Different Representations of Neural Correlates of Spoken Language Comprehension in Local and Global Brain Activation Patterns

Von der Fakultät für Medizin und Gesundheitswissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels

eines Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von Annika Urbschat

geboren am 20.06.1990 in Oldenburg

Gutachter: PD Dr. Stefan Uppenkamp
Weitere Gutachterin: Prof. Dr. Katharina von Kriegstein

Tag der Disputation: 01.09.2021

Summary

The ability to understand spoken language is a remarkable achievement of the human brain. It is very robust against various distortions and adaptable to multiple languages. Regardless of the high interest to scientists in neuroimaging research, this complex process has not been fully understood yet. One reason for this are the limited possibilities to investigate the intact brain and the large amounts of data that are required by these methods. In the past decade methods like multivariate pattern analysis (MVPA) have been established, which have proven to be more sensitive than previous approaches to investigate brain imaging data. Even though the vast variety of statistical methods constantly bears new possibilites, it makes it difficult to compare the results obtained from different studies using different approaches.

This thesis is composed of two parts. In the first part, a new approach to evaluate the results obtained from the searchlight classification analysis of fMRI data is presented, named Searchlight Classification Informative Region Mixture Model (SCIM). The evaluation model assumes that the distribution of classification performance results can be decomposed into informative searchlights, and non-informative searchlights. While BOLD patterns in informative searchlights contain information about corresponding contrasted conditions, the classification of non-informative searchlights corresponding to two contrasted conditions is not significantly above chance level. The decomposition is accomplished with a two-component Gaussian mixture model. The method is compared to commonly used methods like the binomial test and the random permutation test in different statistical evaluations of a simulation dataset and a real auditory fMRI dataset. The SCIM method has shown to be highly robust to the choice of significance levels in comparison to the reference methods and benefits from its comparably low numerical effort. In the second part of the thesis the results obtained from an auditory fMRI study to investigate the cortical activation during the semantic processing of spoken language are presented.

Many datasets only allow for the investigation of one particular part of spoken language processing or include degraded speech as contrast condition to normal speech. The stimulus set used in the study presented in this thesis includes, amongst four other acoustic stimulus conditions, a speech condition which is composed of valid spoken sentences, and a control condition, which is physically identical to normal speech but does not contain any meaningful statements. Therefore, it permits the identification of the cortical activation corresponding to the semantic processing of spoken language, while acoustic differences between the contrasted stimulus conditions are decreased. Results on a group level are obtained by the generation of group result maps and the statistical evaluation of single subject result maps across the group of participants. To disentangle the different stages of spoken language processing, the classification performance results obtained from thirteen condition contrast analyses are correlated with the corresponding differences of the stimulus features. This approach permits the identification of local brain activation patterns corresponding to different stages of spoken language processing. While correlation effects with acoustic property differences are found to be located in the auditory cortex and its belt and parabelt areas and in the inferior frontal cortex, the semantic processing of speech is found to activate a broadly distributed network of structures across the cortex.

The effective connectivity between these areas was investigated with the analysis of their generalized psychophysiological interactions (gPPI). Global activation patterns corresponding to spoken language processing could be identified as three different sub-networks. The left frontal cortex, structures of the temporo-parietal junction bilaterally and the cingulate gyri are part of the semantic processing of speech. The acoustic processing is reflected in connections between the temporo-parietal junction and the temporal lobe structures. The perception of unintelligible speech activates a network between the left temporal lobe, the right inferior frontal lobe and the cingulate gyrus.

The different approaches to analyze the fMRI dataset permit the investigation of the research question from different perspectives. The evaluation of local and global brain activation patterns, obtained from the analysis of a single dataset, supplement the respective other analysis approach.

Zusammenfassung

Gesprochene Sprache zu verstehen ist eine bemerkenswerte Fähigkeit des menschlichen Gehirns. Menschen bewältigen diese Aufgabe bis heute besser als künstliche Algorithmen, sowohl in herausfordernden akustischen Situationen als auch teilweise für mehrere unterschiedliche Sprachen. Trotz des großen Interesses der Wissenschaft im Bereich der kognitiven Verarbeitung von Sprache, konnte dieser Prozess bisher nicht vollständig nachvollzogen werden. Dies liegt unter anderem an den limitierten und indirekten Möglichkeiten, kognitive Prozesse innerhalb des menschlichen Gehirns zu erfassen. Die Methoden setzen große Datenmengen und die komplexe statistische Verarbeitung der Daten voraus. In den letzten zwei Jahrzehnten wurden viele neue Ansätze entwickelt, um neurokognitive Daten effektiver zu analysieren, wie etwa die multivariate Musteranalyse (MVPA). Die Vielzahl statistischer Methoden zur Analyse neurokognitiver Daten macht eine Vergleichbarkeit unterschiedlicher Studien jedoch schwierig.

Diese Arbeit setzt sich zusammen aus einem methodischen Teil und einem Teil, in dem die Daten einer auditorischen fMRI Studie aus verschiedenen Perpektiven analysiert werden. Im methodischen Teil wird ein neues Verfahren zur Evaluierung der Ergebnisse der Searchlight Classification Analyse vorgestellt, die Searchlight Classification Informative Region Mixture Model Methode. Sie baut auf der Annahme auf, dass die Klassifikationsergebnisse einer solchen Analyse in zwei Unterverteilugen aufgeteilt werden können: eine Verteilung, die Voxel und deren Umgebung repräsentiert, in denen Information über die jeweils betrachteten Konditionen enthalten ist, und eine Verteilung von Voxeln, in denen keine Information enthalten ist. Grundlage dieser Trennung von Verteilungen ist ein Gaussian Mixture Model. Die SCIM Methode wird mit zwei gängigen Referenzmethoden, dem Binomial-Test und dem Permutationstest anhand eines Simulationsdatensatzes und eines fMRT Datensatz verglichen. Es zeigt sich, dass die SCIM Methode sehr viel weniger abhängig von der Wahl des Signifikanzniveaus ist und im Vergleich zum Permutationstest zudem einen sehr viel geringen Rechenaufwand erfordert.

Im zweiten Teil werden die Ergebnisse einer auditorischen fMRT Studie präsentiert, die mithilfe der SCIM Methode und weiteren Methoden erlangt wurden. In vielen Studien, in denen die Verarbeitung von Sprache untersucht wird, wird degradierte Sprache als Kontrastkondition zu normaler Sprache verwendet. Wird klare Sprache verwendet, kann oft nur ein Aspekt dieses komplexen Vorgangs abgebildet werden. Neben vier weiteren akustischen Stimuli beinhaltet die Zusammenstellung der Stimuli dieser Studie eine Kondition mit ganzen, alltäglichen Sätzen und als Kontrastkondition einen Stimulus mit den physikalischen Eigenschaften normaler Sprache, jedoch ohne jeglichen semantischen Inhalt. So kann die Verarbeitung gesprochener Sprache als Ganzes untersucht werden, wobei der Einfluss akustischer Eigenschaften der Stimuli reduziert wird. Die Gruppenergebnisse werden sowohl durch über die Gruppe gemittelte Karten, als auch durch statistische Analysen von einzelnen Probanden-Ergebnissen über die Gruppe hinweg ermittelt. Zudem werden nicht nur separate Kontraste betrachtet, sondern die Korrelation von Erkennungsraten aus der Klassifikationsanalyse mit den jeweiligen Unterschieden bestimmter Eigenschaften von Sprache über die Kontraste hinweg betrachtet. So konnten kortikale Areale ermittelt werden, deren unterschiedliche Aktivierung mit den Unterschieden bestimmter Eigenschaften von Sprache einhergehen.

Die effektive Konnektivität dieser Areale wurde anhand ihrer generellen psychophysiologischen Interaktion (gPPI) bestimmt. Anhand dieser globalen Aktivierungsmuster konnten drei kortikale Netzwerke in Zusammenhang mit der Verarbeitung gesprochener Sprache identifiziert werden. Semantisch sinnvolle Sprache geht einher mit der Aktivierung eines Netzwerks zwischen de Arealen innerhalb des linken Frontalkortex, des bilateralen temporo-parietalen Übergangs und des Gyrus cinguli. Bei der akustischen Verarbeitung zeigt sich eine effektive Konnektivität zwischen den bilateral temporalen Gyri und den Arealen des temporo-parietalen Übergangs. Nicht verständliche Sprache scheint ein Netzwerk zwischen dem linken, inferioren Temporallappen, dem rechten Frontalkortex und dem Gyrus cinguli zu aktivieren.

Die unterschiedlichen Ansätze, die zur Analyse dieses fMRT Datensatzes genutzt wurden, ermöglichen eine Bewertung der Ergebnisse aus unterschiedlichen Perspektiven. Es zeigt sich, dass die Untersuchung lokaler und globaler Aktivierungsmuster die Ergebnisse des jeweils anderen Ansatzes gut ergänzen und erweiterte Möglichkeiten zur Interpretation der Daten bieten.

Contents

List of Figures

List of Tables

List of Abbreviations

fMRI	functional magnetic resonance imaging
BOLD	blood-oxygen-level-dependent
TR	repetition time
TE	echo time
EPI	echo-planar imaging
GLM	general linear model
MNI	Montreal neurological institute
MVPA	multivariate pattern analysis
AAL	automated anatomical labeling
SVM	support vector machine
ROC	receiver-operator-characteristic
AUC	area under the ROC curve
FDR	false discovery rate
GMM	Gaussian mixture model
EM	equalization and maximation
SCIM	Searchlight Classification Informative Region Mixture Model
GÖSA	Göttingen sentence test
ISTS	International Speech Test Signal
NVS	noise vocoded speech
LPC	linear predictive coding
SNR	signal-to-noise ratio
SVD	singular value decomposition

gPPI generalized psychophysiological iteractions

IRM informative region map

CRM correlated region map

2C-GMM two-component Gaussian mixture model

3C-GMM three-component Gaussian mixture model

a anterior

p posterior

i inferior

s superior

m medial

l lateral

AnG angular gyrus

CgG cingular gyrus

MCgG middle cingular gyrus

FuG fusiform gyrus

HG Heschl's gyrus

IFG inferior frontal gyrus

OrIFG orbital part of the inferior frontal gyrus

TrIFG triangular part of the inferior frontal gyrus

OpIFG opercular part of the inferior frontal gyrus

MFG middle frontal gyrus

SFG superior frontal gyrus

INS insular cortex

ITG inferior temporal gyrus

MTG middle temporal gyrus

STG superior temporal gyrus

StmP superior temporal pole

LgG lingual gyrus

PrG precentral gyrus

PoG postcentral gyrus

PCun precuneus

PL parietal lobe

RO rolandic operculum

SMG supramarginal gyrus

Sem semantic speech

$NSem$ non-semantic speech

Sem_N semantic speech in noise

$NSem_N$ non-semantic speech in noise

MO modulation

VP voice pitch

SEM semantic

List of Publications

Chapter 2 is based on the following publication:
A. Urbschat, S. Uppenkamp, and J. Anemüller. Searchlight Classification Informative Region Mixture Model (SCIM): Identification of Cortical Regions Showing Discriminable BOLD Patterns in Event-Related Auditory fMRI Data, *Frontiers in Neuroscience*, 14(616906):1470-1491, 2021. ISSN: 1662-453X. doi: 10.3389/fnins.2020.616906.

1. General Introduction

Spoken language processing is a task on which humans perform remarkably well. The integration from acoustic stimuli to an abstract meaning is very robust even in noisy environments and for distorded or degraded speech. Additionally, the mapping from these acoustic features to a meaning is not specfic. Humans can understand different languages that developed from fundamentally different phoneme groups, different idioms or even non-tonal speech like whispering. The neural correlates of this task, in which humans still outperform computers up to date, are therefore of large interest to scientists in neurocognition.

Early statements about the functional segmentation of the cortex go back to lesion studies. Particulary known became the publications by Paul Broca (1861), in which he postulated a localization of the cortical region that permits the verbal expression in the frontal lobe, and Carl Wernicke (1874), who reported a disfunction in the understanding of spoken language in patients with lesions in the left posterior temporal lobe. The respective cortical stuctures, Broca‘s area and Wernicke‘s area, are named after these researchers and also the dysfunctionalities, called aphasia, are still named after them. The expressive aphasia, the inability to form spoken language, is called Broca‘s aphasia, while the receptive aphasia, the inability to understand spoken language, is called Wernicke‘s aphasia. Later research has provided evidence, that more regions than just Broca‘s area and Wernicke‘s area are involved in spoken language processing (Hagoort, 2014). The classic model of langue processing, the Broca-Wernicke-Lichtheim model, derived in the 19th century, has been overtaken by later research (Ben Shalom and Poeppel, 2008).

The introduction of non-invasive neuroimaging methods like fMRI in the early 1990s (Whitten, 2012) has opened possibilities to heavily increase the number of subjects to examine. The possibility to investigate cortical function in healthy participants was also a crucial step towards the understanding of ”normal“ processes in the brain. These studies provided evidence that the brain structures which are involved in the perception and production of spoken language are not as well defined as assumed by the lesion studies by Broca and Wernicke (Tremblay and Dick, 2016).

Visual modalities are very attractive for fMRI since the projection from the retina to the visual cortex is quite linear (Morland et al., 2001) and, therefore, largely intuitive. But also for the auditory modality fundamental representations have been found. There is, e.g., a topology representing spectral cues with spatial location in the cochlea but also in the auditory cortex (Wessinger et al., 1997).

