to investigate whether under these more natural conditions, sensorimotor speech representations also extend to voicing.

Articulatory representations that may speak for sensorimotor integration mechanisms, thus possibly involving the translation between auditory and articulatory codes, included regions within the inferior parietal lobes (IPL). Specifically, the anterior supramarginal gyus (SMG) was found to generalize place of articulation in the left hemisphere and manner of articulation in the right hemisphere. Nearby regions involving the inferior parietal lobe (Kilian-Hütten et al., 2011; Moser, Baker, Sanchez, Rorden, & Fridriksson, 2009; Raizada & Poldrack, 2007) and sylvian-parietal-temporal (Spt, Buchsbaum et al., 2011; Caplan and Waters, 1995; Hickok et al., 2003, 2009) have been implicated in sensorimotor integration during speech perception as well as in mapping auditory targets of speech sounds prior the initiation of speech production (Guenther & Vladusich, 2012; Hickok et al., 2009). Here, we show the sensitivity of SMG to represent articulatory features of spoken syllables during speech perception in the absence of an explicit task triggering an action based

on perception such as speech repetition (Caplan & Waters, 1995; Hickok et al., 2009) or music humming (Hickok et al., 2003). Furthermore, our results suggest the involvement of IPL regions in the perception of clear speech, extending previous findings showing a significant role in the perception of ambiguous spoken syllables (Raizada & Poldrack, 2007) and the integration of ambiguous spoken syllables with lip-read speech (Kilian-Hütten et al., 2011). Also involved in sensorimotor integration mechanisms (Hickok & Poeppel, 2007), the left inferior frontal gyrus (IFG) showed decoding consistency across the three articulatory generalizations, although the individual clusters were found not significant in terms of cortical spread. Besides being involved in syntactic and semantic processing (Hagoort, 2005), the left IFG is implicated in phonetic discrimination (Zatorre et al., 1992) and its neural connections in the speech network make the IFG a central component for normal and fluent speech functions (Aboitiz, 2012; Frey, Mackey, & Petrides, 2014; Glasser & Rilling, 2008; Parker Jones et al., 2014).

Overall, our generalization strategy was crucial to disentangle the neural representation of articulatory and acoustic differences between spoken syllables during passive speech perception. In particular, this generalization strategy allowed mapping the representation of bilabial/labio-dental versus alveolar, stop versus fricative and voiced versus unvoiced articulatory features in cortical regions relevant for language processing. Similar generalization strategies capable to transfer representation patterns across different stimulus classes have been adopted in MVPA studies, for example in isolating the identity of vowels and speakers independent of acoustic variation (Formisano et al., 2008), in isolating concepts independent of language presentation in bilinguals (Buchweitz et al., 2012; J. M. Correia, Jansma, Hausfeld, Kikkert, & Bonte, 2015; J. Correia et al., 2014) and in isolating concepts independent of presentation modality (Shinkareva, Malave, Mason, Mitchell, & Just, 2011; Simanova, Hagoort, Oostenveld, & van Gerven, 2014). Together, these findings suggest that the neural representation of language consists of specialized sub-networks that tune to certain feature characteristics independent of other features within the signal. Crucially, our findings provide evidence for the

interaction of auditory, sensorimotor and somatosensory brain circuitries during speech perception, in conformity with the behavioral link between perception and production faculties in everyday life. The applications of fMRI decoding and generalization methods also hold promise to investigate similarities of acoustic and articulatory speech representations during the perception and production of speech (Christoffels, van de Ven, Waldorp, Formisano, & Schiller, 2011).

References

Aboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Frontiers in Evolutionary Neuroscience*, *4*(February), 2. doi:10.3389/fnevo.2012.00002

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed Neural Representations of Phonological Features during Speech Perception. *Journal of Neuroscience*, *35*(2), 634–642. doi:10.1523/JNEUROSCI.2454-14.2015

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, *105*(1), 158–173.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–12. doi:10.1038/35002078

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 (New York, N.Y. : 1991)*, *19*(12), 2767–96. doi:10.1093/cercor/bhp055

Boersma, P., & Weenink, D. (2001). Praat: doing phonetics by computer. *Glot Int.*, *5*(9), 341–347. Retrieved from http://www.praat.org/

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(13), 4548–57. doi:10.1523/JNEUROSCI.4339-13.2014

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–32. doi:10.1038/nature11911

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–28. doi:10.1016/j.bandl.2010.12.001

Buchweitz, A., Shinkareva, S. V, Mason, R. a, Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language*, *120*(3), 282–9. doi:10.1016/j.bandl.2011.09.003

Caplan, D., & Waters, G. S. (1995). On the nature of the phonological output planning processes involved in verbal rehearsal: evidence from aphasia. *Brain and Language*, *48*(2), 191–220. doi:10.1006/brln.1995.1009

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage*, *56*(2), 582–92. doi:10.1016/j.neuroimage.2010.07.035

Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PloS One*, *6*(5), e18307. doi:10.1371/journal.pone.0018307

Correia, J. M., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(1), 332–8. doi:10.1523/JNEUROSCI.1302-13.2014

Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., & Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Frontiers in Psychology*, *6*(February), 1–10. doi:10.3389/fpsyg.2015.00071

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology : CB*, *19*(5), 381–5. doi:10.1016/j.cub.2009.01.017

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58. doi:10.1016/j.neuroimage.2008.06.037

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(19), 7126–31. doi:10.1073/pnas.1318738111

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *The European Journal of Neuroscience*, *15*(2), 399–402. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11849307

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine : Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *33*(5), 636–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7596267

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science (New York, N.Y.)*, *322*(5903), 970–3. doi:10.1126/science.1164318

Frey, S., Mackey, S., & Petrides, M. (2014). Cortico-cortical connections of areas 44 and 45B in the macaque monkey. *Brain and Language*, *131*, 36–55. doi:10.1016/j.bandl.2013.05.005

Frost, M. a, & Goebel, R. (2012). Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369–81. doi:10.1016/j.neuroimage.2011.08.035

Galantucci, B., Fowler, C. a, & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–77. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2746041&tool=pmcentrez&rendertype=abstract

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(11), 2471–82. doi:10.1093/cercor/bhn011

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401. doi:10.1002/hbm.20249

Guenther, F. H., & Vladusich, T. (2012). A Neural Theory of Speech Acquisition and Production. *Journal of Neurolinguistics*, *25*(5), 408–422. doi:10.1016/j.jneuroling.2009.08.006

Hagler, D. J., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, *33*(4), 1093–103. doi:10.1016/j.neuroimage.2006.07.036

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–23. doi:10.1016/j.tics.2005.07.004

Hickok, G. (2010). The role of mirror neurons in speech and language processing. *Brain and Language*, *112*(1), 1–2. doi:10.1016/j.bandl.2009.10.006

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–82. doi:10.1162/089892903322307393

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22. doi:10.1016/j.neuron.2011.01.019

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*(5), 2725–32. doi:10.1152/jn.91099.2008

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402. doi:10.1038/nrn2113

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*(October), 82. doi:10.3389/fnhum.2011.00082

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, *10*(11), 480–6. doi:10.1016/j.tics.2006.09.002

Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(5), 1715–20. doi:10.1523/JNEUROSCI.4572-10.2011

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, *25*(12), 2179–88. doi:10.1162/jocn_a_00463

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8. doi:10.1073/pnas.0600244103

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(38), 13273–80. doi:10.1523/JNEUROSCI.0584-12.2012

Liberman, a M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4075760

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology : CB*, *17*(19), 1692–6. doi:10.1016/j.cub.2007.08.064

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science (New York, N.Y.)*, *343*(6174), 1006–10. doi:10.1126/science.1245994

Moser, D., Baker, J. M., Sanchez, C. E., Rorden, C., & Fridriksson, J. (2009). Temporal order processing of syllables in the left parietal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(40), 12568–73. doi:10.1523/JNEUROSCI.5934-08.2009

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), 424–30. doi:10.1016/j.tics.2006.07.005

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage*, *56*(2), 593–600. doi:10.1016/j.neuroimage.2010.04.270

Parker Jones, ʻŌiwi, Prejawa, S., Hope, T. M. H., Oberhuber, M., Seghier, M. L., Leff, A. P., … Price, C. J. (2014). Sensory-to-motor integration during auditory

repetition: a combined fMRI and lesion study. *Frontiers in Human Neuroscience*, *8*(January), 24. doi:10.3389/fnhum.2014.00024

Postma, a. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*(2), 97–132. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10986364

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88. doi:10.1111/j.1749-6632.2010.05444.x

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*(5), 351–60. doi:10.1038/nrn2811

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–70. doi:10.1073/pnas.0509989103

Raizada, R. D. S., & Poldrack, R. a. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726–40. doi:10.1016/j.neuron.2007.11.001

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2014). Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. *Cerebral Cortex (New York, N.Y. : 1991)*. doi:10.1093/cercor/bhu274

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), 2418–25. doi:10.1016/j.neuroimage.2010.10.042

Simanova, I., Hagoort, P., Oostenveld, R., & van Gerven, M. a J. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex (New York, N.Y. : 1991)*, *24*(2), 426–34. doi:10.1093/cercor/bhs324

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. *Stuttgart: G. Thieme.*

Vapnik, V. (1995). *The nature of statistical learning theory* (2nd ed.). Springer.

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), 701–2. doi:10.1038/nn1263

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science (New York, N.Y.)*, *256*(5058), 846–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1589767

CHAPTER 5

Information-based connectivity – IBC: a functional connectivity approach for fMRI using multivariate classification

Manuscript in preparation

Correia JM, Bonte M, Hausfeld L, Jansma BM, Valente G

Information-based connectivity – IBC: a functional connectivity approach for fMRI using multivariate classification

Abstract

Functional connectivity is essential to higher-order cognitive processes encompassing neural communication in connected brain networks. In fMRI (functional magnetic resonance imaging), functional connectivity is typically investigated using measures of temporal similarities of signal activations across brain regions. Beyond co-variation in activity strength, the concept of functional connectivity may be extended to include measures of information-sharing among distinct regions involved in distributed neural representation of perceptual or cognitive states. MVPA (multivoxel pattern analysis) permits single trial predictions of experimental conditions or stimuli based on multivariate patterns of brain responses. Combining functional connectivity with the benefits of MVPA is a current challenge within neuroimaging research that promises to find information- instead of activation-based communication between brain regions. Here, we propose a novel functional connectivity method for fMRI and neuroimaging that exploits information representation of single-trials as uncovered by MVPA – IBC (information-based connectivity). IBC investigates mutual information representation abstract to activity correlations by assessing consistencies of binary single-trial predictions across brain regions. Using fMRI simulations we investigate inter-regional signal properties underlying the principles of IBC. Simulation results show that IBC is affected by information fluctuations across trials, but not by activity fluctuations. Furthermore, we validate IBC using fMRI data from a previous MVPA speech perception experiment that highlighted the role of distributed brain regions in the representation of articulatory features during the perception of spoken syllables. The IBC method in such fMRI investigation allowed finding communication within sub-networks of speech and language that may underlie transfer of articulatory information during speech perception, in particular between auditory/sensorimotor and motor regions. Overall, IBC shows the ability to find modulation of information availability across regions, thereby providing a new measure of information sharing within neural networks supporting neural processing and cognition.