Sinusoids are represented in the tonotopy in the auditory cortex with respect to the frequency of their periodic sound pressure level (SPL) changes. With decreasing frequency, temporal amplitude modulations are perceptable and are represented along the auditory pathway in different stages (Giraud et al., 2000) up to the auditory cortex including the primary auditory cortex in the medial Heschl’s gyrus. The recognition of an auditory stimulus as speech induces specific activation in the lateral part of the Heschl‘s Gyrus (Uppenkamp et al., 2006) that showed no activation for non-speech sounds. With increasing information encoded in a speech signal, the corresponding neural activation can be located increasingly in areas around the primary and secondary auditory areas.

Since communication via spoken language is so crucial for the social interaction between humans, neural correlates of language processing can be found in broad networks across the cortex. In many different studies neural activations are ivestigated, which correspond to very fundamental tasks like the processing of temporal modulation over the processing of more speech specific characteristics like formants or a voice pitch to abstract representations like semantic content or syntactic structure. While the acoustic processing is understood very well thanks to multiple studies that examined these processes from different perspectives, the abstract interpretation of spoken language is not fully understood yet.

There are some models, that summarize the current status of research on spoken language processing like those by Hickok and Poeppel (2007) and Friederici (2002). The models propose an acoustic processing of spoken language in the superior temporal lobes bilaterally, while the activation corresponding to further, more abstract processing is located left lateralized in the frontal lobe. Further research on this topic has enlarged these models with increased informational detail. However, the research on spoken language processing, in particular the processing of sentences as they occur in everyday context, is still limited in the possibilities of evaluating the cognitive processes due to two major obstacles:

1) In comparison to studies in the viusal domain, the amount of data that can be collected in auditory fMRI studies is rather small due to the information that can be mediated within a specific time frame and due to the sparse imaging paradigm that is often used for auditory fMRI studies. This paradigm is benefical for auditory studies, since it temporally separates the scanner noise from the target stimuli (Edmister et al., 1999; Hall et al., 1999). On the other hand, sparse imaging leads to long times of repeat (TR) and simultaneously to smaller data sets. Therefore, robust methods for the identification of regions that are involved in the processing of complex auditory stimuli are required. Norman et al. (2006), Kriegeskorte et al. (2006) and others proposed multivariate pattern analyses (MVPA) as an alternative for the univariate general linear model (GLM, Friston et al., 1995). The statistical power of highly noisy single voxel time series is accumulated to increase the sensitivity of the analysis. Even though the method is up to day commonly used, it lacks an established framework to evaluate the outcome from MVPA methods like the searchlight classification that is one very intuitive approach to evaluate local blood oxygen level dependent (BOLD) patterns (Etzel et al., 2013). The usage of different methods, evaluation tools and thresholds (ranging between p <0.001 to p <0.05) increases the variability of results obtained from different studies and impede a good comparability of those. However, a good comparability of findings from different studies is essential to understand the highly complex processes of the brain.

2) Spoken sentences are very complex acoustic structures. With increasing complexity of a stimulus, the risk of mixing different characteristics of speech in the conditions increases. Therefore, it is difficult to design a stimulus set that separates the effect of interest from confounds. The number of studies, in which the the neural correlates of spoken language processing on the sentence level have been investigated, is small in comparison to the studies that provide evidence for the models that describe speech processing on the acoustic level or on a phoneme, phrase or word level. Study designs that exclusively include natural spoken language often only permit the investigation of one aspect of spoken language interpretation upon the acoustic processing (e.g., syntax vs. semantic). Degraded speech stimuli that still hold spectral characteristics of speech, like noise vocoded speech (NVS) or spectrally rotated speech, can be manipulated in their intelligibility with regard to the context in which they are presented or the number of channels they have been created with. However, there is no sufficient evidence that the use of degraded speech for the investigation of abstract speech features like semantic content or syntactic structure does not lead to confounds and that the findings from these studies are specific for the effect of interest.

The work described in this thesis shall depict some possibilities how to deal with these obstacles and supplement the current knowledge about the neural spoken language processing in local and global patterns of BOLD activity in the brain.

1.1. The identification of information regions

Multivariate pattern analyses (MVPA) have been proven to have some advantages over mass-univariate analysis methods like the general linear model (GLM) (Norman et al., 2006; Kriegeskorte et al., 2006). One promising possibility to explore neural correlates of specific tasks is the supervised classification of BOLD patterns (Mahmoudi et al., 2012). Instead of statistically evaluating rather noisy single voxel time courses, machine learning models are trained with patterns of multiple voxels. From the classification performance on unseen data the informational content about the specific cognitive task encoded in the BOLD patterns can be inferred. Whole brain maps that reflect the informational content of local BOLD patterns can be derived, e.g., from the searchlight algorithm (Kriegeskorte et al., 2006). Spatially overlapping subsets of imaging data are extracted from the data set and analyzed separately. The classification performance is mapped to the respective center-voxel, providing a map of classification performance values. The further estimation of the statistical power of single searchlight results, however, still lacks a general test method. In **Chapter 2** an algorithm is presented named Searchlight Classification Informative Region Mixture Model (SCIM) that is built on the assumption that the distribution of voxels can be decomposed into informative voxels and non-informative voxels based on their searchlight classification performance value. Subsequent to the searchlight classification and a spatial smoothing of the classification performance maps, the distribution of performance values is decomposed with a two-component Gaussian mixture model (GMM, Dempster et al., 1977). While the non-informative values are distributed around chance level, the informative searchlights build a sub-distribution around higher classification performance values, while both distributions overlap. The separation of informative and non-informative searchlights is finally based on the a-posteriori probability for the non-informative searchlight distribution. Comparable methods based on a GMM have previously been proposed, e.g, by Everitt and Bullmore (1999), for the evaluation of activation maps in mass-univariate analysis algorithms. The reference tests for the evaluation of multivariate searchlight analysis results are the binomial test (e.g., used in Abrams et al., 2012; Akama et al., 2012, 2014; Oosterhof et al., 2010) and the random permutation test (e.g., used in Allefeld and Haynes, 2014; Arsenault and Buchsbaum, 2015; Hausfeld et al., 2014). The binomial test, separating informative searchlights from non-informative searchlights based on their binomial probability of the classifier's performance, is built on strong assumptions about the distribution of performance values, that are already violated when using cross-validation during the classification analysis (Stelzer et al., 2013). The random permutation test has been described to be advantageous to the binomial test (Stelzer et al., 2013; Allefeld et al., 2016; Pereira and Botvinick, 2011). However, it is attached to a very high numerical effort since it requires a large amount of repetitions of the whole classification analysis (usually about 1000 times).

1.2. The disentanglement of stimulus features

The intelligibility of speech cannot be predicted by a single feature of the acoustic stimulus (Arai and Greenberg, 1998). Temporal amplitude modulations are essential for the comprehension of spoken utterances (Riecker et al., 2002). However, they are not sufficient to understand speech (Giraud, 2004), because, amongst other properties, also the formant structure has a considerable impact on speech intelligibility (Lattner et al., 2005; Heinrich et al., 2008). The complex nature of spoken language makes it, accordingly, very difficult to build stimulus sets in which contrasts between the different conditions are limited to one specific feature of speech, in particular when the research question focuses on the more abstract cognitive tasks in spoken language comprehension. The recognition of speech, inferred from vowel perception, is accompanied with neural activation in the lateral part of Heschl‘s gyrus (Belin et al., 2000; Uppenkamp et al., 2006). Subsequently, the information that is encoded in speech is initially represented in words. Different patterns of cortical activation can be found for the perception of existing words in comparison to non existing words as reported by Binder et al. (2000b), Poldrack et al. (1999) and Kouider et al. (2010). Existing words, in turn, are categorized, which is reflected in cortical activation in the parietal lobe (Hwang et al., 2009) and in the inferior frontal cortex (Seghier et al., 2004). The meaning of ambiguous words can be clarified during the comprehension on the sentence level. Many studies, in which the neural correlates of the spoken sentence comprehension are explored, focus on the explicit difference between the processing of semantic content and syntactic structure (Humphries et al., 2006; Friederici et al., 2003; Rogalsky and Hickok, 2009). Of comparable interest is the effect of semantic priming or predicitability on the intelligibility of speech and the corresponding cortical activations (Obleser et al., 2007a; Rothermich and Kotz, 2013; Blank and Davis, 2016).

In **Chapter 3** the results of an fMRI study are presented, which differs in its composition from the above mentioned studies as follows: the stimuli that have been used to compare the semantic and syntactic processing from the physical recognition of speech are based on a natural speech recording. While being physically both similar to natural speech, one condition comprises valid spoken utterances (named semantic speech throughout this work) and the other does not hold any meaningful semantic or syntactic structure (named non-semantic speech in the following). In addition to these two conditions, the study design includes four further acoustic conditions. The idea of comparing classification results obtained from the MVPA across conditions, as done in Okada et al. (2010), is adapted with a correlation analysis across different contrasts between conditions. This permits a more sophisticated separation of different speech features entangled in single conditions.

Another problem that is considered in that chapter, is the high inter-subject variability of functional foci (Poldrack et al., 1999; Laumann et al., 2015). This is, in particular, observable for increasing complexity of cognitive tasks due to increasing individuality of strategies to solve the corresponding task (Seghier et al., 2004, 2008). This problem is accounted for with a statistical evaluation across the group of subjects instead of only presenting group result maps.

1.3. Task related connectivity between cortical structures

Cognitive processes are often not only reflected in a segregated neural activity, but rather in integrational information across spatially distinct cortical areas. The strength of the connection between cortical structures can be derived from two measures, namely structural and functional connectivity. The structural connectivity can be quantified with diffusion weighted imaging methods (Uddin, 2013), reflecting the density of fibers within the white matter of the brain between gray matter cortical structures. Functional connectivity, on the other hand, is derived from temporal coactivation, as inferred in BOLD responses, in separate brain regions (Aertsen et al., 1989; Friston and L. Harrison, 2003). Disregarding the long-term plasticity of the brain, structural connections between cortical regions are anatomical structures that cannot rapidly change for specific tasks. Synchronized spontaneous fluctuations of brain activity in spatially distant brain strcutures, as identified in resting state fMRI, also reflect task-independent correlations of neural activity that give insight to the general functional organization of the brain (van den Heuvel and Hulshoff Pol, 2010). Some synchronized neural activation patterns are, however, specific for cognitive tasks and reflect the interaction between those cortical structures in order to solve the corresponding task. Those functional connections are identified with functional connectivity and effective connectivity. Functional connectivity is inferred by temporal covariances of pairs of voxel or time-series of voxels in a region, often determined by the first component of a singular value decomposition (SVD) of the set of voxel-timeseries in a spatial neighborhood of voxels (Friston2004). The effective connectivity is, on the other hand, represented by the influence of spatially distant structures‘ activations (Friston 1994). One very intuitive model of the several effective connectivity approaches is the model of the psychophysiological interactions (PPI, Friston et al., 1997). The GLM that is fitted to the time-series of voxels or averaged time-series of regions includes the activation of distant cortical areas as regressors. The resulting regression-coefficients depict the influence of the cortical activation in a specific region on the activation in the considered region or voxel. McLaren et al. (2012) expanded this approach to the general psychophysiological interactions (gPPI) that permit the parallel investigation of multiple conditions.

There are many different studies investigating the cortical connectivity during semantic processing in the visual domain, using picture categorization tasks (de Zubicaray and McMahon, 2009) or reading tasks (Hauk et al., 2004; Mashal et al., 2009). However, there are only few studies that focus on the functional or effective connectivity during the abstraction from acoustic input to a meaningful statement. From studies that investigated the segregated activity during spoken language processing, activity in the following cortical structures contributes considerably to this abstract cognitive task, since they have been repeatedly reported to show activation during different speech related tasks: superior temporal lobe (Davis and Gaskell, 2009; Friederici, 2011; Obleser and Kotz, 2010; Heim et al., 2009), the middle temporal gyrus (Davis and Gaskell, 2009; Mashal et al., 2009; Rogalsky and Hickok, 2009), the inferior temporal lobe (Sharp et al., 2009; Akama et al., 2012; Awad et al., 2007), the inferior frontal gyrus (Gonzalez-Castillo and Talavage, 2011; Newman and Twieg, 2001; Rodd et al., 2005), the superior frontal gyrus (SFG Kouider et al., 2010; Birn et al., 2010; Tremblay and Dick, 2016), the middle frontal gyrus (Whitney et al., 2009; Sharp et al., 2009; Heim et al., 2009), the angular gyrus (Obleser et al., 2007a; Mashal et al., 2009; Clos et al., 2014), the supramarginal gyrus (Raettig et al., 2010; Friederici et al., 2010; Richardson and Price, 2009), the fusiform gyrus (von Kriegstein et al., 2003; Davis and Gaskell, 2009; Abrams et al., 2012) and the cingulate gyrus (Evans and Davis, 2015; Smirnov et al., 2014; Obleser et al., 2007a).

In Chapter 4, the gPPI method has been used to investigate the cortical activation during the processing of spoken language in spatially distinct regions and their influence on the respective other regions to identify neural networks that reflect interaction specific for spoken language processing.

1.4. Outline of the dissertation

This work is composed of a methodological part (Chapter 2) and an experimental part (Chapter 3 and 4). The neurophysiological results that are presented troughout this work are based on one auditory fMRI study, investigating the neural correlates of spoken language processing and the comparison of those on an acoustic and an abstract level. To acoount for both, local and global patterns that might be encoded in the BOLD reponses, the data set was investigated with different methods: 1) searchlight classification that permits the identification of local activation patterns and 2) effective connectivity analysis based on generalized psychophysiological interactions that permit the identification of integrational brain patterns, reflecting the interaction between distant brain regions.

In Chapter 2, the mathemathical background of the proposed SCIM method is presented. The comparison of the performances of the SCIM method and the reference methods on a subset of the fMRI dataset (only one contrast has been investigated between semantic speech and non-semantic speech) and the simulation data revealed some advantages of the SCIM method. The results obtained from the different methods are compared with respect to their sensitivity and specificity, their robustness against different applied p-value thresholds and the plausibility of their neurophysiological interpretation. Considering the advantages of the SCIM method that have been observed on a part of the data set, the method is used for the further analysis.

The results obtained from the searchlight classification analysis with the SCIM method are presented in Chapter 3. In addition to the contrast between clear semantic speech and clear non-semantic speech, further contrasts between the BOLD responses corresponding to six acoustic conditions in total are described. To account for the different characteristics of the acoustic conditions, that differ in their content of temporal modulation, human voice pitch and semantic content, the analysis was expanded to a comparison across the contrasts. A Spearman rank-correlation between the quantified differences of the conditions and the corresponding classification performance values of the respective searchlight was used for the indentification of regions that show BOLD responses whose differentiablility correlates with specific differences of speech features across conditions.

In Chapter 4, the results obtained from the comparison of local activation and integrated brain activity are presented. Statistical parametric maps obtained from the standard GLM, which are presented here for the contrast semantic speech and non-semantic speech, are compared to the gPPI analysis. The identification of gPPIs is also based on the GLM method, but includes the activation in different brain regions as potential regressors to explain the temporal variance of cortical activation in the respective voxel or brain region. The results obtained from this analysis, combined with an interpretation of results from several previous studies on the topic of spoken language processing, allow for the identification of separate neural networks for different stages of the processing of spoken language.

The findings that are presented in Chapter 3 and Chapter 4 are put into relation to each other in Chapter 5. The advantages and limitations of the stimulus set and the analysis methods are discussed. An outlook is given which presents possibilities with the proposed SCIM method from Chapter 2 and future studies that might improve and supplement the results presented here.

2. Searchlight Classification Informative Region Mixture Model (SCIM): Identification of Cortical Regions Showing Discriminable BOLD Patterns in Event-Related Auditory fMRI Data

This Chapter is based on the following publication:
A. Urbschat, S. Uppenkamp, and J. Anemüller. Searchlight Classification Informative Region Mixture Model (SCIM): Identification of Cortical Regions Showing Discriminable BOLD Patterns in Event-Related Auditory fMRI Data, *Frontiers in Neuroscience*, 14(616906):1470-1491, 2021. ISSN: 1662-453X. doi: 10.3389/fnins.2020.616906.

Author contributions

SU contributed the fMRI experiment design. AU and SU acquired the fMRI data. AU and JA analyzed the fMRI data and developed the proposed method. AU implemented the algorithms and analyses, and wrote the initial manuscript. JA proposed the GMM and a-posteriori model and wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

2.1. Abstract

The investigation of abstract cognitive tasks, e.g., semantic processing of speech requires the simultaneous use of a carefully selected stimulus design and sensitive tools for the analysis of corresponding neural activity, that are comparable across different studies investigating similar research questions.

Multi-voxel pattern analysis (MVPA) methods are commonly used in neuroimaging to investigate BOLD responses corresponding to neural activation associated with specific cognitive tasks. Regions of significant activation are identified by a thresholding operation during multivariate pattern analysis, the results of which are susceptible to the applied threshold value. Investigation of analysis approaches that are robust to a large extent with respect to thresholding, is thus an important goal pursued here.

The present paper contributes a novel statistical analysis method for fMRI experiments, searchlight classification informative region mixture model (SCIM), that is based on the assumption that the whole brain volume can be subdivided into two groups of voxels: spatial voxel positions around which recorded BOLD activity does convey information about the present stimulus condition and those that do not. A generative statistical model is proposed that assigns a probability of being informative to each position in the brain, based on a combination of a support vector machine searchlight analysis and Gaussian mixture models.

Results from an auditory fMRI study investigating cortical regions that are engaged in the semantic processing of speech indicate that the SCIM method identifies physiologically plausible brain regions as informative, similar to those from two standard methods as reference that we compare to, with two important differences. SCIM-identified regions are very robust to the choice of the threshold for significance, i.e., less “noisy”, in contrast to, e.g., the binomial test whose results in the present experiment are highly dependent on the chosen significance threshold or random permutation tests that are additionally bound to very high computational costs.

In group analyses, the SCIM method identifies a physiologically plausible pre-frontal region, anterior cingulate sulcus, to be involved in semantic processing that other methods succeed to identify only in single subject analyses.

2.2. Introduction

Multi-voxel pattern analysis is a tool that has been established in functional magnetic resonance imaging (fMRI) analyses investigating acquired data obtained from cognitive studies. The approach provides multiple advantages compared to conventional univariate analyses strategies, e.g. general linear models (GLM, Friston et al., 1995) due to its' higher sensitivity (Norman et al., 2006). Information from comparably weak functional BOLD signals in single voxels are accumulated to better discriminable patterns of BOLD responses, which can increase the statistical power (Kriegeskorte et al., 2006). However, a standard for evaluation and interpretation of outcomes from theses multivariate analyses has not been established yet. Since the statistical nature of results from multivariate analyses (e.g., classification accuracies or area under the ROC curve for classification analyses) differs from those obtained by univariate analyses (e.g., z-scores, t-scores, beta-values), different statistical tests need to be applied to distinguish statistically significant results. In this paper we present the searchlight classification informative regions mixture model (SCIM) algorithm, a procedure to statistically evaluate multivariate pattern analysis (MVPA) results obtained from fMRI data that is robust against threshold choices while being less computationally expensive in comparison to commonly used random permutation tests.

To identify cortical regions that show distinguishable BOLD patterns for contrasted conditions, one method is the searchlight classification algorithm (Kriegeskorte et al., 2006). Local patterns of BOLD responses in spherically shaped spatial data subsets are evaluated in a classification analysis, resulting in three dimensional maps representing the local informational content about the contrasted conditions, with a classification performance value for each searchlight‘s center-voxel. To separate informative searchlight regions from those without information, classification results need to be tested for statistical significance. Different approaches have been used for evaluating classification performance results.

In some neuroimaging studies the single subjects results of secondary interest compared to group-level analyses due to the high variability across humans. In these cases one approach to extract informative regions from searchlight analyses are voxelwise t-tests across subjects for the classification performance against chance level (Bode and Haynes, 2009; Carlin et al., 2012; Kahnt et al., 2010). Since the number of subjects and therefore the samples per test are limited to low numbers in most studies, this approach for information-like measures was often criticized (Stelzer et al., 2013; Brodersen et al., 2013; Allefeld et al., 2016).

Under the null hypothesis that a classifier cannot find information about differenes between two conditions in the BOLD data for underlying cognitive tasks, the classification of these data can be modelled as a Bernoulli trial, resulting in a binomial distribution for n independent tests, requiring independence of trials (Pereira et al., 2009). The binomial test was utilized by multiple fMRI studies (Abrams et al., 2012; Akama et al., 2012, 2014; Oosterhof et al., 2010). Most fMRI MVPA studies, however, compensate for the low number of trials accessible per subject with cross-validation in the classification analysis, violating the independence of trials in the analysis leading to too optimistic results in statistical evaluation with the binomial test. Random permutation test are a frequently used alternative to binomial tests (Allefeld and Haynes, 2014; Arsenault and Buchsbaum, 2015; Hausfeld et al., 2014) motivated by the few assumptions about the data they require (Stelzer et al., 2013; Allefeld et al., 2016; Pereira et al., 2009; Pereira and Botvinick, 2011). Under the assumption that data samples are independent of class labels, the null hypothesis expects the original classification performance to be drawn from a distribution derived by repetitions of classification analysis with randomly permuted class labels. The probability for the null hypothesis, respective, p-value is determined by the number of permutations that lead to an equally high or higher classification performance than the original analysis. However, the smallest p-value that can be achieved is one divided by the number of repetitions. Due to the high dimensionality of fMRI data, these test are computationally very expensive.

Instead of artificially creating a distribution of classification performance values that are obtained from classification of non-informative searchlight volumes by permutation of class labels, we propose to use the assumption that, for cognitive tasks, only specific brain regions will be involved while large cortical regions remain unaffected. The distribution of classification performance values obtained from all searchlight regions from the brain can then be decomposed into a non-informative searchlight distribution and an informative searchlight distribution with a two-component Gaussian mixture model (GMM, Dempster et al., 1977), assuming a Gaussian nature of the sub-distribution due to the high dimensionality of the searchlight numbers (about 10^5 searchlights/voxels respectively). In mass-univariate fMRI analyses similar approaches have been applied to decompose activated voxel distributions and non-activated voxel distribution, using, e.g, fundamental power frequency for decomposition (Everitt and Bullmore, 1999; Vincent et al., 2010; Hartvig and Jensen, 2000) or activation clusters (Kim et al., 2010; Penny and Friston, 2003; Oikonomou and Blekas, 2013). Pendse et al. (2009) used three-component Gaussian mixture models for this purpose - in addition to the non-activated and activated distribution, and they assumed a deactivated distribution with a decreased BOLD response for specific conditions. Given the non-directional nature of classification results, we propose to apply a two-component GMM for MVPA results.

A reliable and robust statistical evaluation is of increasing importance for the investigation of rather

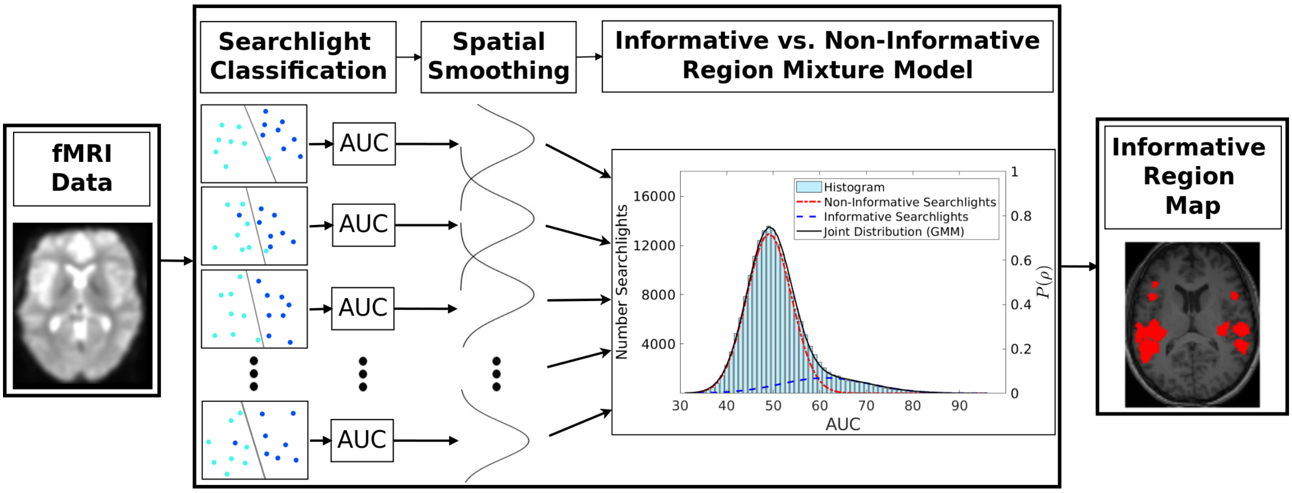

Figure 2.1.: Proposed searchlight classification informative region (SCIM) algorithm procedure. Subsequent to fMRI data classification with the searchlight algorithm, the resulting area-under-curve (AUC) performance values are spatially smoothed and decomposed into a non-informative and an informative searchlight distribution using a two-component GMM. Searchlights with a-posteriori probability for the informative distribution above threshold, equivalent to the non-informative distribution posterior below threshold, define the informative region map (IRM).

complex and abstract cognitive tasks, e.g., the semantic interpretation while listening to spoken language. We therefore show the applicability of the proposed method not only on artificial simulation data but also on data from an auditory fMRI study, investigating the differences in cortical regions involved in the processing of semantically valid speech utterances compared to an acoustic signal that is physically identical to normal speech but without any semantic content. The amount of literature covering speech processing by humans indicates the importance of this topic for research on human communication. Studies range from very fundamental tasks like voiced pitch or vowel perception (Uppenkamp et al., 2006; Liebenthal et al., 2010; Formisano et al., 2008) and speech recognition to more abstract tasks like phoneme recognition and finally semantic interpretation on a lexical word level (LoCasto et al., 2004; Handjaras et al., 2016) and sentence level (Friederici et al., 2000). However, abstract tasks require very careful study designs and more research is required to obtain reliable results to understand human communication basics.

The comparison of the proposed SCIM method to the binomial test shows a high robustness of SCIM against threshold choices, leading to similar results to those obtained by the frequently proposed permutation test. However, the computational cost is considerably reduced, allowing for an increased number of comparisons of conditions and better insights into cognitive processes.

2.3. Methods

2.3.1. Algorithm architecture

The proposed algorithm computes for each voxel the a-posteriori probability of how likely it is that the small brain volume surrounding this voxel conveys information about the experimental condition. The resulting three-dimensional probability map is subsequently referred to as the informative region map (IRM). Fig. 2.1 provides an overview of the algorithm's main steps, described in detail in subsequent sections.

A-posteriori probabilities are computed based on a two-component Gaussian mixture model, which models the distribution of decoding accuracy (area-under-curve values, AUC) across spatial analysis positions. Decoding is performed using searchlight classification with linear support vector machines (SVM), from which the classification accuracy for each voxel position, averaged across all stimulus presentations, is obtained.