Introduction

Over the past decade functional connectivity and multivariate statistics have significantly contributed to advance brain research. In fMRI (functional magnetic resonance imaging), MVPA (multivoxel pattern analysis) has allowed uncovering spatial patterns of fMRI responses related to several aspects of neural processing and cognition that would have otherwise remained unknown when using traditional univariate analysis (Haxby et al., 2001; Formisano et al., 2008b; Mitchell et al., 2008). Regional MVPA methods allow assessing brain-states representation within independent brain regions. For example, the searchlight method (Kriegeskorte et al., 2006) uses a moving focal multivoxel selection approach in combination with MVPA to map information content on the cortex (Kriegeskorte et al., 2006). How distributed informational patterns are connected during neural processing is however not addressed in MVPA methods. Functional connectivity approaches investigate temporal signal correlations between brain regions (Friston, 2011), but are nonetheless mute to information representation unraveled by MVPA. Investigating the transfer of information representation across independent cortical regions, while exploiting technical advances brought by MVPA, is essential to understand the neural circuitry responsible for neural processing and cognition.

Functional connectivity in fMRI includes a broad range of methods based on the analysis of the time-series of BOLD (blood oxygen level dependency) responses across brain areas. Correlational methods allow finding temporal similarities between BOLD time-series (Friston et al., 1997). Auto-regressive methods, such as Granger causality (Roebroeck et al., 2005) or coherence and partial coherence analyses (Sun et al., 2004) additionally provide a measure of causality or effective connectivity. Furthermore, hypothesis driven methods, such as DCM (dynamic causal modeling; Friston et al., 2003) model fMRI signals in an interconnected network as designed deterministic input-output systems. In addition to activity-based connectivity approaches, exploiting the sensitivity of MVPA has allowed identifying inter-regional temporal

correlations emerging from spatial informational patterns across time-points (Coutanche and Thompson-Schill, 2013). Overall, these methods are dependent on the temporal dynamics of the fMRI time-series, which introduce technical limitations related to the slow hemodynamics of neural activity and low temporal resolution of fMRI (Eichler, 2005), and related to spatial modulations of neurovascular tissue properties (Vigneau-Roy et al., 2014). Less dependent on temporal fMRI dynamics however, alternative methods summarize the BOLD time-series to single-trial estimates (Rissman et al., 2004, Waldorp et al., 2011), and may provide an advantage to assess functional connectivity in event-related fMRI designs. Here, we propose and investigate a new method to assess inter-regional functional connectivity in experimental designs that exploits MVPA to find neural representations of cognitive states in slow event-related fMRI. This method is referred as IBC (information-based connectivity) and measures information-sharing by assessing consistencies of binary single-trial classification predictions across brain regions.

In MVPA, the decoding performance of multivariate patterns is typically reported as averaged accuracies across multiple single-trial classification predictions. In IBC, single-trial predictions are used to investigate inter-regional connectivity at the level of information representation. Accordingly, brain areas whose decoding predictions co-vary are assumed to be modulated by common physiological states related to the availability of information representation across single-trials. Conversely, in activity-based connectivity approaches, to uncover such information covariance may be more difficult to attain due to contamination of other sources of signal covariance, such as noise or activity fluctuations independent of the information of interest. The goal of IBC is thus to assess functional connectivity related to information representation of brain-states while abstracting from more general and uninformative signal covariance. We investigate the principles of IBC using single-trial response simulations across multivariate patterns within spatially non-overlapping regions. Furthermore, we apply and validate the IBC method on real data of an empirical fMRI speech perception study of spoken syllable decoding.

Methods

Information-based connectivity (IBC)

IBC captures inter-trial consistency of binary information decoding across two brain regions. When a testing trial produces the same MVPA decoding prediction in two regions, this is counted as a match in IBC. The total number of matches indicates a measure of inter-regional decoding consistency (IBC_{metric}) of two regions (e.g., *A* and *B)* across trials (*i)* from a total of *N* trials.

(1) $$IBC_{metric} = N - \sum_{i}^{N} XOR(A_i, B_i)$$ where,

$$XOR(Ai, Bi) = A.\overline{B} + \overline{A}.B$$

IBC_{metric} is however not an absolute measure, as the total number of possible matches is constrained by the decoding success among both regions. To assess the effect size of IBC, we therefore compare the observed IBC_{metric} with its distribution under the null hypothesis (H0) that there is no decoding consistency between regions. In a single subject setting, it is possible to estimate non-parametrically the null distribution of IBC_{metric} values, assuming exchangeability of the trials (under the null hypothesis) and permuting the trials (exhaustively or with a few thousands random permutations) in one of the two regions. In group studies, a second level random-effects analysis is instead performed, comparing the observed IBC_{metric} with its expected value, under H0, across all subjects. An analytical expression for the expected value is provided in equation 2, and is based on the expected overlap from *N* draws from two independent Bernoulli random variables (e.g., *A* and *B*).

(2) $$IBC_{chancelevel} = \frac{(N - sum(A)).(N - sum(B)) + (sum(A).sum(B)}{N}$$

Where, $$sum(A) = \sum_{i}^{N} A_i$$ and $$sum(B) = \sum_{i}^{N} B_i$$

(3) $$IBC = IBC_{metric} - IBC_{chanceleve}$$

The IBC principles, *IBC*, IBC_{metric} and $IBC_{chance\text{-}level}$ calculations are illustrated in Figure 1.

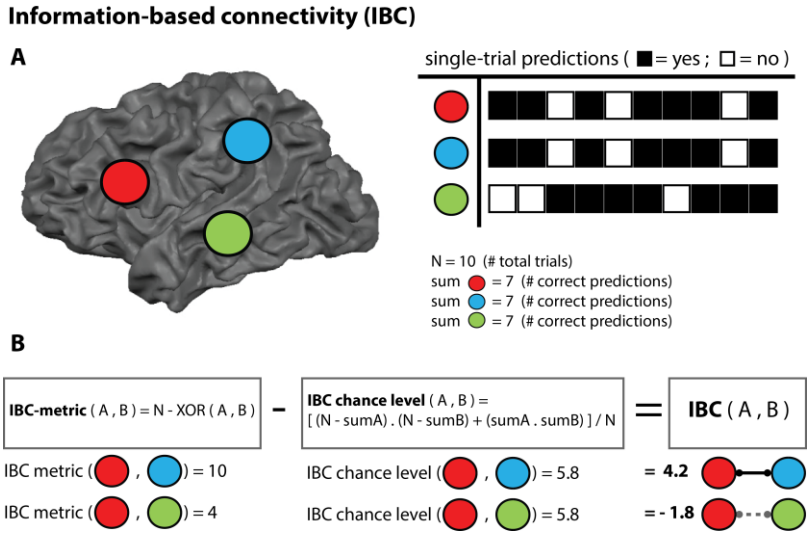


Figure 1. Principles of information-based connectivity (IBC) illustrated using three independent brain regions (red, blue and green). IBC connectivity is simulated between the red and blue regions , but not between the red and green regions. A) Given a decoding analysis that uncovers information representation, the consistency of single-trial predictions across the regions is treated as indication for shared information representation. B) Calculation of the IBC_{metric} and $IBC_{chance\text{-}level}$. Computations between red and blue regions suggest existence of IBC since the IBC_{metric} is larger than $IBC_{chance\text{-}level}$. Computations between the red and green regions do not suggest existence of IBC since the IBC_{metric} is smaller than $IBC_{chance\text{-}level}$. XOR is the exclusive logical OR (XOR(0,0)=0; XOR(1,1)=0; XOR(1,0)=1; XOR(0,1)=1).

Simulations

To evaluate the principles underlying IBC as a measure of inter-regional functional connectivity for fMRI data, we simulated single-trial voxel estimates within three independent regions (A, B and C). A single-trial estimate refers to a measure that summarizes the BOLD time-series of voxels within single-trials

by means of regression of a hemodynamic response function (HRF) (Ley et al., 2012; De Martino et al., 2008) or by means of averaging within a time-window (Formisano et al., 2008b; Correia et al., 2014). The three regions contained five voxels each and simulated binary discriminative patterns between two classes (Figure 2). IBC connectivity was manipulated between region A and B, but not between region A and C. Hence, region A was used as a seed, region B as a target to assess IBC, and region C as a control to the target region. Within each region, single-trial voxel estimates were simulated using three independently manipulated parameters: 1) the signal contrast between classes, which directly affected the decoding performance within each of the simulated regions; 2) the covariance of information representation between regions, which simulated inter-trial information fluctuations; and 3) the noise covariance between regions, which added noise fluctuations independent of information representation.

(4)

$$data = (information \times information\ fluctuations) + noise\ fluctuations$$

Construction of simulated data

The signal within each voxel was constructed as a combination of information, information fluctuation and noise fluctuation (Figure 2A and Equation 4). Different levels of information contrast within each region were simulated, producing several CNR (contrast-to-noise ratio) values reflecting the contrast between the patterns of the classes and the standard deviation of the introduced noise in the signals (SD =1). From no information (CNR=0) to high information (CNR=1), we simulated six CNR values for the three regions (CNR = [0, .2, .4, .6, .8, 1.0]). Next, information fluctuation was introduced to simulate an inter-trial information modulation that weighted the representation patterns generated in the information layer. Information fluctuation across the three regions dictated the consistency of decoding predictions assessed by the IBC_{metric}. To simulate information fluctuation between the seed and the target regions, we simulated six covariance parameters ranging from no information

covariance (=0) to high information covariance (=.5) in steps of .1 (information covariance = [0, .1, .2, .3, .4, .5]). Information covariance between the target and control regions was kept at zero for all simulations. Within regions, information covariance was maintained constant at maximum level (=1) since within region effects were not under investigation. Finally, information fluctuation for each voxel was generated using a Gaussian distribution (mean=0, SD=1) bound to the different simulated information covariance matrices (Figure 2C).