2.3.2. Informative vs. non-informative region mixture model

We hypothesize that brain volumes, containing results from searchlight analyses, can be divided into two populations: (1) searchlights that carry information about a condition contrast and (2) searchlights that do not contain this information. Classification performance of the former is expected to be on average higher than chance level, albeit it may fluctuate considerably among informative searchlight volumes. Classification performance associated with the non-informative searchlight volumes, instead, necessarily fluctuates around chance level.

To construct a generative probabilistic model that reflects the diversity of observed classifier performance within the two groups, we adopt a two-component Gaussian mixture model where one mixture component models the non-informative searchlight distribution and the second component models the informative distribution.

Component distributions for searchlight area-under-curve (AUC) performance values, shown in Fig. 2.2, for the informative ($\mathcal{N}_I$, blue) and non-informative ($\mathcal{N}_N$, red) components are computed from the whole brain AUC histogram using the expectation-maximization (EM) algorithm (Dempster et al., 1977). The underlying mixture model links the component distributions to the joint distribution $\mathcal{P}$ (black) according to

$$\begin{aligned}\mathcal{P}(\rho_k|\mu_I,\mu_N,\sigma_I,\sigma_N,\pi_I,\pi_n) &= \pi_I\mathcal{N}_I(\rho_k) + \pi_N\mathcal{N}_N(\rho_k)\\ &= \pi_I\mathcal{N}(\rho_k|\mu_I,\sigma_I) + \pi_N\mathcal{N}(\rho_k|\mu_N,\sigma_N),\end{aligned} \tag{2.1}$$

where ρ_k is the classification performance AUC reached by the k-th SVM classifier, operating on the k-th searchlight volume. Estimated values of prior probabilities π_I, π_N, distribution means μ_I, μ_N, and standard deviations σ_I, σ_N are obtained from subsequent iterations of expectation-step (E-step) and the maximization-step (M-step) of the EM-algorithm that maximizes the logarithmic likelihood function

$$\begin{aligned}&\ln\mathcal{P}(\{\rho_k\}_{k=1}^{K}|\mu_I,\mu_N,\sigma_I,\sigma_N,\pi_I,\pi_n)\\ &=\sum_{k=1}^{K}\ln\left\{\pi_I\mathcal{N}(\rho_k|\mu_I,\sigma_I)+\pi_N\mathcal{N}(\rho_k|\mu_N,\sigma_N)\right\}.\end{aligned} \tag{2.2}$$

The a-posteriori probability for the k-th searchlight to belong to the subset $\mathcal{C}_I$ of informative searchlight volumes is given by

$$\begin{aligned}&p(k \text{ informative}|\rho_k) \equiv p(\mathcal{C}_I|\rho_k)\\ &= \frac{p(\mathcal{C}_I)p(\rho_k|\mathcal{C}_I)}{p(\mathcal{C}_I)p(\rho_k|\mathcal{C}_I)+p(\mathcal{C}_N)p(\rho_k|\mathcal{C}_N)}\\ &= \frac{\pi_I\mathcal{N}(\rho_k|\mu_I,\sigma_I)}{\pi_I\mathcal{N}(\rho_k|\mu_I,\sigma_I)+\pi_N\mathcal{N}(\rho_k|\mu_N,\sigma_N)}.\end{aligned} \tag{2.3}$$

Conversely, the probability of k being from the subset $\mathcal{C}_N$ of non-informative searchlight volumes is

$$\begin{aligned} p_{\mathrm{SCIM}}(k) &\equiv p(k \text{ non-informative}|\rho_k) = p(\mathcal{C}_N|\rho_k) \\ &= \frac{\pi_N \mathcal{N}(\rho_k|\mu_N,\sigma_N)}{\pi_I \mathcal{N}(\rho_k|\mu_I,\sigma_I) + \pi_N \mathcal{N}(\rho_k|\mu_N,\sigma_N)} \\ &= 1 - p(\mathcal{C}_I|\rho_k). \end{aligned} \tag{2.4}$$

The latter quantity, p_{SCIM}, is used throughout the manuscript since it facilitates comparison with the classic p-value of reference methods that denotes the probability of accepting the Null-hypothesis. Thus, searchlight volumes with p_{SCIM} below threshold indicate informative searchlight volumes and in their entity constitute the informative region map (IRM).

2.3.3. Searchlight classification

The searchlight algorithm requires the spatial division of the data set into overlapping, near spherically shaped searchlight volumes centered around each voxel, the center-voxel of the respective searchlight sphere, with the radius set to three voxels. BOLD activations of voxels within a particular searchlight span a multidimensional feature-space vector $\mathbf{x}$ from which the corresponding experiment condition label $y = \pm 1$ is to be predicted.

Subsequent to all searchlight volumes being analyzed independently of each other, classification performance results are mapped to the respective center-voxel of each searchlight, resulting in a three-dimensional information-based map (Kriegeskorte et al., 2006) that reflects the information conveyed within local BOLD regions about the experimental contrast. Information-based maps reflect the informational content in local BOLD patterns based on their separability in a high-dimensional feature space. The absolute activation strength is not important for the interpretation.

2.3.4. Support vector machine

The classification analysis was based on a linear support vector machine analysis (SVM, Schölkopf and Smola, 2001), which is a suitable and robust classification method for fMRI data (e.g., Misaki et al., 2010). A logistic regression model in pilot experiments led to comparable but slightly lower classification performance results. SVMs are discriminative classifiers, finding a separating hyper-plane in feature space with maximum distance to the respective class-samples. The resulting model is parameterized by an optimum weight-vector $\mathbf{w}^*$ that projects data-samples $\mathbf{x}_i$ orthogonally to the separating hyper-plane. To allow for overlapping class-distributions, the soft-margin linear SVM solution is obtained by minimizing a cost function that includes the projection term $\mathbf{w}^T\mathbf{x}_i$ as well as a regularization term $\mathbf{w}^T\mathbf{w}$, resulting in the optimum

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \Big(\frac{1}{2}\mathbf{w}^T\mathbf{w} + C \sum_{i=1}^{l} \max(1 - y_i \mathbf{w}^T \mathbf{x}_i, 0)^2 \Big). \tag{2.5}$$

To avoid overfitting, the regularization parameter C is determined from experimental data by nested cross-validation in which an inner cross-validation loop is employed to find the optimal C through grid-search, and an outer cross-validation loop repeatedly estimates classifier performance on held-out data.

2.3.5. Area under the curve (AUC) analysis

Performance of the classifier at each searchlight's spatial position is measured as the area under the curve (AUC), a quantity that is independent of a specific classifier threshold value since it is computed by integrating the area under the receiver operating characteristic (ROC) curve of true- and false-positive rates. AUC has been shown to provide a reliable performance measure with advantageous properties in a number of classification problems, as confirmed by, e.g., Bradley (1997), and can be interpreted as the probability of a correct classifier decision in a pairwise comparison task of one positive and one negative example being drawn at random from the data ensemble (Green and Swets, 1966). In a number of analyses performed here (cf. results), the overall accuracy measure of the percentage of correct classifications is used as an alternative to AUC, to investigate the impact on the overall SCIM system's analysis results.

2.3.6. Spatial smoothing

To decrease the effect of inter-individual anatomical differences across participants and to avoid destroying potential fine grained structure that might support classification, classification performance maps were spatially smoothed with a Gaussian kernel (FWHM $3mm$) instead of a spatial smoothing step during the preprocessing as it is common in multivariate analysis procedures.

2.3.7. Metrics for separation of informative and non-informative distributions

The degree to which our hypothesis of underlying informative and non-informative voxel distributions is fulfilled can be estimated by the separation of the informative distribution (mean μ_I, standard deviation σ_I) and the non-informative distribution (μ_N, σ_N) after the two-component mixture model has been fit to the searchlight AUC performance values. The classic metric for the separation of two normal distributions is the sensitivity index, which is given by

$$d' = \frac{\mu_I - \mu_N}{\sqrt{\frac{1}{2}(\sigma_I^2 + \sigma_N^2)}}. \tag{2.6}$$

The resulting d' values are included as a model selection parameter in the cross-validation procedure for regularization.

We note that a number of separation criteria have been evaluated as alternatives to the sensitivity index, including several mean- and variance-based measures, geometric distribution overlap, and Kullback-Leibler divergence. The corresponding results showed no systematic differences to the d' sensitivity index.

2.3.8. Baseline statistical tests

Previous studies have applied a number of different statistical tests to obtain thresholded result maps from multivariate fMRI analyses, two commonly used tests being the binomial test (Abrams et al., 2012; Akama et al., 2012; Oosterhof et al., 2010) and the random permutation test (Allefeld and Haynes, 2014; Hausfeld et al., 2014; Kumar et al., 2016).

In the binomial test, p-values are computed as the probability for n coincidentally correct classifications in N trials according to Eq. 2.7, with p_T the a priori probability of the target class, i.e., the prior probability of a semantic speech stimulus, and $p_F = 1 - p_T$ the probability of a non-target stimulus,

$$p_{\text{bin}}(k) = \binom{N}{n} p_T^n \, p_F^{(N-n)}. \tag{2.7}$$

The random permutation test is based on repeated application of the entire classification procedure (cf. searchlight classification) on data with label-independent data. In each of N_r repetitions, target

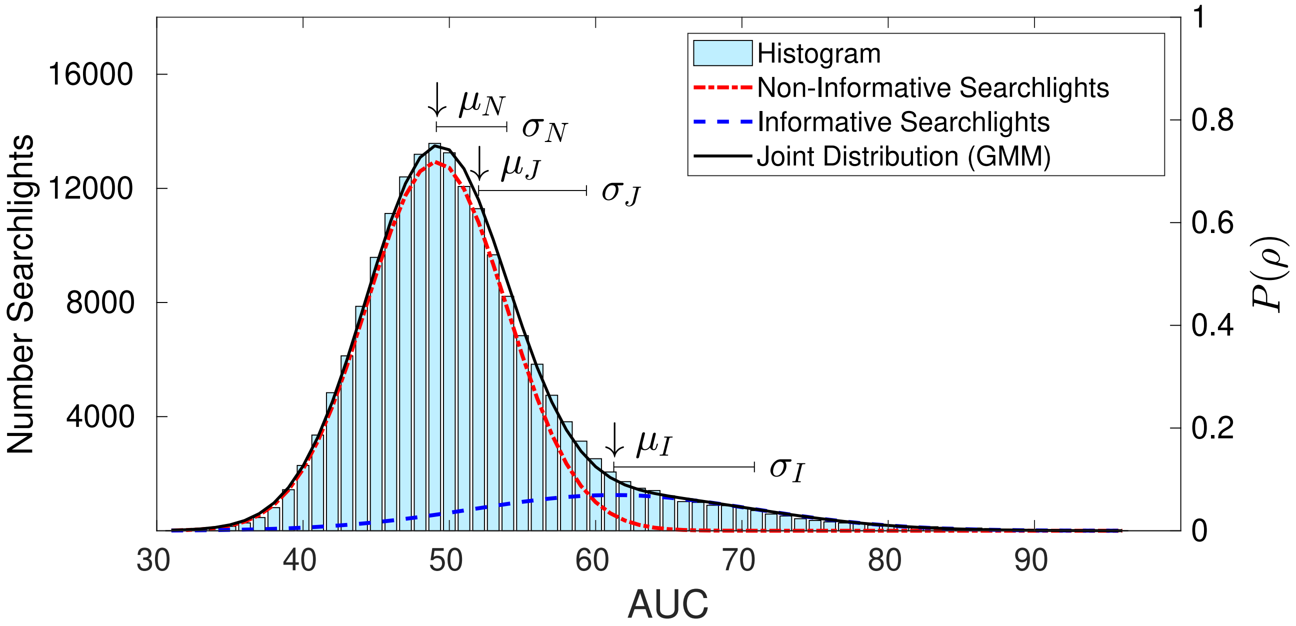

Figure 2.2.: The histogram of searchlight area under the curve (AUC) values in an example map from single-subject results analysis is overlain with the respective GMM and the corresponding informative and non-informative searchlight distributions. Additionally means and variances of distributions as basis of metrics for separation criteria are displayed.

labels are shuffled randomly to simulate independence of samples and targets. p-values are subsequently obtained from the number of repetitions n_h, that led to an equal or higher classification performance than the performance obtained from the original (unshuffled) data set, divided by the total number of repetitions N_r,

$$p_{\mathrm{rp}}(k) = \frac{n_h}{N_r}. \tag{2.8}$$

A direct comparison for the two reference evaluation tests can be found in Stelzer et al. (2013).

2.3.9. Speech stimuli

The aim of the present study is the identification of cortical structures that are engaged in semantic processing of speech. To disambiguate simultaneous physical and semantic stimulus differences that occur, e.g., when contrasting speech with noise, we employed two stimulus sets, semantic and non-semantic speech, that are characterized by largely identical acoustic properties while differing only in the presence vs. absence of semantic meaning.

Non-semantic speech utterances were taken from the “International Speech Test Signal” (ISTS, Holube et al., 2010), originally designed as a test signal for language-independent hearing aid evaluation. ISTS is constructed from speech material from six female speakers with different native languages (Arabic, English, French, German, Mandarin and Spanish), each reading a text in her mother tongue. It has been subdivided into segments of 100 ms - 600 ms duration, that were subsequently rearranged in a pseudo-random order to form a continuous stream of speech utterances. The resulting ISTS generates the percept of nonsense speech that does not contain any semantically valid statements.

Semantic speech stimuli are sentences chosen at random from the Göttingen sentence test (Kollmeier and Wesselkamp, 1997), a speech intelligibility test comprised of phonetically balanced sentences, that each convey a short semantically valid statement. To achieve perceptual comparability to the ISTS, the sentence test's male voice was transformed to a female voice percept by pitch-shifting and digitally changing the vocal tract using the tandem straight (Kawahara and Morise, 2011) method.

2.3.10. Data aquisition

FMRI data were recorded in a 3 T Siemens MRI scanner. 19 subjects participated in the study (11 male, 8 female, 23.5 ± 2.6 years in age), 18 of them with German as mother tongue. The latter participants were considered for further data analysis. All subjects participated voluntarily with an expense allowance.

Subjects were presented with semantic and non-semantic speech stimuli, as described above, in a passive listening paradigm. A sparse imaging design was employed with a time of repetition (TR) of 9 s, including 6.1 s of sound presentation followed by 2.9 s of EPI sequence data acquisition of a complete brain volume in 21 slices with a voxel size of 3.125 mm × 3.125 mm × 3.9 mm, a field of view of 20 cm × 20 cm, a matrix size of 64 × 64 and an echo time (TE) of 55 ms. Sparse imaging allows for the separation of the presentation of auditory stimuli and the scanner noise in time. In addition, the temporal overlap of measured BOLD responses for different stimuli is decreased, which has proved to be an advantage for auditory fMRI experiments (Edmister et al., 1999; Hall et al., 1999) and is also a big advantage for fMRI classification analysis. We note that semantic and non-semantic trial conditions were interleaved with five additional acoustic stimulus conditions, whose analysis is beyond the current scope and will be reported in a subsequent publication. One session comprised 50 min, including four runs with 70 trials (10 trials per condition). A T1-weighted anatomical image was recorded for each participant to allow for localization of resulting active brain regions. Preprocessing including fMRI time series motion correction, realignment and normalization to the standard MNI brain, was performed with SPM8 software (Friston et al., 1995).

2.3.11. Simulation data

Ten simulation data sets have been created to evaluate how accurate the proposed SCIM method and the reference methods can identify regions in a data set that has been manipulated by position information about different conditions at specific locations. These spatial locations define a template map that is compared to the result maps obtained from the different evaluation methods.

A total of 80 experimental trials, 40 each per target and non-target condition, were simulated that carried information about the experimental condition only within a spatially limited template mask area, resembling the SCIM method's informative region map from one subject. In voxel regions outside of the template mask, simulated voxel activations were generated at random from a normal distribution with voxel-wise mean and variance that was identical to mean and variance computed across all experimental fMRI data across target and non-target condition. For voxels within the template mask region, voxel-wise class-specific mean and variance values were identical to mean and variance computed across all experimental fMRI data computed separately for the target and non-target condition, respectively. Obtained normal distribution voxel activations in the template mask region were spatially smoothed with a Gaussian kernel, full-width-half-maximum 3mm, to simulate dependencies across adjacent voxels.

2.3.12. Group level analysis

For all three considered analysis methods, SCIM, permutation test and binomial test, group results are obtained by pooling classification performance values separately for every voxel-position.

For the SCIM method, classification performance results are averaged across subjects, resulting in one map that represents for each voxel the mean classification performance value. Subsequently, the distribution of averaged performance values is decomposed into a non-informative and an informative searchlight distribution, similar to single subject analysis, resulting in a-posteriori probabilities for the non-informative searchlight distribution that are comparable to p-values from other statistical evaluation methods.

Permutation test group results are calculated similar to the proposed method in Stelzer et al. (2013). Classification performance values are averaged voxelwise. For every subject a set of $r = 100$ analysis repetitions with randomized labels is performed and the classification performance results are stored in r separate maps per subject. In the subsequent voxelwise analysis, one random sample from the set of r samples per subjects is selected and the corresponding classification performance at the spatial location is averaged across subjects. This procedure is repeated 100000 times, resulting in a Null-distribution containing 100000 samples per voxel. The resulting p_{rp}-value is calculated as the number of samples within this Null-distribution that are higher or equal to the original average classification performance divided by the number of trials (100000).

For the binomial test, the number of correct samples per searchlight/voxel is summed voxelwise. Now the assumption for one subject that the probability for the null-hypothesis' is equal to the binomial probability for n correct classifications in N samples is adapted to the the sum of all correct classifications $\sum n_m$ in $M \times N$ samples with a group size of M subjects, with n_m the number of correct classifications from the data of subject m. The resulting p_{bin}-value is determined by

$$p_{bin} = \binom{M \times N}{\sum n_m} p_T^{\sum n_m} \, p_F^{(M \times N - \sum n_m)}. \tag{2.9}$$

2.4. Results

2.4.1. Simulations

The reliability of the searchlight classification informative region mixture model (SCIM) was verified with a classification analysis of simulated data. We compared informative region maps (IRMs) obtained from the SCIM method analysis to maps from searchlight classification with subsequent binomial test, random permutation test (n=100 repetitions) and to the template map that underlies the simulation data.

Ten repetitions of simulation data analysis were carried out, with procedures identical to those used for experimental fMRI data, including both smoothed (Fig. 2.3) and unsmoothed AUCs (Fig. 2.4).

Both figures show results from one simulation run for the SCIM method, the random permutation test and the binomial test (red maps), as well as the ground-truth template map (cyan map). After spatial smoothing of the AUC maps (Fig. 2.3), simulation data analyzed with the SCIM method and the random permutation test lead to comparable results. Informative regions obtained from these methods are slightly larger than those in the template map, which can be explained by the searchlight algorithm that spatially smears over information contained in voxels and additional spatial smoothing of AUC maps subsequent to the searchlight classification step. However, the random permutation test result map shows small additional informative regions that are not present in the template map. Results obtained with the binomial method are only valid for high significance thresholds ($p < 0.01$).

For unsmoothed AUC maps (Fig. 2.4), both reference methods, the random permutation test and the binomial test, exhibit informative regions not present in the template map. While these false positive results could be handled with cluster thresholds for the random permutation test results, the binomial test leads to invalid results. Results obtained with the SCIM method on unsmoothed AUC maps are slightly less sensitive compared to those obtained from smoothed AUC maps. Still, in comparison to the reference methods, the SCIM method best reproduces the template map.

The overlap of simulation result maps and the template map was defined as the number of voxels active in template and result maps, relative to the average total number of active voxels,

$$\text{overlap} = \frac{n_{t \cap r}}{0.5(n_t + n_r)}, \tag{2.10}$$

with $n_{t \cap r}$ the number of voxels that were active in both maps, n_t active voxels in the template map, and n_r active voxels in the result map.

The statistical evaluation of simulation analysis results supports the advantage of the SCIM method compared to the reference methods. Spatial overlap of IRMs with the underlying template map for different significance thresholds is shown in Fig. 2.5 with correction for the false discovery rate (FDR, Benjamini and Hochberg, 1995, panel A) and without correction for multiple comparisons (panel B). Medians across ten repetitions are displayed as lines and the inter-quartile range is shown as semi-transparent plane, however, not visible due to very small variance across repetitions. For low p-value thresholds, the overlap values are comparable for the SCIM and the binomial results, while the random permutation test results include no informative regions for very low p-values due to the upper limitation of resulting p_{rp}-values, restricted by the number of repetitions ($p_{min} = \frac{1}{n_{rep}}$). For increasing threshold values, the overlap with results obtained by the reference methods decreases significantly, while the overlap of SCIM results with the template map stays almost constant. Even though p-values larger than 0.05 have little relevance in practice, the corresponding result range is shown for values up to $p = 1$ in order to prove the robustness of the proposed method.

The sensitivity, the specificity and the ROC curves for the different methods are depicted in Fig. 2.6 for result maps with and without correction for multiple comparison. Except for the SCIM method with unsmoothed AUC maps, all methods reach a high sensitivity for p-values larger than 0.01. The difference between the specificity for smoothed and unsmoothed maps in the SCIM algorithm, however, is comparably small. The specificity of the permutation test and the binomial test decreases comparably fast for p-values larger than 0.01. The ROC curves for the corrected tests show an advantageous curve course of the permutation test with the smoothed AUC maps for p-values larger than 0.01. However, for smaller p-values the sensitivity of the permutation test is zero. For unsmoothed AUC maps, the curve courses can be separated in two groups, where the methods are applied to smoothed and unsmoothed maps, respectively.

2.4.2. Single subject results

In this section analyses of single subject results are presented. For the spatial distribution of classification analysis results, single slices from single subject results are displayed for three different participants. Quantitative analyses are performed across all subjects.

Spatial p-value distribution

In Fig. 2.7 the spatial distribution of a-posteriori probabilities from the SCIM analysis and p-values from the random permutation test and the binomial test are displayed for a single slice (at $z = 6mm$) for three single subject results. Transparent slices are located at $p = 0.05$, separating informative from non-informative searchlights for non-corrected analyses. The SCIM analysis provides plateaus of high

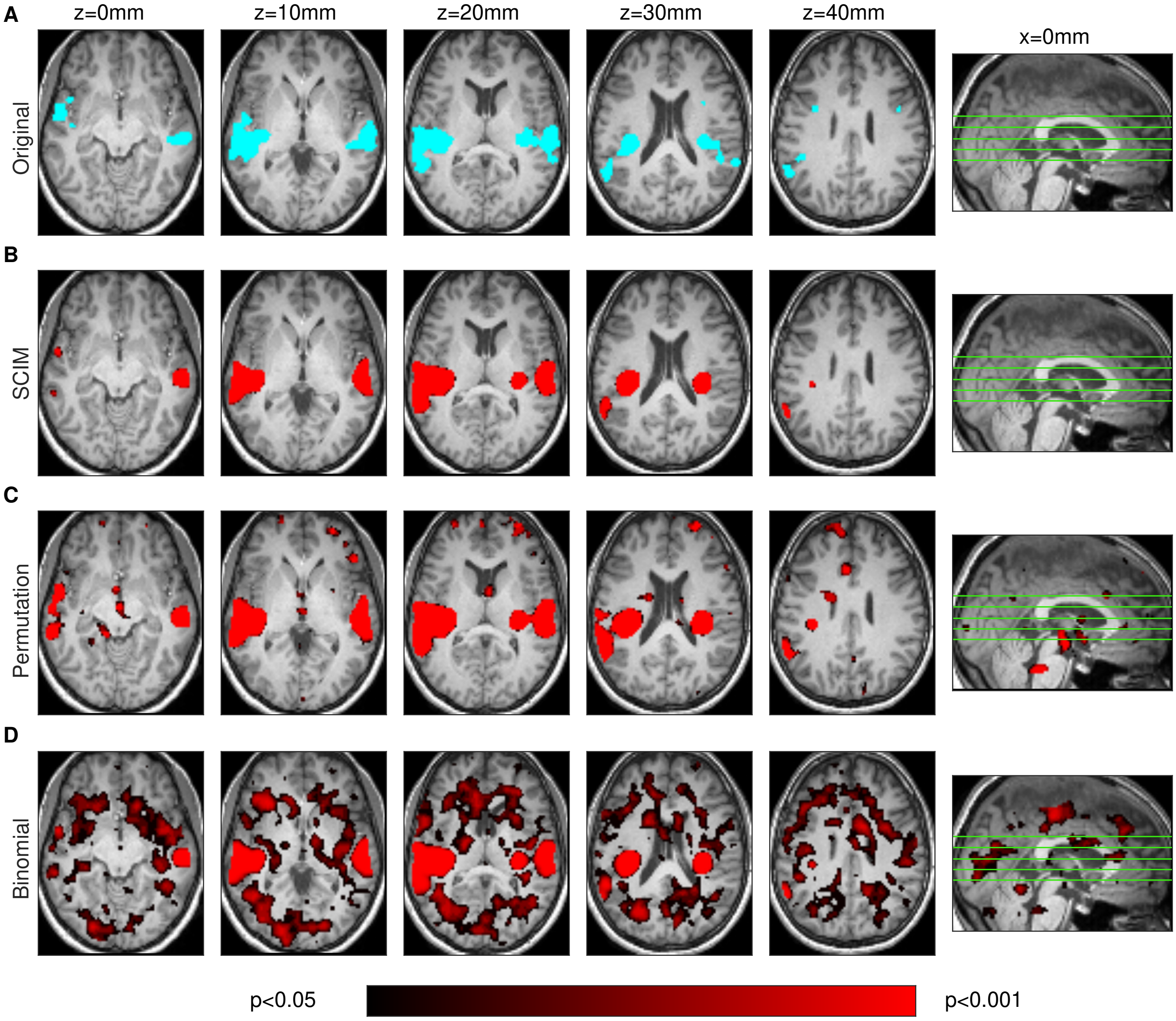

Figure 2.3.: Results obtained from simulated data on smoothed AUC maps. Template map (A) for comparison with simulation result maps obtained from analysis with SCIM method (B), random permutation test (C) and binomial test (D) of spatially smoothed AUC maps. Due to spatial smearing effects based on the searchlight algorithm, results obtained from all methods show larger spatial extent than the template map. The map based on SCIM analysis is most similar to the template map. The map obtained from random permutation test shows larger smearing effects, while binomial test results in informative regions that are not present in the template map. The locations of the transversal slices are depicted on a sagittal slice ($x = 0$).