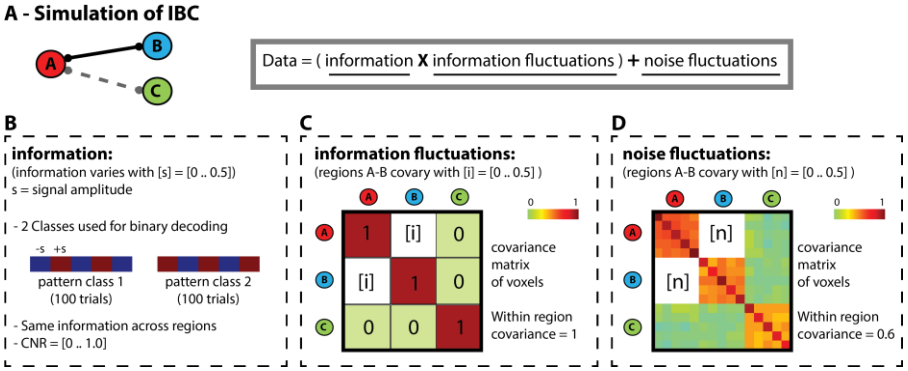


Figure 2. Simulation of single-trial fMRI responses in the three regions (A, B and C). A) Region A and B are simulated to share information representation across trials; region A and C not. Data is generated as a linear combination of three signal parts: information; information fluctuations; and noise fluctuations. B) The information layer consists of a multivariate pattern of 5 voxels discriminative of class 1 and class 2. In the generation of the simulated signals (simulation 1), the contrast of the information signal varied from 0 (no information, CNR=0) to 0.5 (high information, CNR=1). C) The information fluctuations introduced a trial-to-trial covariance of information representation across the three regions. D) The noise fluctuations introduced additional variation to the individual voxels. Within-region covariance of the noise was kept at 0.6 for all regions. In total, 216 signals (6 information levels by 6 information fluctuation covariance levels by 6 noise fluctuation covariance levels) were simulated and tested.

In order to introduce noise fluctuation independent of information fluctuations or information representation, we added Gaussian noise (mean=0; SD=1) to the simulated signals (Figure 2D). The covariance of the noise

fluctuation across voxels and regions was used as a parameter to assess how vulnerable the IBC method is to inter-regional signal correlations that do not modulate inter-trial information content. Similarly to information fluctuations, noise covariance was parametrically modulated between the seed and target regions (0, .1, .2, .3, .4, and .5), but not between the seed and control regions (covariance=0). Within-region covariance was kept constant at .6 in order to simulate signals with realistic spatial correlations commonly observed in fMRI recordings.

Furthermore, two simulation versions were used: simulation 1 in which CNR, thus information content, was manipulated in the same way across the three regions; and simulation 2 in which CNR was kept constant for the seed region at a representative value ensuring information decoding (CNR=.5; leading to averaged accuracies between .75 and .8), while CNR manipulation for the regions B and C varied parametrically. For each simulation version, a total of 100 simulations were tested and the difference between the IBC_{metric} and $IBC_{chance\text{-}level}$ were averaged. Differences in the connectivity between the seed and target region versus the seed and control region reflect the parametric manipulations that aimed to investigate the principles underlying IBC as a functional information-based connectivity method exploiting MVPA at the level of single-trial predictions. The classification algorithm employed was SVM (support vector machines, Cortes and Vapnik, 1995), although other binary classification algorithms are suitable for IBC as well.

Statistics on simulated data

We statistically assessed the effect of IBC in the connections of interest as a modulation of information fluctuation covariance across CNR and noise fluctuation covariance (as reported in average values in figures 5 and 6). Using a linear regression of the observed IBC values across the information covariance levels used in the simulations, we statistically tested differences between the regression coefficients underlying the connectivity between seed and target regions versus seed and control regions (paired t-tests, one-sided). These tests illustrate the modulatory effect of information fluctuation covariance across

the different conditions of CNR and noise fluctuation covariance (Figure 7). Steeper slopes for the modulatory effect of information fluctuation covariance between the seed and target region in comparison to the seed and control region reflect the success of IBC across different conditions of information availability and noise covariance.

fMRI data validation

To evaluate the potential of IBC empirically we applied our method to the fMRI experiment reported in Chapter 4. Here, fMRI responses to spoken syllables were analyzed using surface-based searchlight-MVPA (Kriegeskorte et al., 2006; Chen et al., 2011) to investigate articulatory feature representation of spoken syllables during speech perception. Statistically significant regions (cluster-size corrected at $p<0.05$) resulting from the cortical mapping of multivariate classification were used as seeds to compute IBC connectivity maps. IBC employed to this data set aimed to investigate speech networks related to MVPA information transfer involved in sensorimotor integration processes (Hickok and Poeppel 2007; Pulvermuller and Fadiga, 2010; Hickok, 2011; Guenther and Vladusich , 2012; Buchsbaum et al., 2011; Galantucci et al., 2006; Meister et al., 2007).

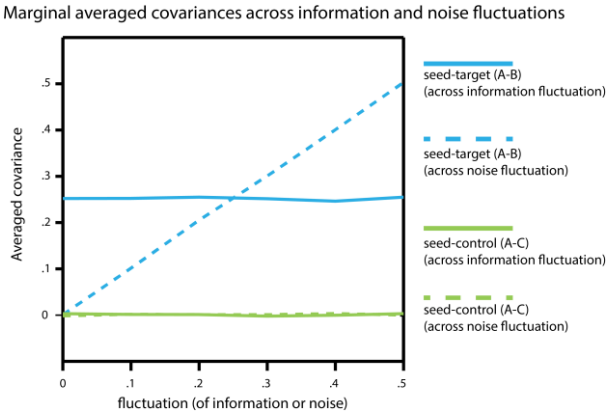


Figure 3. Averaged covariance between the seed region A and target region B (blue lines), and the seed region and the control region C (green lines) across the information fluctuation dimension (continuous lines) and across the noise fluctuation dimension (discontinuous lines). The non-averaged covariance for all simulated variations are depicted in figure S1 of the supplementary material.

fMRI experiment

The fMRI experiment aimed to investigate the transformation of acoustic to articulatory features of spoken syllables during speech perception. Articulatory features relate to vocal tract gestures necessary in the execution of speech. We used a total of 24 syllables that were grouped based on articulatory features across three articulatory dimensions. The dimensions were place of articulation, manner of articulation and voicing. An MVPA based generalization strategy allowed investigating the representation of articulatory features beyond acoustic variation across other articulatory dimensions and was performed for (1) place of articulation across manner of articulation, (2) manner of articulation across place of articulation, and (3) voicing across manner of articulation. Overall, the generalization analyses successfully isolated articulatory features of speech from acoustic variation. In this chapter we apply IBC to analyze functional connectivity in the representation of one of the articulatory features, place of articulation. For simplicity of results, we further restricted the IBC analysis to regions of the left hemisphere that showed significant generalization, and

included the medial and the posterior STG (superior temporal gyrus), the posterior post-central gyrus and anterior SMG (supramarginal gyrus). Importantly, the searchlight method generated single-trial predictions across the cortical surface. These accuracy maps are used here to calculate IBC maps based on seed regions previously identified by the generalization analyses (Figure 8A).

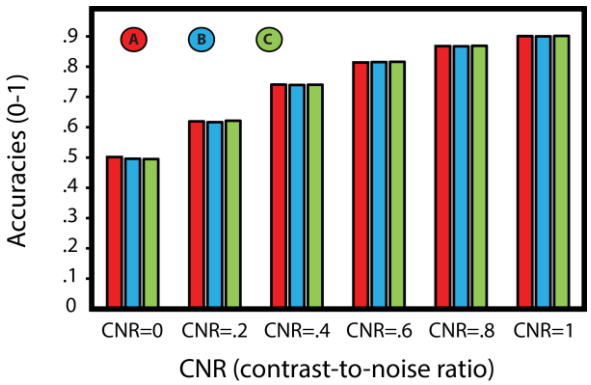


Figure 4. Averaged accuracies for the three simulated regions (A, B and C), for the different CNR levels. One combination of information and noise covariance is showed, since accuracies only depend on CNR level.

IBC maps of fMRI experiment

One connectivity map per seed region was computed. For each cortical vertex we calculated the IBC_{metric} and $IBC_{chance\text{-}level}$ in reference to vertices within the seed region. The average IBC across all vertices of the seed regions and across cross-validation folds were used as individual connectivity values. The maps reflect group level statistics of IBC_{metric} values in comparison to $IBC_{chance\text{-}level}$ values using a one-sided non-parametric Wilcoxon's test, with type I error rate alpha=5%, after setting an initial vertex-level threshold (p<0.05, uncorrected) and submitting the maps to a correction criterion based on the estimate of the spatial smoothness of the maps (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006).

Time-series maps of fMRI experiment

To compare the produced IBC maps with a functional connectivity approach based on activity correlation, we calculated the average of the time-series within seed regions as predictors in a GLM (General Linear Model) model. For each seed-region, time-series connectivity maps were statistically assessed at the group level, threshold ($p<0.05$, uncorrected) and submitted to a correction criterion based on the estimate of the spatial smoothness of the maps (Forman et al., 1995; Goebel et al., 2006; Hagler et al., 2006).

Results

Results of IBC on simulated data

We simulated multiples signal versions for the three regions using parametric modulations of information contrast (CNR), information fluctuation covariance and noise fluctuation covariance. For each signal, we calculated the temporal covariance between regions, the classification accuracy and the IBC. Temporal covariance between regions is directly related to activity-based functional connectivity. We observed a modulation of inter-regional signal covariance due to the increase of noise covariance, but not due to the increase of information fluctuation covariance (Figure 3). The full illustration of signal covariance between the three regions across information and noise modulations can be found in the supplementary material (Figure S1). Classification accuracies were directly modulated by CNR within each region (Figure 4). In *simulation-1* (Figure 5), no modulation of information fluctuation covariance or noise covariance existed on the classification accuracy within the regions.

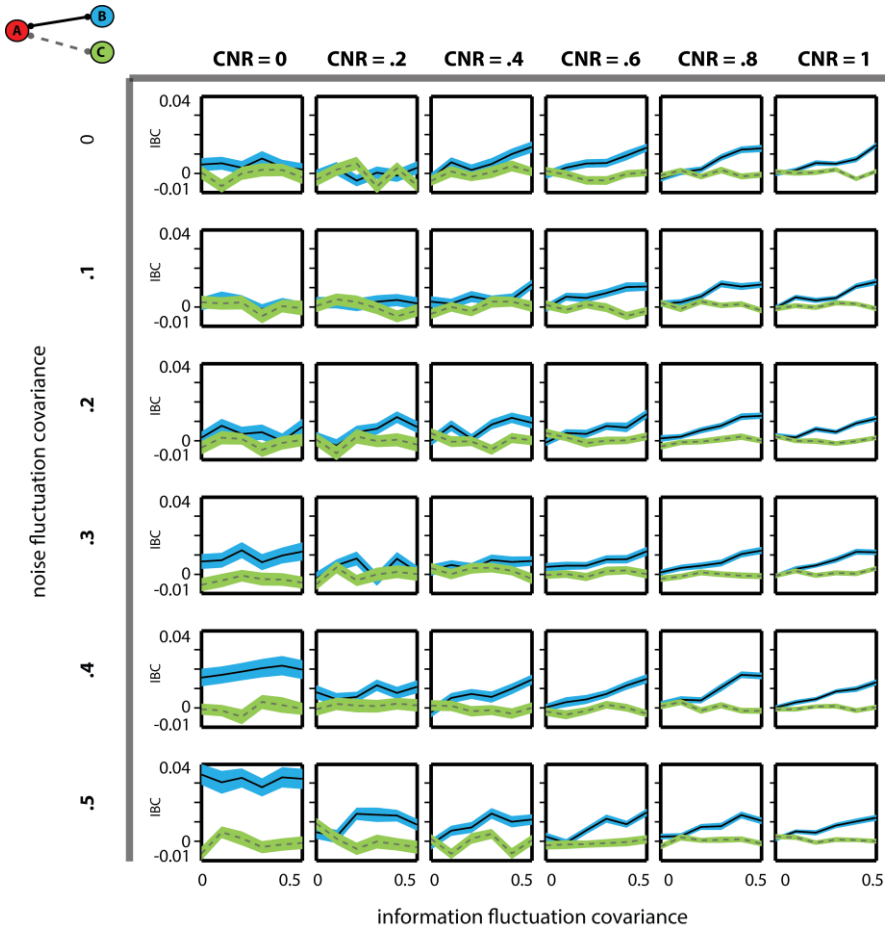


Figure 5. Results of IBC for *simulation-1* (variable CNR across regions A, B and C). Rows are different levels of non-task related noise. Columns are different levels of CNR for all regions. Within each result chart, the continuous black line with blue shaded area (SEM, standard error of the mean) refers to IBC between region A and B, and the discontinuous gray line with green shaded area (SEM) refers to IBC between region A and C. On the x-axis the different information fluctuation levels are depicted.