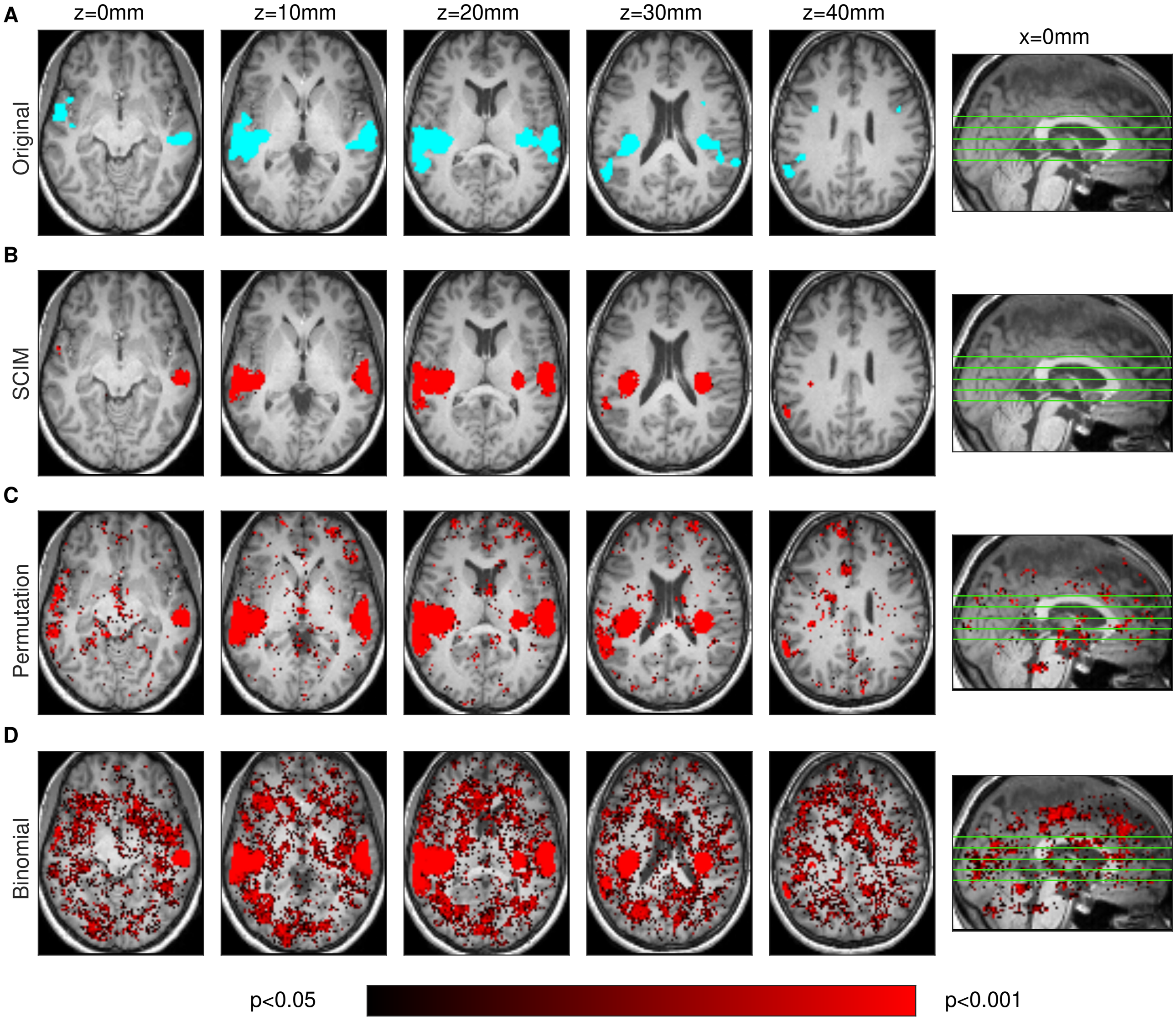

Figure 2.4.: Results obtained from simulated data on unsmoothed AUC maps. The SCIM result maps (panel B) are comparable to those obtained with spatial smoothing (Fig. 2.3) while in the permutation test results (panel C) and in the binomial test results (panel D) numerous small informative regions can be found that are not in line with the template map (panel A).

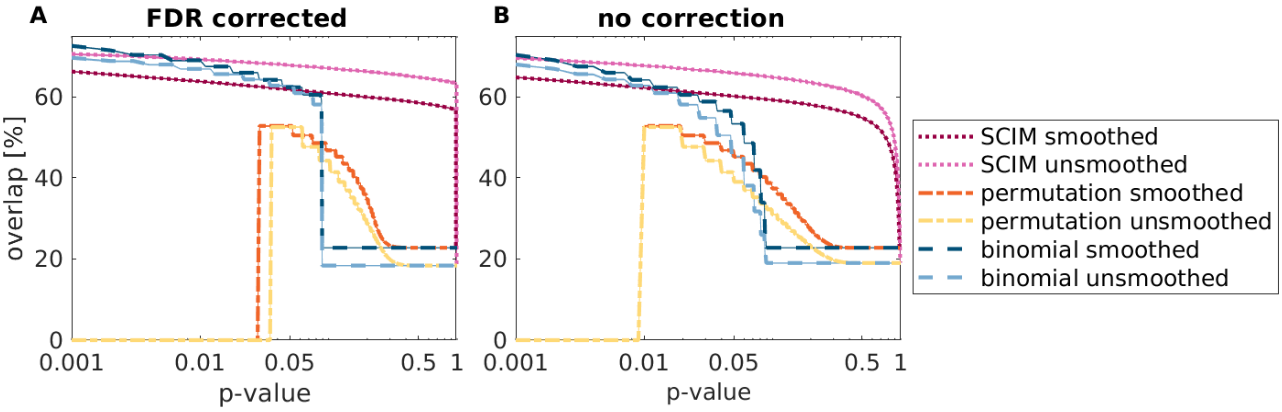

Figure 2.5.: Overlap of simulation result maps with underlying ground truth map obtained from SCIM analysis, random permutation test and binomial test for ten repetitions of simulation analysis and different applied significance p-value thresholds, respectively. Median values across ten repetitions are presented as lines, inter-quartile ranges are displayed as semi-transparent plane but not visible due to the very small variance across repetitions. For very small p-values, SCIM and binomial test results show comparable overlap with ground truth maps. However, overlap decreases for binomial results with increasing p-values, while SCIM results stay almost constant. Result maps obtained from random permutation test show minimum p-values of $p_{rp} = 0.01$ (resulting from 100 repetitions) and exhibit no informative regions for lower p-value thresholds. For p-values higher than 0.5, both reference methods, random permutation test and binomial test, are limited by the additional criterion of $AUC > 0.5$ for searchlights to be informative and overlap values converge to a constant value. For results obtained with the SCIM method, this value is achieved for p-value threshold close to 1.

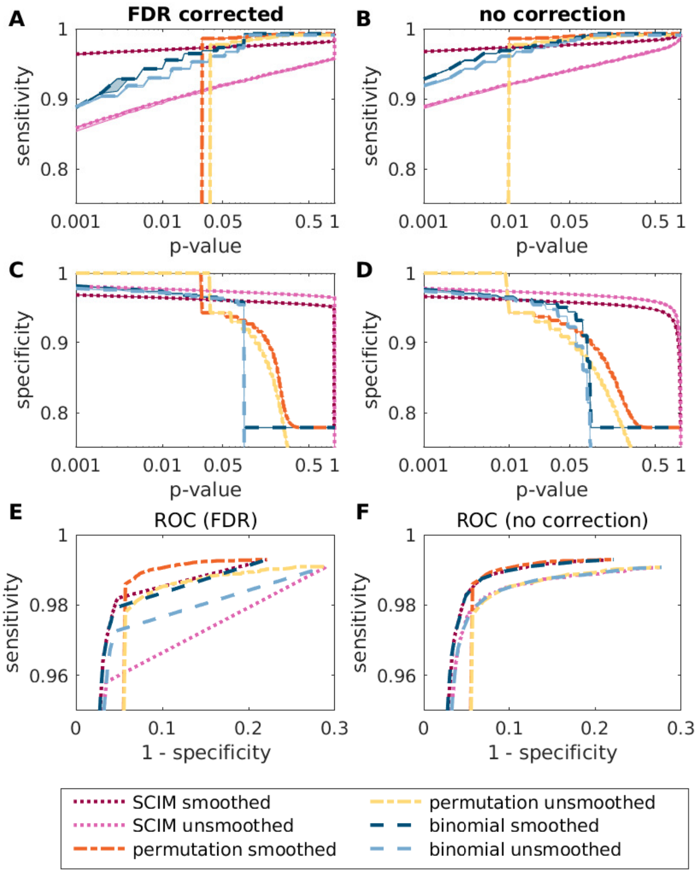

Figure 2.6.: Comparison of the sensitivity (panels A and B), specificity (panels C and D) and ROC curves (panels E and F) for the different methods SCIM, permutation test and binomial test. All methods were tested with smoothed and unsmoothed AUC maps. The results with FDR correction are depicted in the left panels and results without correction for multiple comparison are depicted in the right panels.

significance levels for all displayed single subject results, while non-informative searchlights correspond to regions of significance levels larger than $p = 0.4$, that are not displayed in the plots. The described effects below are true for all slices and all subjects, the plots are limited to one slice of three subjects, respectively, due to space limitations.

Informative regions obtained from permutation tests are predominantly similar to those obtained from the SCIM analysis, with the exception of isolated small informative regions that occur in the permutation test results. However, significance levels for informative and non-informative regions are not as clearly separated as in the SCIM analysis and the presence of the previously mentioned small informative regions is dependent on the applied significance thresholds.

The distribution of p-values resulting from the binomial test analysis shows a gradual transition from informative to non-informative searchlight areas in a narrow interval of p-values. This leads to high dependence of informative regions on the applied significance threshold.

Single subject informative region maps

Single slices (at $z = 6mm$) of IRMs obtained from single subject analyses and the corresponding statistical distribution of results in the whole result map are displayed in Fig. 2.8. Result maps obtained with the SCIM method, the random permutation test and the binomial test were thresholded at $p < 0.05$, respectively. Informative regions obtained without correction for multiple comparisons are colored in red, the corresponding informative regions with FDR correction are colored in orange. Permutation test results show no informative regions after FDR correction for subjects 2 and 3, while with other methods (and for subject 1 also with random permutation) FDR correction leads to slightly decreased sizes of informative regions. Anatomical regions identified to be engaged in the semantic processing task are qualitatively similar for all methods. The right panels of Fig. 2.8 show the histograms of classification performance values across all brain-searchlights, the assumed Null-distribution for the SCIM method (blue line) and the corresponding a- posteriori probabilities for classification performance values obtained with the SCIM method. Black dots show p-values obtained from random permutation tests, and the dashed lines show the Null-distributions for one voxel with high (red), mid (green) and low (blue) classification performance, respectively. Null-distributions obtained from the random permutation test are comparable for all three subjects, while the original distribution of AUC values is different. AUC maps resulting from the analysis of data from Subject 1 (A) are shifted towards lower values, maxima of distributions from Subject 2 (B) and Subject 3 (C) are located at chance level ($p_{chance} = 50\%$). Only searchlights with $AUC > 50\%$ are considered to be informative, and less searchlights satisfying this criterion involve less comparisons to be corrected for in the FDR procedure. Therefore, the FDR corrected IRM for Subject 1 is comparable to the uncorrected one, while IRMs for Subject 2 and 3, with more searchlights being associated to the distribution of searchlights with $AUC > 50\%$ show no informative regions after FDR correction. The dark red dotted lines show the binomial distribution resulting from study design with $N = 80$ samples and $p_{chance} = 0.5$. This distribution also shows p-values for accomplished classification performance results as well as the assumed Null-distribution for the binomial test.

Influence of significance threshold

Quantitative analyses of single subject results are displayed in Fig. 2.9. The left panels show the portion of informative searchlights from all brain searchlights for different applied significance thresholds for the three evaluation methods, (1) SCIM, (2) random permutation and (3) binomial test and classification performance measures, AUC and accuracy. Lines represent the median across subjects, while semi-transparent areas display the inter-quartile range. FDR corrected SCIM analysis results