For each simulated signal, IBC was calculated for the target connection between regions A and B, while the connection between regions A and C served as control. To depict IBC for both inter-regional connections, $IBC_{chance\text{-}level}$ values were subtracted from the IBC_{metric} values and the resulting difference was

averaged across the multiple simulated signals. Unexpectedly however, when no information was available within the tested regions (CNR=0), noise covariance seemed to directly modulate IBC (Figure 5, left column). To further investigate such consistency of single-trial predictions expressed by IBC as modulated by noise covariance in the absence of information, we hypothesized this effect emerges from a model learning in the absence of information, which reflects consistency between trial patterns dependent on noise fluctuations and in turn on noise fluctuation covariance. Therefore, we employed *simulation-2* (Figure 6) in which the seed region always contained discriminant information patterns (CNR=.5), whereas the information content for the target region and control region was manipulated parametrically (CNR varied from 0 to 1 in intervals of .2). Our findings show that when one of the regions (e.g., the seed region) contains information representation, noise modulation of IBC was absent (Figure 6, left column). Importantly, the modulatory effect of information fluctuation covariance was tested using linear regression parameters (i.e., slope) representing its effect on IBC for *simulation-1* (Figure 7A) and *simulation-2* (Figure 7B). Confirming our hypothesis, information fluctuation covariance led to a significant modulation of IBC ($p<0.05$) in all CNR conditions except when no information was available (CNR=0).

Results of fMRI experiment

IBC statistical maps for five seed regions-of-interest previously identified (chapter 4) in the decoding analysis of place of articulation for the left hemisphere (Figure 8A) are depicted in figure 8B. The selected regions-of-interest of the left hemisphere were the medial STG (superior temporal gyrus), posterior STG, posterior inferior post-CG (central gyrus) and anterior SMG (supramarginal gyrus).

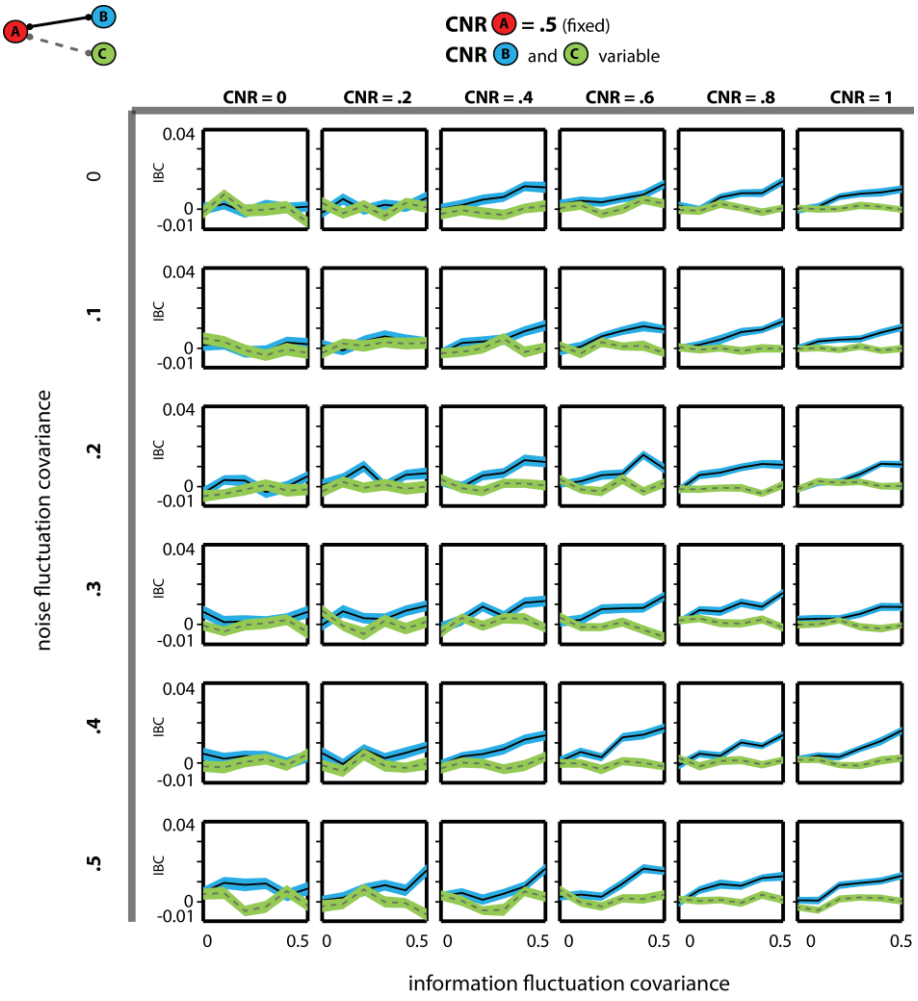


Figure 6. Results of IBC for *simulation-2* (fixed CNR for region A, variable CNR across regions B and C). Rows are different levels of noise fluctuations. Columns are different levels of CNR for regions B and C. CNR of region A was kept constant to test IBC in scenarios comprising seed regions containing discriminant information representation. Within each result chart, the continuous black line with blue shaded area (SEM) refers to IBC between region A and B, and the discontinuous grey line with green shaded area refers to IBC between region A and C. On the x-axis the different information fluctuation levels are depicted.

The IBC results show the existence of inter-regional connectivity that is beyond a local spread of mutual information related with the spatial interdependency of overlapping *searchlight* patterns. The IBC statistical maps indicate mutual articulatory feature representation, such as between medial STG and insular and motor regions (Figure 8B-1), posterior inferior post-CG and posterior STG/STS (superior temporal sulcus)/MTG (middle temporal gyrus) and pre-central gyrus (Figure 8B-3), and between anterior SMG and motor regions in the anterior bank of pre-central gyrus (Figure 8B-4). The time-series maps depict temporal signal correlations of the seed regions across the cortex (Figure 8C). Visual inspection allows identifying apparent differences between the time-series correlations and IBC. In particular, the connectivity to (pre)-motor areas (pre-central gyrus) found in IBC was not found when using such univariate temporal correlation approach. In turn, also time series show areas connected to the language network, that were not detected by IBC, for example the SMG in connection to the posterior MTG.

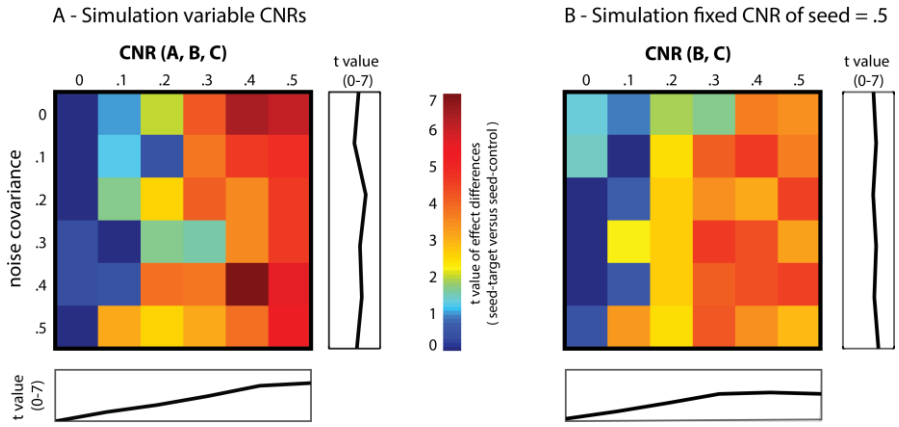


Figure 7. Modulatory effect of information fluctuation covariance between region A and B under different CNR and noise covariance conditions. A) Results for simulation 1 (CNR of regions A, B and C are varied conjointly). B) Results for simulation 2 (CNR of region A is fixed to .5 and CNR of regions B and C are varied conjointly). T values (paired t-test, one sided) are depicted from blue to red (0<t<7), t(99)=1.66 corresponds to p=0.05 and t(99)=2.36 corresponds to p=0.01. Marginal averaged t values across noise fluctuation and information contrasts are shown at the side and below the matrices.

Discussion

In this study we introduce a new method to investigate brain connectivity that relies on shared information representation of brain-states across regions – 'information-based connectivity' (IBC). Information representation obtained by MVPA in focal regions-of-interest (e.g., in *searchlight* approach, Kriegeskorte et al., 2006) is used to find region-pairs that encode the same information in a consistent manner across single-trial predictions. This method allows identifying brain networks that rely on information instead of activation-based inter-trial covariance. Because IBC uses single-trial fMRI estimates instead of fMRI time-series, it is less dependent on temporal dynamics of fMRI responses (Rissman et al., 2004; Waldorp et al., 2011). The advantage of information- over activation-based connectivity relates to its potential to identify information-sharing among distinct regions involved in distributed neural representation of perceptual or cognitive states. At the informational level, IBC across brain regions suggests an interdependent processing of information underlying specific brain-states. Moreover, the low temporal resolution of fMRI recordings and the inherent slow hemodynamic changes of neuronal activity are current limitations to functional connectivity analysis based on fMRI time-series (Eichler, 2005). When BOLD responses are summarized to single-trial estimates, this temporal information is omitted (Rissman et al., 2004; Waldorp et al., 2011). Furthermore, IBC varies considerably from the IC (*informational connectivity*) method proposed by Coutanche and Thompson-Schill (2013) that correlates averaged decoding accuracies between regions obtained from time-point patterns of the fMRI responses. IC is thus directly dependent on the temporal dynamics of the fMRI time-series. IC may be particularly advantageous when performing connectivity relying on MVPA in block-designed experiments, whereas the IBC method we propose directly profits from slow event-related designs.

Using simulated single-trial fMRI estimates, we evaluated the feasibility and validity of IBC. Importantly, the simulations show that IBC is particularly

sensitive to covariance of information fluctuations across regions, but not on the overall signal covariance. Our simulation findings further show that IBC must be employed using seed regions containing information representation, to avoid full-scale pair-wise connectivity computations. When information content is not present among the seed or target regions (CNR=0), an undesirable IBC_{metric} larger than $IBC_{chance\text{-}level}$ emerges in the presence of large, spurious correlations (see left column of Figure 5). We propose that such effect relates to the fact that model learning in the absence of information reflects existing consistencies between single-trial patterns. These consistencies relate to high noise fluctuation covariance between regions, leading to covariance of single-trial patterns that deviate from the learned patterns. To test such hypothesis, we carried a second set of simulations (*simulation-2*) in which the seed region always contained information content (CNR=.5) and the CNR of the target and control region were manipulated parametrically as before (CNR ranged from 0 to 1). This approach allowed us to confirm that when one region (e.g., the seed region) contains information representation, spurious IBC disappears (see left column of Figure 6). These conclusions suggest the need to use seed regions selected based on their information content prior the IBC analysis. This procedure is safe from circularity of analyses given the fact that 1) MVPA analysis and IBC analysis are independent under the perspective that MVPA is performed region-wise, whereas IBC is performed pair-wise. Selecting regions based on their elevated MVPA accuracies does not posit a bias for strength of IBC.

IBC employed to empirical fMRI data shows its potential as a tool to identify functionally connected neural networks based on transfer of information when perceiving spoken syllables. The IBC maps are meaningful within the context of speech perception and in particular of sensorimotor integration, and bring new insights onto how neural articulatory representations of spoken syllables are transferred across the brain. Using real fMRI data originated from a speech processing investigation reported in chapter 4, we computed IBC statistical maps that show mutual information representation across cortical regions. The

representation and transfer of articulatory features of speech (Hickok and Poeppel, 2007) may subserve an efficient and robust speech perception and production capacity in normal adults (Guenther and Vladusich, 2012; Hickok et al., 2011). Moreover, we visually compared connectivity statistical maps produced by IBC (Figure 8B) and produced by an activity-based approach using fMRI time-series as predictors in GLM (Figure 8C). The unique capacity of IBC to highlight shared information between the seed-regions and (pre)-motor areas in the anterior pre-central gyrus is observed, suggesting that information representation, but not activity correlation, occurs between these regions.

This study places IBC as a potential method of functional connectivity for fMRI. However, further aspects underlying the statistical assessment of IBC remain unexplored to date. One aspect that requires further consideration deals with the interdependency of IBC variability and MVPA accuracy. The closer the MVPA accuracies of two regions are to chance-level (i.e., 0.5), the larger their possible IBC range. In extreme cases, if two regions have perfect classification accuracies (i.e., averaged accuracy = 1), the IBC_{metric} would be equal to $IBC_{chance\text{-}level}$. Since neuroimaging studies that exploit the sensitivity of MVPA to investigate cognitive processing typically do not comprise large univariate effects, classification accuracies usually do not exhibit such extreme values. Hence, IBC provides a potentially useful new connectivity method for MVPA research in neuroimaging. IBC range variability can be observed in the IBC simulations, in particular in the progressive reduction of SEM (standard error of the mean) with the increase of information contrast (Figure 5 and Figure 6). To prevent modulations of IBC due to classification accuracy, a normalization of the IBC_{metric} in respect to the IBC range limits can be performed, for example, using odds-ratio normalization (Szumilas, 2010).

Beyond brain connectivity, the conceptual methodology of IBC promises to find mutual information across different imaging modalities. In particular, a similar IBC method may allow combining simultaneous fMRI and EEG recordings (De Goldman et al., 2009; Martino et al., 2011; Huster et al., 2012).

This approach may help uncovering information rather than activation that is immune to the disparate nature of the acquired signals (Uludag and Roebroeck, 2014). Furthermore, inter-trial decoding consistencies in simultaneous fMRI-EEG recordings may unravel both complementary and shared spatio-temporal cortical representation patterns that are exploited in modality-specific analyses, such as including aspects related to the spatial organization of fMRI and to temporal and oscillatory dynamics of EEG signals.

Conclusions

In conclusion, IBC combines two important characteristics to investigate functional connectivity in fMRI, extending available connectivity measures. First, IBC exploits consistency of information representation instead of activation across brain regions. To better assess the neural circuitry responsible for higher-order cognitive processes, it is crucial to include aspects related to information representation and transfer (Coutanche and Thompson-Schill, 2013). Second, IBC uses single-trial estimates instead of BOLD time-series. Connectivity measures using single-trial estimates are less dependent on the temporal dynamics of fMRI signals and provide advantage to assess functional connectivity in slow event-related designs (Rissman et al., 2004). IBC may be used in combination with other methods of functional and effective connectivity to better describe the neural networks involved in certain cognitive processes and tasks.

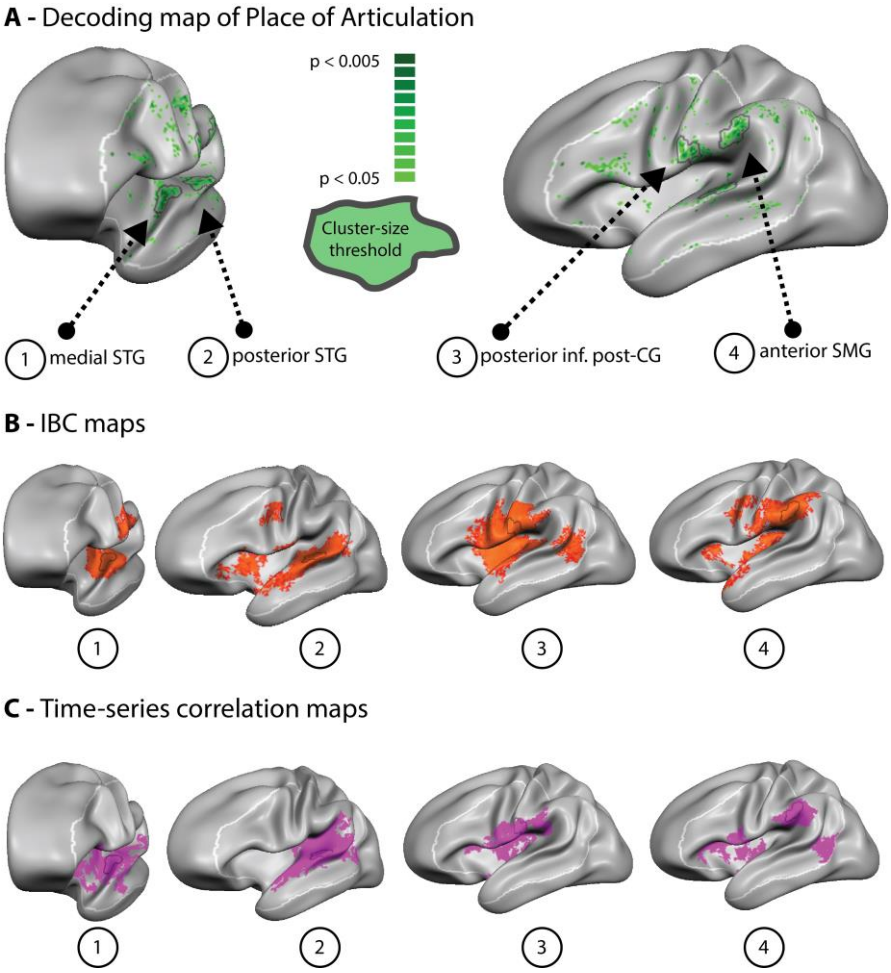


Figure 8. Results of IBC applied to real fMRI data. A) Decoding of articulatory features of speech (place of articulation) in the left hemisphere. White line refers to the region of interest originally used in the decoding analysis of articulatory features of speech. Green colored regions depict group level decoding statistical values ($p<0.05$, uncorrected). Regions circled by a black line were statistically corrected for multiple comparisons using cluster size threshold correction. B) IBC statistical maps for each of the selected regions-of-interest. Maps (1 to 4) reflect IBC in the respective seed regions circled by a black line. C) Time-series correlation maps for each of the selected regions-of-interest. Maps (1 to 4) reflect the temporal correlation based on activity strength in the respective seed regions circled by a black line. Maps show vertices connected to the seed region ($p<0.05$, uncorrected, cluster size correction with alpha = 5%).

References

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–28.

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage*, *56*(2), 582–92.

Correia, JM., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(1), 332–8.

Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning 20*:273–297.

Coutanche, M. N., & Thompson-Schill, S. L. (2013). Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Frontiers in Human Neuroscience*, *7(2)*, 15.

De Martino, F., de Borst, A.W., Valente, G., Goebel, R., Formisano, E., 2011. Predicting EEG single trial responses with simultaneous fMRI and relevance vector machine regression. NeuroImage 56, 826–836.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58.

Eichler, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1457), 953–67.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, NollDC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of cluster-size threshold. *Magnetic Resonance Medicine 33,* 636–647.

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

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13–36.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–29.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302.

Frost MA, Goebel R (2012) Measuring structural–functional correspondence: spatial variability of specialised brain regions after macroanatomical alignment. *Neuroimage 59,* 1369 –1381.

Galantucci, B., Fowler, C. a, & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–77.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping 27,* 392– 401.

Goldman, R. I., Wei, C.-Y., Philiastides, M. G., Gerson, A. D., Friedman, D., Brown, T. R., & Sajda, P. (2009). Single-trial discrimination for integrating simultaneous EEG and fMRI: identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *NeuroImage*, *47*(1), 136–47.

Guenther, F. H., & Vladusich, T. (2012). A Neural Theory of Speech Acquisition and Production. *Journal of Neurolinguistics*, *25*(5), 408–422.

Hagler DJ Jr, Saygin AP, Sereno MI (2006) Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage 33,* 1093–1103.

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, 2425–2430.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing, *8*(May), 393–402.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Huster, R.J., Debener, S., Eichele, T., Herrmann, C.S., 2012. Methods for Simultaneous EEG-fMRI: An Introductory Review. The Journal of Neuroscience 32, 6053–6060.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8.

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(38), 13273–80.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology : CB*, *17*(19), 1692–6.

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 (New York, N.Y.)*, *320*(5880), 1191–5.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*(5), 351–60.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*(2), 752–63.

Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, *25*(1), 230–42.

Sun, F. T., Miller, L. M., & D'Esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *NeuroImage*, *21*(2), 647–58.