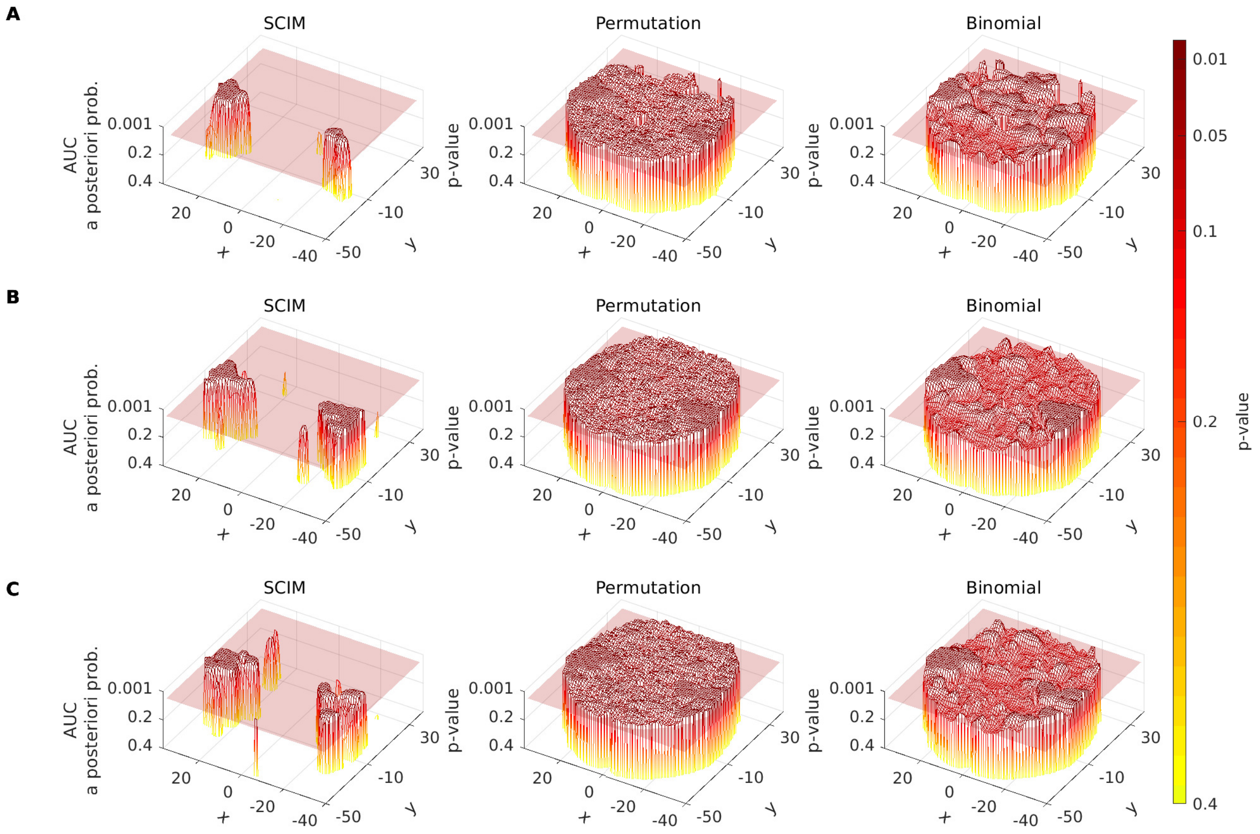

Figure 2.7.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutation test and binomial test) across a single slice ($z = 6mm$, single subject, evaluation measure AUC, spatial smoothing for SCIM, random permutation and binomial) of single subject results from three different subjects. Left panels: distribution of p_{SCIM}-values resulting from the Searchlight Classification Informative Region Mixture Model (SCIM, semi-transparent plane located at $p_{\text{SCIM}} = 0.05$) has plateaus of high significance levels for informative searchlight regions and a low noise floor across non-informative areas. Center panels: p-values resulting from random permutation analysis (semi-transparent plane located at $p_{\text{perm}} = 0.05$). Right panels: The distribution of binomial-test p-values (semi-transparent plane located at $p_{\text{bin}} = 0.05$) shows gradual transition from informative to non-informative searchlight areas in a narrow p-value interval. Differences between informative and non-informative areas are best delineated by the SCIM method and less pronounced with random permutation and binomial methods.

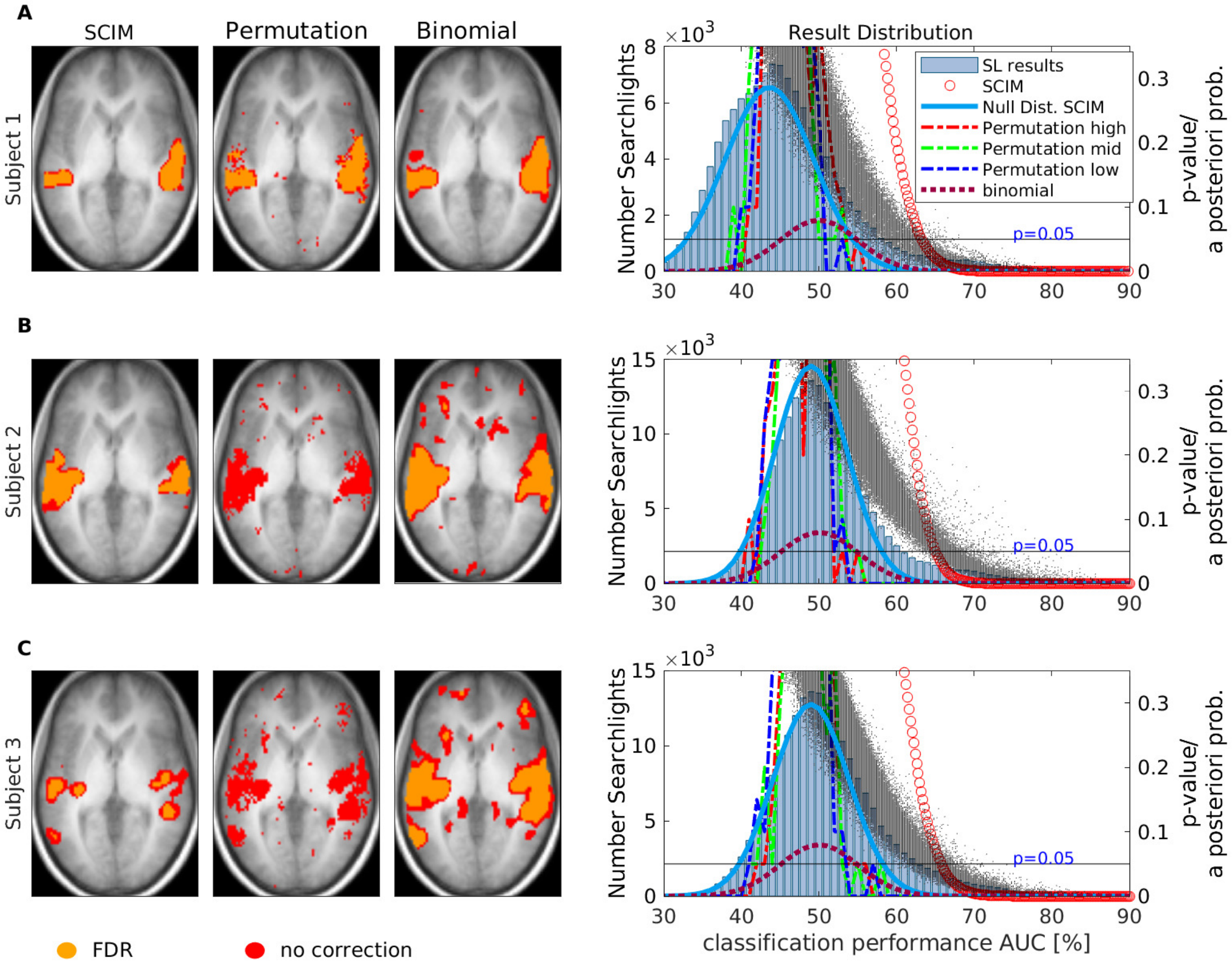

Figure 2.8.: Single slice (at $z = 6mm$) of subject informative region maps (IRMs). IRMs obtained from single subject results with SCIM methods (first column), Permutation test (second column) and binomial test(third column) for three different subjects at a significance level $p_{SCIM} < 0.05$, $p_{rp} < 0.05$ and $p_{bin} < 0.05$, without correction for multiple comparison (red) and with FDR correction (orange), respectively. For subjects 2 (panel B) and 3 (panel C), no informative voxels can be found when FDR correction is applied. For the SCIM method and binomial method, maps obtained with FDR correction exhibit slightly smaller informative regions. Right panels show the results' distribution across voxels. A histogram of AUC values is presented as bar-plot. A-posteriori probabilities obtained from SCIM analysis (red circles) and p-values obtained from permutation test (black dots) and binomial test (dark red dashed line) are displayed, as well as underlying assumed Null-distributions for the different tests, SCIM method (dark blue line) and permutation test with one distribution for a voxel with high (blue), middle (green) and low (red) performance, respectively. For subject 1 (panel A) the distribution peaks for AUC values lower than 50%, while for the other subjects the maximum is located at chance level AUC=50%.

are approximately constant up to threshold levels of $p = 0.1$. For higher thresholds the number of informative searchlights increases only marginally. The number of informative searchlights obtained from the random permutation tests differ across subjects. For higher thresholds than $p = 0.05$ the upper quartile of the group shows a strongly increasing number of informative searchlights with increasing p-value thresholds, while the median of the group shows an almost constant trend, similar to the results obtained with the SCIM method.

Since searchlights that lead to very low classification performance results are associated with low p-values but are not expected to be highly informational, only searchlights with higher classification performance than 50% are considered to be informative. This additional criterion is the limiting factor for thresholds around $p_{rp} = 0.5$ for the random permutation test and $p_{bin} = 0.08$ for the binomial test. For higher significance thresholds all searchlights with a higher performance value than 50% are considered to be informative, irrespective of the exact applied p-value thresholds.

Corresponding statistics for non-corrected maps in the lower left panel show strong dependencies on applied p-value thresholds for both performance measures, AUC and accuracy for the binomial test, and AUC measure for the random permutation test. SCIM method results and random permutation results obtained from the accuracy analysis are nearly constant up to p-values of 0.05. The number of informative searchlights obtained from the SCIM analysis increases slowly for higher thresholds, while IRMs obtained from the random permutation analysis show a sharply increasing number of searchlights with increasing p-value thresholds.

Usual p-value thresholds vary between $p = 0.01$ and $p = 0.05$ across studies. The impact of different applied thresholds for significance is represented in the right panels (B and D) of Fig. 2.9 as the inverse slope of median curves from panels A and C in the range between $p = 0.01$ and $p = 0.05$, respectively. The red lines show the median, the boxes the inter-quartile range, the dashed lines 5%- and 95%-quantiles and the red crosses show outliers. For FDR corrected maps (panel B), the SCIM results show very low differences in this range. Random permutation tests lead to no informative searchlight with FDR correction, except for two subjects, that show larger differences in the mentioned range than SCIM results. Binomial tests show larger differences across thresholds than both, SCIM and random permutation results. Without correction for multiple comparisons (panel D) inverse slopes for SCIM results only differ marginally from those obtained with FDR correction. For the random permutation test on AUC values, the absolute value of the inverse slope and therefore the change of numbers of informative searchlights is significantly larger than for the SCIM analysis. Results obtained with the accuracy measure are comparable for SCIM analysis and random permutation tests. Binomial test results lead to significantly larger absolute values of inverse slopes for both, AUC and accuracy measure.

2.4.3. Group level results

Summary of evaluation methods and classification measures

Fig. 2.10 shows the group level informative region maps (IRMs) obtained with the three different approaches SCIM method, random permutation test and binomial test. Informative regions obtained with an applied p-value threshold of $p < 0.05$ are colored in red, respective informative regions for $p < 0.01$ in dark violet and $p < 0.001$ in pale violet. For AUC measures (panel A) the informative regions do not differ considerably for the different applied thresholds and with or without FDR correction when significance evaluation is performed with SCIM or random permutation test. Binomial test results show additional informative regions to those obtained with the previous two methods in anatomical areas that do not overlap with results known from literature for the investigated cognitive task, both with and without FDR correction. However, informative regions obtained with the binomial test methods with significance thresholds lower than $p = 0.01$ are similar to those obtained with the

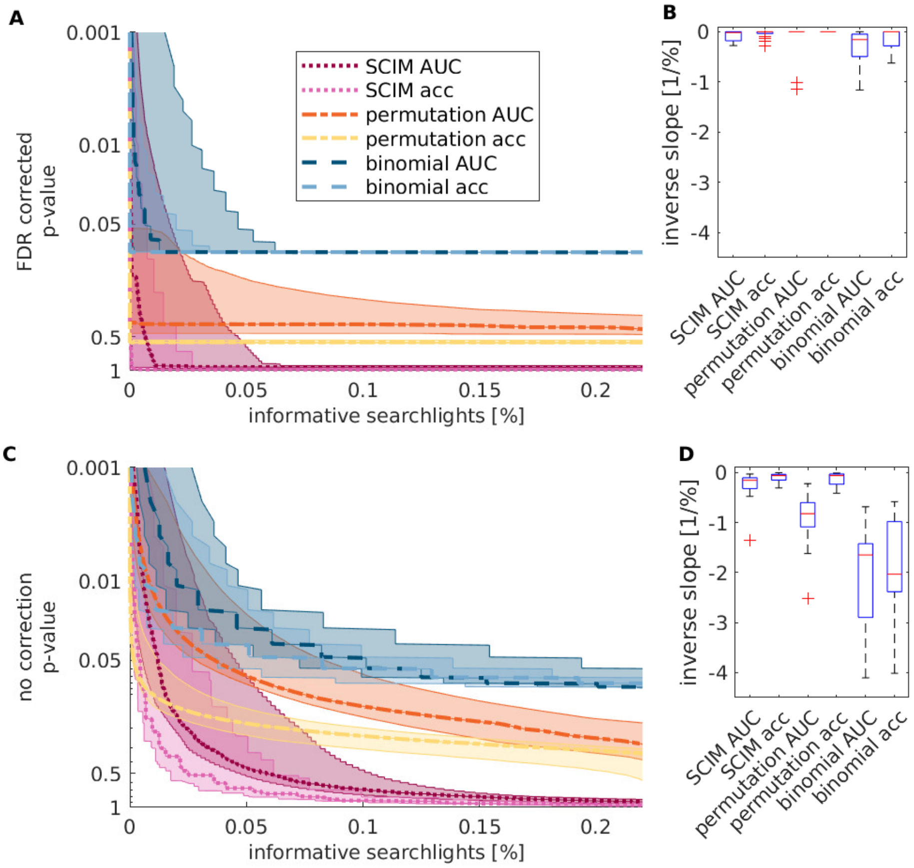

Figure 2.9.: Quantitative analysis of single subject maps across subjects. (A) and (C) Cumulative histograms of fraction on searchlight volumes (in %, abscissa) whose p_{SCIM}-value is below a chosen threshold p_{thr}-value (ordinate), i.e., which are considered informative. Curves indicate group median for SCIM method, random permutation method and binomial method with area-under-curve (AUC) and accuracy (acc) measures, respectively. Semi-transparent areas depict the inter-quartile range. (B) and (D) Average inverse slopes of curves in panels (A) and (C) within the interval $0.05 > p_{\mathrm{thr}} > 0.01$. Upper panels (A and B) show FDR corrected results, lower panels (C and D) show respective non-corrected results. Results indicate that the SCIM method is characterized by a strong separation of informative and non-informative searchlight volumes, both for FDR corrected and non-corrected maps, while results obtained with AUC measurement and random permutation test are highly dependent on the applied thresholds. Binomial test results show this dependency in all cases.