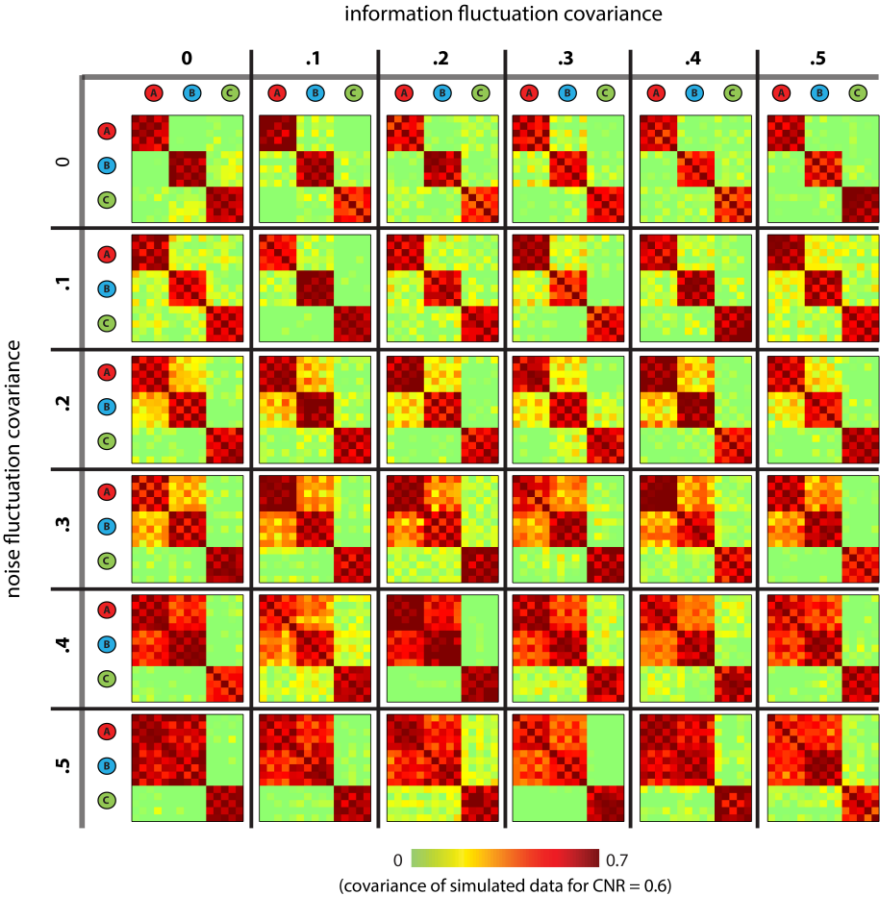
Szumilas M (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie Canadienne de Psychiatrie de L'enfant et de L'adolescent*, *19*(3), 227–9.

Uludağ, K., & Roebroeck, A. (2014). General overview on the merits of multimodal neuroimaging data fusion. *NeuroImage*, 1–8.

Vigneau-Roy, N., Bernier, M., Descoteaux, M., & Whittingstall, K. (2014). Regional variations in vascular density correlate with resting-state and task-evoked blood oxygen level-dependent signal amplitude. *Human Brain Mapping*, *35*(5), 1906–20.

Waldorp, L., Christoffels, I., & van de Ven, V. (2011). Effective connectivity of fMRI data using ancestral graph theory: dealing with missing regions. *NeuroImage*, *54*(4), 2695–705.

Suplementary figure



Supplementary figure S1. Covariance matrices of simulated signals for a representative information contrast (CNR = 0.6). Rows are different values of noise fluctuation covariance. Columns are different values of information fluctuation covariance. Each signal covariance matrix consists of 15 by 15 voxels, with 5 voxels per region.

CHAPTER 6

Summary and future directions

Perceiving and using spoken language is a fast and effortless capacity that is essential for verbal communication in everyday life. The function of speech perception transcends our ability to listen and understand someone else's speech. It is part of the cognitive machinery that enables us to speak in a connected manner, to think using a linguistic 'currency', to access (verbal) memories and to enjoy the benefits of a conscious mind. Efforts to understand the neural mechanisms of speech perception are essential to face the challenges underlying language disorders, to develop brain computer interfaces or to understand human cognition. Speech perception involves both the processing of lower-level perceptual features of speech as well as their link to higher-order speech and language networks. This thesis investigates two levels of neural representation during online speech perception, those at the interface with the semantic memory network (Chapters 2 and 3) and the sensorimotor systems (Chapter 4).

Speech comprehension demands the access to lexical-semantic networks that activate semantic-conceptual representations distributed throughout the cortex. The neural mechanisms supporting these semantic-conceptual representations remain largely unresolved (Damasio, 1989; Patterson et al., 2007). Sensorimotor integration links the auditory and the motor systems of speech enabling the exchange of relevant speech information during acquisition and development of speech and language, and speech production in everyday life (Guenther and Vladusich, 2011). The need for sensorimotor integration during speech perception is highly debated (Hickok et al., 2011; Pulvermuller and Fadiga, 2010; Galantucci, 2006). In particular, methodological constraints related to experimental design and analysis methods have so far prevented the disentanglement of neural responses to acoustic versus articulatory features of speech.

This thesis uses fMRI (functional magnetic resonance imaging) and EEG (electroencephalography) to image brain responses to particular properties of speech utterances, namely the semantic identity of words and concepts, and articulatory representations of spoken syllables. Multivariate pattern analysis

(MVPA) was employed to exploit the interaction between multiple brain imaging responses in combination with its capacity to predict information from complex response patterns (Formisano et al., 2008a; Pereira et al., 2009). Most crucially, we have employed generalization strategies that show the capacity to classify speech utterances across variation of other acoustic dimensions (Formisano et al., 2008b), such as concepts across different word forms from English and Dutch (Chapters 2 and 3), and articulatory features of syllables across other articulatory dimensions (Chapter 4). Furthermore, we have investigated a new approach to find connectivity between distributed brain regions encompassing the representation of brain-states as uncovered by fMRI-based MVPA (Chapter 5). This approach enables to study neural communication in brain networks based on the transfer of information representation instead of covariance of fMRI activity strength.

In chapter 2 and chapter 3 we investigate the neural representation of semantic concepts evoked by equivalent but acoustically distinct spoken Dutch and English words in bilingual subjects. We analyzed brain responses to several presentations of four monosyllabic animal words in each language. Using MVPA, we trained classifiers to predict the individual words participants were listening to within each language and thereafter tested whether this learning was generalizable to the other language. In chapter 2, we used fMRI and showed the feasibility of both within-language discrimination and across-language generalization using a focal multivariate approach – the *searchlight* method (Kriegeskorte et al., 2006). Our findings indicate that in bilingual listeners, language-independent semantic-conceptual knowledge is organized in abstract form in focal regions of the cortex. In particular, we observed language-independent representations of spoken words in the left anterior temporal lobe (ATL), which converges with previous findings showing fMR adaptation effects in this region for cross-language semantic priming of visual words in bilinguals (Visser et al., 2012). Furthermore, it corroborates the neuropsychological findings in semantic dementia (Damasio et al., 1996) and indicates a role of the left ATL as a central-hub in abstract semantic processing (Patterson et al., 2007).

In chapter 3 the same experimental design was adapted to EEG to investigate the temporal and temporal-oscillatory dynamics of spoken word comprehension using MVPA. We identified specific time windows and frequency bands enabling discrimination and generalization of individual spoken words. We demonstrated within-language word decoding in a broad time-window from ~50 to 620 ms after word onset with a strong contribution of slow oscillations (below 12 Hz). Most importantly, we were able to isolate specific time windows, including the 550-600 ms window, in which EEG features enabled the generalization of the meaning of the words across their Dutch and English word forms. The results of chapter 3 demonstrate the feasibility of using MVPA to identify individual word representations based on speech evoked EEG signals. Furthermore, they indicate the advantage of feature selection approaches in assessing temporal (Hausfeld et al., 2012; Chan et al., 2011) and temporal-oscillatory EEG response features in classification.

In chapter 4, we used fMRI in combination with similar cross-class generalization strategies to investigate the representation of articulatory features of spoken syllables during speech perception. In this study we explored whether during the perception of speech, the brain extracts motoric information of speech signals, such as related to their articulatory gestures instead of acoustic information alone. Using a *searchlight* approach we highlighted several cortical areas that are sensitive to articulatory features, such as place of articulation, manner of articulation and voicing. Crucially, using a generalization strategy we were able to decode articulatory features of spoken syllables, across acoustic variation of other articulatory dimensions. This strategy is important since most articulatory gestures of speech sounds also vary in acoustic properties, thus restricting the possibility to disentangle both types of representations. Our cortical generalization maps suggest invariant neural representations of manner and place of articulation within a distributed network of auditory, somatosensory and motoric neural populations and of voicing in a more restricted region in the right anterior temporal cortex. Furthermore, they support the notion that the representation of spoken

syllables during on-line speech perception includes the transformation of acoustic input to abstract articulatory codes.

In chapter 5, inspired by the sensitivity of *searchlight* to find focal activity patterns informative of stimulus related brain-states, we developed and validated a novel approach for functional connectivity that relies on the availability of information instead of activation across brain regions. In this *information-based connectivity* (IBC) approach information representation obtained by MVPA in focal regions-of-interest is used to find region-pairs that encode the same information in a consistent manner across single-trials. The advantage of information- over activation-based connectivity relates to its potential to identify neural representation transfer across brain regions that originate from information fluctuations beyond co-variation in activation strength. Furthermore, current limitations of functional connectivity analysis based on fMRI time-series (activation strength) include the low temporal resolution of fMRI recordings and the inherent slow hemodynamic changes of neuronal activity (Eichler, 2005). IBC uses single-trial estimates to investigate functional connectivity at the informational level, which minimizes the risk for spurious interpretations of temporal correlations of BOLD responses across brain regions (Rissman et al., 2004; Waldorp et al., 2011).

Overall, the findings in this thesis follow on the success of the *searchlight* method as an approach to map information representation in the cortex (Kriegeskorte et al., 2007). Specifically, we have exploited the potential of *searchlight* and its characteristic of using patterns of small number of voxels (features), to generalize information across different stimulus dimensions. We propose that reduced feature spaces may maximize the ratio of generalizable features, and that this factor is crucial in designs that we refer here as *cross-class generalization*. Beyond discrimination, this strategy allows findings commonalities between information patterns. In chapters 4 and 5, the *searchlight* method was employed on the cortical surface (Chen et al., 2011) instead of in the volume, allowing a selection of focal features based on their geodesic rather than volumetric proximity. This methodological change posits an advantage to avoid

cross-gyral overlap, and was particularly relevant to reduce the possible contamination of activity between auditory and motor regions. The *searchlight* method also served as an inspiration for the functional connectivity approach described in chapter 5. We are interested to further pursue the IBC method as a potential method to combine, at the informational level, recordings from different imaging modalities acquired simultaneously, such as fMRI and EEG. There, similar principles as in IBC may allow finding spatial and temporal associations between patterns from fMRI and EEG using their information-based single-trial predictions rather than relying on the disparate physiological nature of these signals.

Researching speech using multivariate classification was fundamental in this thesis. The findings with regard to conceptual and sensory-motor representations can help to formulate new and finer grained hypothesis for speech processing models. They will also bring a better understanding of functional (activation and information based) neural networks, representation, and computation. From a methodological point of view, to further our understanding on neural mechanisms of speech, we believe it is crucial to 1) increase the statistical sensitivity and specificity of data analysis methods, and 2) to pursue the within-category investigation of neural representations of individual speech items (e.g. syllables, words, etc), instead of their average activity across categories (e.g., words versus non-words). In fact, the capacity to discriminate between two items from the same category is crucial to understand and comprehend speech in everyday communication. MVPA provides the means to ask such questions with the adequate statistical sensitivity. Its capacity to unravel information discrimination together with information commonalities using generalization strategies employed in this thesis, also promises to study the interlinked representations between speech perception and speech production faculties. When combined with well-designed stimuli and experiments, this approach shows potential to answer a number of unresolved questions in speech research, such as 1) the link of speech perception and speech production with a network for conceptual knowledge, and 2) the similarity of articulatory representations during perception and production.

Our results also indicate potential to assess the contribution of temporal-oscillatory EEG response features in the representation of speech. Beyond spatial and temporal signatures of speech perception, temporal-oscillatory features relate to the neural mechanisms underlying spoken word and concept representation (Peelle and Davis, 2012; Giraud and Poeppel 2007; Lakatos et al., 2005). In the future, methods able to assess the multivariate contribution of more sophisticated aspects of temporal-oscillatory dynamics, such as phase-phase-coupling and phase-amplitude-coupling (Voytek et al., 2013) may allow identifying crucial aspects of speech perception and production that may relate to an entrainment of neural oscillations to temporal features of speech signals and synchrony mechanisms across different frequency bands over time.

References

Chan AM, Halgren E, Marinkovic K, Cash SS (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*(4), 3028–39.

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage*, *56*(2), 582–92.

Damasio AR (1989) Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*. 33, 25-62.

Damasio, H., Grabowski T. J., Tranel D., Hichwa R. D., & Damasio A. (1996). A neural basis for lexical retrieval. *Nature*, 380, 499-505.

Eichler M (2005) A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1457), 953–67.

Formisano E, De Martino F, Valente G (2008a) Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging, 26*, 921–934.

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

Galantucci B, Fowler Ca, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–77.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–7.

Hausfeld L, De Martino F, Bonte M, Formisano E (2012) Pattern analysis of EEG responses to speech and voice: influence of feature grouping. *NeuroImage*, *59*(4), 3641–51.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America ,104*, 20600 –20605

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology, 94*, 1904–1911.

Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews. Neuroscience*, *8*, 976–87.

Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*(9), 320.

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage, 45*, 199–209.

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, *11*(5), 351–60.

Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*(2), 752–63.

Visser M, Jefferies E, Embleton KV, Lambon-Ralph MA (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience, 24*, 1766 –78.

Voytek B, Esposito MD, Crone N, Knight RT (2013) A method for event-related phase / amplitude coupling, *Neuroimage*, *64*, 416–424.

Waldorp L, Christoffels I, van de Ven V (2011) Effective connectivity of fMRI data using ancestral graph theory: dealing with missing regions. *NeuroImage*, *54*(4), 2695–705.

CHAPTER 7

Knowledge Valorization

This thesis aims to understand fundamental neural mechanisms underlying the human ability to communicate via language. Such ability empowers people from all countries, cultures and religions to structure and communicate their thoughts. Effective verbal communication requires a neural system able to interact with and integrate sensory information from multiple modalities, such as audition and vision, to translate perceptual input to meaning-based memory representations during speech perception and to transform the intended message to motor actions for speech production. This thesis builds upon an intrinsic curiosity for this fascinating and highly efficient cognitive capacity. Understanding basic neural mechanisms of speech perception in the healthy adult brain will provide an important benchmark for studying normal and problematic language development throughout the lifespan. In the long run, the knowledge and method development generated by this thesis will help addressing unresolved language and speech disorders that impair millions worldwide. Persons with communication disorders, such as dyslexia, stuttering, apraxia of speech or specific language impairment (SLI) suffer from an impoverished quality of life. Additional fundamental knowledge on the neural systems supporting speech and language may in the future lead to a significant improvement of treatments and prevention strategies that are at present only partly effective. In particular, understanding the complex mechanisms underlying speech and language may empower clinicians with individually tailored solutions that may reveal imperative to achieve successful results in such diverse and variable clinical conditions.

Overall, our understanding of neural information hidden in small electric signals and chemical changes, revolving around billions of microscopic neurons and neural connections in the human working brain remains critically limited. Necessary progress in unravelling the neural mechanisms of speech and language in the healthy brain demands a fusion of well-designed cognitive/psycholinguistic paradigms and state-of-the-art non-invasive brain imaging techniques. This thesis aims to advance on both these directions: first, by designing more optimal ways to investigate at fine-grained level of detail the processing of individual spoken sounds; and second, by applying novel analysis

schemes that permit unraveling information content from brain responses to spoken sounds across different stimulus dimensions and cortical regions.

Translation towards speech and language development and disorders

Recent developments in brain imaging and brain stimulation technologies have inspired researchers to adopt novel approaches towards the treatment and identification of biomarkers for speech and language disorders. Technologies such as EEG, (f)MRI, TMS (transcranical magnetic stimulation) and TCS (transcranical current stimulation) in combination with neurofeedback solutions show potential to target normal and abnormal speech and language development with unprecedented detail. As these technologies improve at a fast rate, it is crucial to find direct and mechanistic evidence for their potential translation to clinical settings. Naturally, no animal testing is possible in speech and language research and thus, clinical approaches must have strong scientific foundations. Our work aimed to set an additional building brick on such foundations. The experimental paradigms and analysis approaches could be used to study the refinement of perceptual, articulatory and meaning based neural representations during normal and deficient language development. For example, they can help in clarifying why individuals show such large differences in the ability to learn a second/new language, and in understanding underlying deficits in children that struggle in learning to read due to poor decoding skills (developmental dyslexia) and/or more general comprehension deficits.

Further uncertainty exists in the relationship between speech perception and production. Speaking and listening seem to be highly interactive components of the same speech processing system. However, brain imaging studies find variable activations when subjects perceive or produce speech. Analysis schemes based on pattern generalization approaches used throughout this thesis promises to investigate commonalities between such systems that may

reveal crucial to understand their interdependence in everyday life, normal and abnormal development. For example, the foundations of speech production impairments may include a deficient activation of predictive auditory templates during online production and/or a deficient parsing of auditory speech sounds onto articulatory representations at some point of the person's development. Moreover, dyslexia may include a similar erroneous mapping between written, auditory and articulatory representations during a critical developmental stage.

Methodological innovation

The projects in this thesis adopted a combination of carefully controlled experimental stimuli and designs with multivariate fMRI/EEG decoding techniques. This provides an innovative approach to investigate information content from patterns of brain responses to speech, language and reading. Generalization approaches are valuable to investigate higher-order representations across several fields of neuroscience, such as object recognition, language, memory as well as feelings and emotions.

We further investigated a new method that aimed to assess communication between brain regions during specific cognitive states. Complementary to overall activity of different brain regions, inter-regional communication during high-order cognition, such as speech perception is fundamental to describe the operational networks of speech and language processing. Similarly, other high-order cognitive capacities rely on distributed brain networks. Our method seeks to unravel the transfer of information availability across the brain, thus promises to better characterize such cognitive capacities.

Finally, as our understanding of brain patterns across fMRI and EEG responses increase, it becomes attractive to develop translation mechanisms of fMRI patterns to more accessible and portable imaging techniques, as provided by EEG or fNIRS (functional near-infrared spectrocospy), which promise to play a central role in therapeutic solutions for speech and language.

Acknowledgments

Special thanks to those who read or are considering reading this thesis. Although the success of my work is not merely measured by the impact it has on other researchers but also by the impact it has on me and on my future, I feel an immense honor if someone is willing to devote her/his time to open this book and dive with me into investigations conducted on the nature of speech and language.

I would first like to thank my supervisors' "great vision" for believing in me after a job interview. Everyone is a potential liar during a job interview. A PhD pursued in any other place than Maastricht could have formed me in a complete different manner. I'm truly happy with the transformation this time had on me, which has, hopefully, launched a committed junior scientist to research neural mechanisms providing us the necessary faculties for verbal communication. Milene Bonte and Bernadette M. Jansma, your knowledge and personality has guided me well and inspired me up to the present day. I thank you for always letting me pursue my own ideas, but never without the right amount of argumentation. Dear Milene, once you accused me of being stubborn. You did so in such a positive way that made me happy and even proud that day. It took me a while to understand that I could profit from being a bit less stubborn in professional matters (at least). I appreciate your way of communicating with me on the everyday basis. It worked very well. I recognize your ability to do so, smoothly and effectively. Your growing confidence in my work is very important to me - thank you. Dear Bernadette, I always feel even more proud of working together with you when I presence your pertinent questions during highly complex scientific talks. No matter how far from your scientific area of specialization a topic may be, when you decide to pose a question I know it is wrapped by a unique capacity to see neuroscience from a clear and global perspective, where nothing stands unlinked. I try to follow your sense for integration of knowledge. Just starting here…

Next, I would like to thank some of the co-authors in the studies I've conducted during my PhD. Elia Formisano, Giancarlo Valente and Lars Hausfeld, I know that sometimes I may have sounded confused with technical

decisions on analysis methods. Your different ways of parsing complex processes onto easy to follow explanations, enriched with pragmatic and experience-based, almost emotional insights (Elia), complete theoretical contemplations (Giancarlo) and practical suggestions (Lars), was crucial to me during the past four years.

I would also like to share appreciation to members of Maastricht University that although not working directly with me, listening to their lectures and talks always functioned as a pill of extra motivation for research. Among these, I remember especially a course by Kamil Uludäg, talks by Rainer Goebel and Elia, the PhD defenses from Alex Goulas, Marieke Mur and Bert Jans, as well as several progress reports from other fellow Phd colleagues, such as Matteo Bastiani and Rosanne Rademaker that made me want to go immediately back to my desk and work harder on my goals.

During this time I also had the pleasure to collaborate, in the position of supervisor, with former bachelor and master students of Maastricht University. Thank you Sanne Kikkert, Elena Vieth, Miriam Löhr, Lukas Hemmers, Selma Kemmerer, Nora Krott and Deniz Dohmen. Sanne Kikkert, you were the first student I supervised and will always remember the time we spent together. I was extremely afraid of guiding you wrong. You were a very complete student, full of knowledge and motivation. Seeing you pursuing your own academic/scientific career is fantastic - good luck. Miriam Löhr, hope you are enjoying your master's internship with us as much as I do working with you. To share snacks and fruits over long fMRI scanning sessions while discussing a bit of everything, from science to music and travelling was great. Hope you have an adventurous summer break somewhere in Greece and then engage yourself in a PhD in neuroscience next academic year.

To my colleagues and friends with whom I shared many days and nights during the PhD time: Alex, Lars, Gojko Žarić, Aline De Borst, Marin Been, Matteo, Rosanne, Martin Havlicek, Britta Gräwe, Marieke and Kamil. I would love to keep you all in the same city for a long long time, but this is insofar not possible. Counting on Britta to find us that special corner of paradise with

sunshine all-year round and make us an offer we cannot refuse - “Britta?”. Marieke, you inspired me a lot when I first arrived to Maastricht and keep inspiring me now as a colleague and friend. Your energy is incredible. Sorry for eating your cookies in our Wycker Brugstraat flat. Alex Goulas, when you left Maastricht you left a hole in my social calendar that was hard to overcome. You know that, so just come to the party, you can do your laundry some other day. Marin, the boxing was fantastic. Getting occasionally punched in the face felt really good, I do admit now! Rosanne, you are a wonderful office mate. It took me a while on google translator to follow everything in the beginning, but I guess now it's easier! Jumpthedesk.com could have had become a viral trend on internet and we could have become rich (easily) or fired (easily). Kamil, 'El Clásico' is not the same without you. Watching those Portuguese guys from Real Madrid loosing big time is not so bad in your company given your notable sensitivity for joke making! But… when… unexpectedly, although non-surprisingly, Ronaldo makes impossible goals in Camp Nou, the environment gets as tense as probably Barcelona's bench at the same time – one gotta love those moments, many thanks for sharing. Aline, stay. We need your energy as MRI needs RF pulses. Hope to still share many more good times together. Prof. Dr. Bastiani, all the best in Oxford and for your defense. You are a scientific-success magnet and can only expect you to rock in science as much as you rock on your everyday life. Playing ‘You’ on your movie is an honor. Just imagine… being Bastianni for a few scenes. Who could resist?

Many thanks to all of those who were always available for a nice chat during coffee or lunch break. Felix Dücker and Anna Gardumi: it feels good to know your office is always available to me (or perhaps I force it), from early in the morning to late in the afternoon. Anna, Matteo ate your chocolate-made Santa (I suspect). Roberta Santoro, good luck for your own defense and all the best in Switzerland. I know “the dress is a pretty dress”. Sriranga Kashyap, sharing the office with you is a pleasure. Hope we find more time in the future for socializing. Jan Zimmermann, you used to seat on Sri’s chair. Missing our casual (aka. naked) Fridays with Rosanne! Good luck to your new professional challenge, hope you manage to find what you are looking for in science. Inge

Timmers, spending time with you is a recipe for having a good time. The trip to California will always be remembered. Alongside with Katie Wheat, Sanne ten Oever, Michelle Moerel, Arne Seehaus and Marieke, we had a great time singing 'California here we come' while sipping home-made super alcoholic fruit smoothies. Thank you Sanne for so many productive scientific discussions and Michelle for taking me to 'Ala' upon my first visit to Maastricht. Mario Archilla Melendez, compañeiro, en hora buena te casaste cuando te quedaban todavia algunos años sin dolor en la espalda. My sincere admiration to both of you. Valentin Kemper, it's always a pleasure to discuss together.

To Gojko Žarić and Lars Hausfeld, my paranymphs. I love you. I thank you in advance for the work you may have preparing for the day of my defense. You may be nervous alongside with me on that day, you may even shed a tear in moments of unbarable emotion load. Fear not. No matter what happens it will always end in a party with appropriate levels of joy. Gojko (I wrote your name correctly, for once at least), I hope you manage to arrive comfortably on time on this day carrying a bottle of your strongest Serbian liquor to help me out. Lars, don't forget the rings. Joking, this is a defense mate. Overall, thank you both for baring with me these years and specially the last months, days, hours, and final exciting minutes of the defense. Your sense of humor is a delight to me, even when it's bad. I guess that could be an alternative definition of friendship. Lars, we will do great with the crazy EEG-fMRI project. Gojko, I'm working on our project together right after finishing these acknowledgments.

To my Portuguese friends that are planning to travel to support me on the day of my defense. Knowing that time passes and our friendship remains in solid shape is a relief. Bruno, Tiago and Rita, Maastricht will be yours again for a weekend. Your lives grow busy, but we must make sure we keep them loaded with exciting plans together. Sofia (Tiny), the 'Xourissus' may be back together, no words. I will do my best for it to be an unforgettable weekend. Marisa, remember I threw a frisbee into your forehead when we were like 13 years old? It was an accident, ok? I hope this is not some kind of master revenge plan. I will ask the security to keep an eye on you and check your purse for killing

frisbees just in case. Many thanks for coming in representation of the 'Norte'. Tiago and Bruno, after this is done, let's set a time for a trip together. You two are probably the best travelling buddies one can ask for. Bruno. Obrigado pela tua amizade que parece mais um laço familiar, and for the cover of this thesis.

To Luís Xarelho, Valentina Petroulia, Harris Georgiopoulos and Veit Kubik. You were extremely important during those years prior the PhD, which involved taking such decision. Thank you for your friendship. Nuno Costa e Margarida Carneiro, ouvi rumores que me deixam especialmente feliz por vocês. Sou fan da vossa história e quero seguir atento a nova temporada. Bruno Andrade, Bruno Pinto e Leonor Costa, desejos de muitos mais verões no Algarve com boa comida e boa companhia. Malta do Kayak-Polo, é este ano que se ganha o Campeonato? Um abraço sentido para o Daniel Ferreira e mais uns quantos para o Ricardo Pereira, João Ribeiro e Diogo Andrade. Desculpas por não seguir sendo o atleta participativo que gostaria de ter sido nos últimos anos. Bem sei que comigo a todo-o-gás teriamos feito mais um pouco (i.e., ganhado e perdido mais um pouco).

Mãe, Pai, Sofia, Chris, Julia and Francisco. Love you with all my heart. Obrigado por todo o apoio durante estes últimos anos, principalmente na decisão de mudar o rumo da minha carreira. No entanto, essa mudança não culmina com esta tese, mas a busca por entender a complexidade do cerebro humano vai continuar por muitos mais anos. A parte mais difícil é essencialmente estar longe de vocês, não ver o Francisco e a Julia crescerem diariamente. Obrigado por me deixarem viver a vossa casa como se fosse minha durante as férias, e por vêces até, transformá-la num parque de diversões para fins agricolas.

To Carolin. The most special thanks of all. You 'enjoyed' the stress to have the thesis ready on time. I'm ready to do the same for you in two years and a few months from now. No house is too small when it's shared with an enormous amount of love! Thank you for you kindness, friendship and love. Yours, João.

PUBLICATIONS

Correia JM, Jasma BM, Hausfeld L, Kikkert S, Bonte M (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Frontiers in Psychology,* 6(2), 1-10.

Correia JM, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(1), 332–8.

Correia JM, Jansma BM, Bonte M (*Submitted*). Brain-based decoding of spoken syllables isolates articulatory representations during passive listening.

Correia JM, Bonte M, Hausfeld L, Jansma BM, Valente G (*in Preparation*). Information-based connectivity – IBC: a functional connectivity approach for fMRI using multivariate classification.

Zander, T.O., Lehne, M., Ihme, K., Jatzev, S., **Correia, JM.**, Picht, B., Nijboer, F. (2011). A dry EEG-system for Scientific Research and Brain-Computer Interfaces. *Frontiers in Neuroscience, 5(53).*

Dai, Y., Yu, X., Xiao, Z., Xu, Y., Zhao, M., **Correia, JM.**., Reed, G. M. (2014). Comparison of Chinese and international psychiatrists' views on classification of mental disorders. *Asia-Pacific Psychiatry : Official Journal of the Pacific Rim College of Psychiatrists*, *6*(3), 267–73.

Evans, S. C., Reed, G. M., Roberts, M. C., Esparza, P., Watts, A. D., **Correia, JM.**, M., Saxena, S. (2013). Psychologists' perspectives on the diagnostic classification of mental disorders: results from the WHO-IUPsyS Global Survey. *International Journal of Psychology : Journal International de Psychologie*, *48*(3), 177–93.

Reed, G. M., Roberts, M. C., Keeley, J., Hooppell, C., Matsumoto, C., Sharan, P., Robles, R., Carvalho, H., Wu, C., Gureje, O., Leal-Leturia, I., Flanagan, E. H., **Correia, JM.**, Maruta, T., Ayuso-Mateos, J. L., de Jesus Mari, J., Xiao, Z., Evans, S. C., Saxena, S., and Medina-Mora, M. E. (2013). Mental health professionals' natural taxonomies of mental disorders: implications for the clinical utility of the ICD-11 and the DSM-5. Journal of Clinical Psychology, 69(12), 1191–212.

Reed, G. M., **Correia, JM.**, Esparza, P., Saxena, S., & Maj, M. (2011). The WPA-WHO Global Survey of Psychiatrists' Attitudes Towards Mental Disorders Classification. *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *10*(2), 118–31.

Dias, J. M., Bastos, R., **Correia, JM.**, Vicente, R. (2005). Semi-automatic 3D reconstruction of urban areas using epipolar geometry and template matching. *Computer-Aided Civil and Infrastructure Engineering , 21 (7), 466*.

CURRICULUM VITAE

João Miguel Mendonça Correia was born in Lisbon, Portugal, on November 29th 1981. After completion of the secondary education in 1999 he enrolled at the University to study Telecommunications and Computer Science Engineering in ISCTE (Instituto Superior de Ciências do Trabalho e da Empresa), Lisbon. While obtaining his degree in 2005, he started an engineering career in *Siemens* where he worked on international projects concerning the research and development of 3rd Generation telecommunication platforms. In September 2008, João decided to take a sabbatical time to pursue a Master program in Biomedical Engineering at the University of Patras, Greece. The master education attracted him to brain research, which was further amplified during an internship at the Technical University of Berlin, working in the field of Brain Computer Interfaces using EEG. Later, he continued the development of his Master thesis project at Stockholm University and Karolinska Institute, Stockholm, researching EEG correlates of working memory performance in young and elderly populations. After the master studies, João could only envision himself working in the field of brain sciences. He was involved in a series of 'mental health' projects conducted in the World Health Organization (WHO), Geneva, in 2010 and later that year decided to pursue scientific goals involving human brain imaging in the research of neural processes required for verbal communication. In the framework of such goals, João conducted his PhD under the supervision of Prof. Bernadette M. Jansma and Dr. Milene Bonte from October 2010 to October 2014 at the department of Cognitive Neuroscience, Maastricht University. During this time he investigated brain mechanisms crucial for the perception of spoken language, specifically on brain response patterns to individual spoken items employing computational models exploiting multivariate analysis. From October 2014 to the present date, João is a post-doctoral researcher involved in multiple brain imaging projects on speech perception and speech production within the same collaboration.