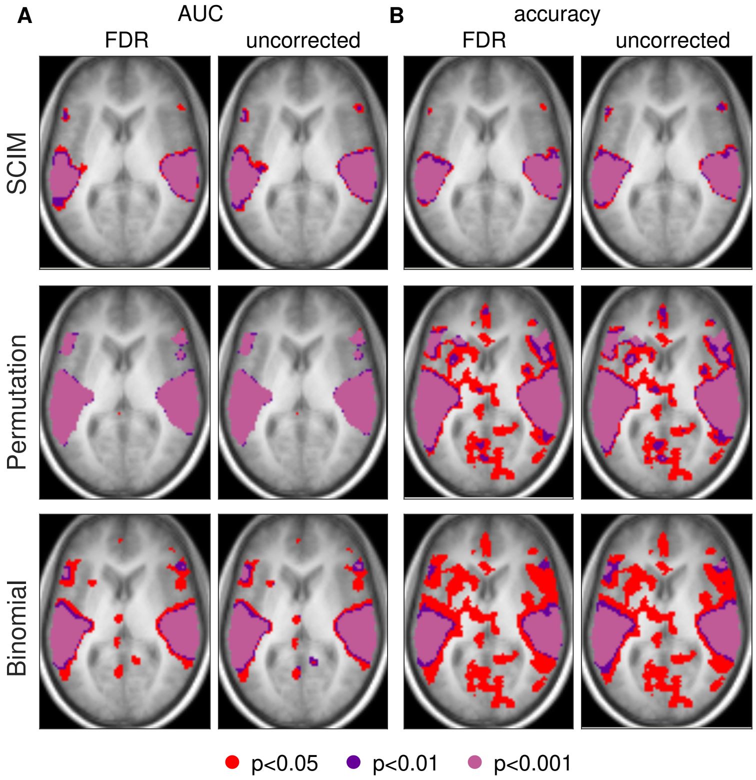

Figure 2.10.: Single slices ($z = 6mm$, respectively) of group maps with AUC measure (panel A) and accuracy measures (panel B). First and third column show results with FDR correction, second and fourth column respective results without correction for multiple comparison. Group results obtained with the proposed SCIM methods are displayed in the first row, random permutation test results in the second row. Third row represents results from binomial test group results. Informative regions at a threshold with $p < 0.05$ are colored in red, respective results for thresholds $p < 0.01$ in dark violet and $p < 0.01$ in light violet.

two other methods. For accuracy measures (panel B), the IRMs obtained with the random permutation test and the binomial test for p-value thresholds at $p < 0.05$ include informative regions in anatomical areas that are not consistent with areas known from the literature. Informative regions obtained with thresholds $p < 0.01$ are consistent with those obtained from AUC analysis and SCIM analysis.

Group level p-value distribution

The spatial distribution of a-posteriori probabilities p_{SCIM} obtained from the SCIM analysis and p_{rp}-values from the random permutation test as well as p_{bin}-values from the binomial test are displayed for a single slice ($z = 6mm$) of group result maps in Fig. 2.11. Other slices show similar effects, but are not shown here due to space restrictions. The semi-transparent slices are located at thresholds of $p = 0.001$. Informative regions, i.e. , the segments above a semi-transparent plane, are comparable for results obtained from SCIM and the random permutation test. However, the random permutation test on accuracy measures leads to just below threshold results that are not as clearly separated from informative regions as compared to results obtained from AUC measures or the SCIM analysis on both measures, AUC and accuracy. The p_{bin}-values obtained from the binomial test lie in a very small value range that does not permit a reliable separation of informative from non-informative regions. While the other methods, SCIM and random permutation test, show spatially smooth plateaus of high significance (respectively low p-values), binomial test results show very homogeneous spatial distributions for both, AUC and accuracy measure.

Statistical distribution of group level results

The distribution of group results obtained with the different approaches can be found in Fig. 2.12 for the AUC measure (A) and the accuracy measure (B). Light blue bars display the histogram of the respective average classification performance results across all subjects. Red circles show the corresponding a-posteriori probabilities obtained with the SCIM method, with less data points for the accuracy measure due to the limited resolution of 80 samples per searchlight. The underlying assumed Null-distributions for the SCIM analyses are shown as dark blue lines. The p_{rp}-values resulting from the random permutation tests are displayed as black dots. In comparison to single subject results (cf. Fig. 2.8), group result random permutation p_{rp}-values show less variance for respective performance measures. However, low p_{rp}-values are associated with lower classification performance values compared to a-posteriori probabilities (p_{SCIM}) obtained from SCIM analyses. For the binomial test, the assumed distribution is equal to the resulting p_{bin}-values for respective classification performance values. These are displayed as dashed dark red lines.

For the AUC measure, the assumed Null-distributions from SCIM analyses and the random permutation analyses do not peak at the expected chance level at $p_{chance} = 0.5$ in contrast to the assumed Null-distribution for accuracy measure. However, while the SCIM Null-distribution is determined by the histogram of AUC values the Null-distribution resulting from random permutation tests does not follow the histogram of achieved AUC values. The binomial Null-distribution, on the other hand, that only depends on the study design but not on analysis results, is distributed around the expected chance level. For the accuracy measure, all Null-distributions are centered around chance level. For both measures, AUC and accuracy, the SCIM analysis provides a more stringent selector for informative regions in classification performance maps than the random permutation test and the binomial test.

Group level informative region maps

In Fig. 2.13 informative regions are displayed for the contrast semantic speech vs. non-semantic speech, emerging from fMRI analysis with the proposed SCIM method based on AUC measure (A), on accuracy measure (B) and the corresponding maps obtained from commonly used random permutation test (C and D).

Informative regions arising from the proposed method (SCIM) overlap with those arising from random permutation test with AUC measure to a large extent, with slightly larger informative regions in the random permutation test results. The location of informative regions obtained from the described analyses are in primary auditory and adjacent regions in Heschl's gyrus and the superior temporal gyrus, in Broca's area in the inferior frontal gyrus region and posterior to the auditory cortex in Wernicke's area. Additionally informative regions for semantic processing have been found in fronto-cortical regions in anterior cingulate gyrus.

The previous statistical evaluation has shown that the statistical power of the binomial test is considerably lower than the power of the SCIM method and the permutation test. Therefore, and in order to be able to display several slices of the real data within a reasonable amount of space, the result maps are focused on those obtained from the SCIM method and the permutation test. The corresponding result maps obtained from the binomial test can be found in the supplementary material.

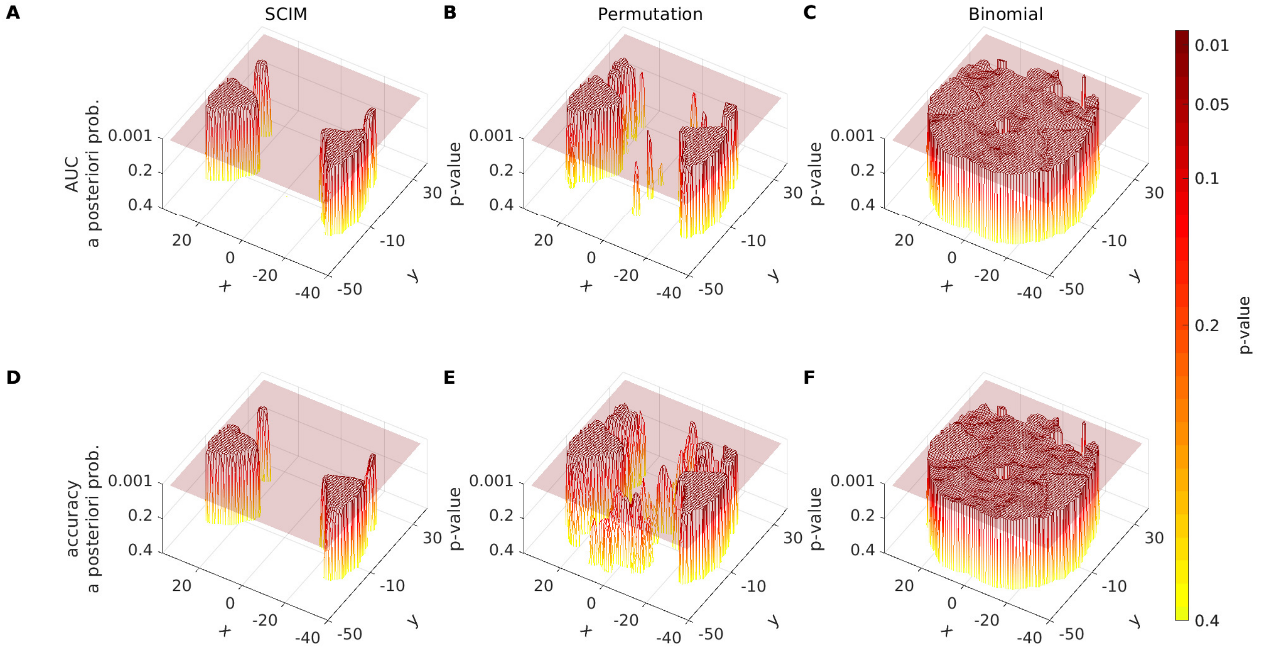

Figure 2.11.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutation test and binomial test) across a single slice from group result maps ($z = 6mm$, group results, evaluation measure AUC (A-C) and accuracy (D-F), spatial smoothing). Left panels (A and D): distribution of p_{SCIM}-values resulting from the Searchlight Classification Informative Region Mixture Model (SCIM, semi-transparent plane located at $p_{\text{SCIM}} = 0.001$) has plateaus of high significance levels for informative searchlight regions and a low noise floor across non-informative areas. Center panels (B and E): p-values resulting from random permutation analysis (semi-transparent plane located at $p_{\text{perm}} = 0.001$). Right panels (C and F): Distribution of binomial-test p-values (semi-transparent plane located at $p_{\text{bin}} = 0.001$) in a very narrow p-value interval. Differences between informative and non-informative areas are best delineated by the SCIM method, however, very similar to those in results obtained from random permutation test. For accuracy measure, the random permutation test exhibits sub-threshold non-informative regions, that are not as well separated from informative regions as compared to results map from AUC analysis or SCIM analysis. Results obtained from the binomial method are almost non-separable into informative and non-informative regions, since the the range of emerging p-values is very small.

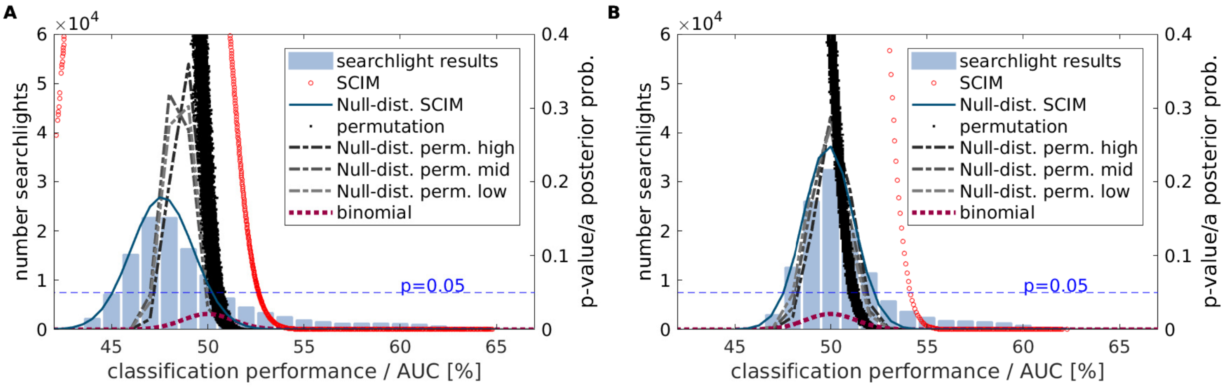

Figure 2.12.: Statistical evaluation of group results based on (A) AUC measures and (B) accuracy measures. Histograms show the distribution of classification performance results emerging in group mean maps. Red circles show a-posteriori probabilities obtained from SCIM analysis for respective classification performance values and the blue line the underlying Null-distribution. In random permutation test p-values are calculated independently for all voxels that are shown with black dots. p-values obtained with the binomial test result from the binomial distribution that also represents the assumed Null-distribution for this test.

2.5. Discussion

This paper presents a novel method for the evaluation of results obtained from multivariate searchlight classification analysis of fMRI data. Simulation data and data from a real auditory fMRI experiment are analyzed with the proposed SCIM method and results are compared to those obtained from two references methods, the random permutation test and the binomial test. The evaluation and comparison of the methods is based on the spatial distribution of obtained p-values, robustness of results for different significance thresholds and classification measures (AUC and accuracy) and consistency with results described in previous studies investigating semantic processing of acoustic stimuli.

2.5.1. SCIM method for a-posteriori probability estimation

The analyses of simulation data in Sec. 2.4.1 confirm the general applicability of the SCIM method and advantages over reference methods, random permutation test and binomial test, when results are compared to ground-truth. All methods reproduce the template map for low p-values with minor differences depending on certain processing stages, in particular spatial smoothing, cf. Fig. 2.5. Without spatial smoothing, the reference methods exhibit false positive informative regions, not included in the template map, while the SCIM method shows largest consistency with the template map for both, smoothed and unsmoothed AUC maps with lower sensitivity but higher specificity with unsmoothed maps. As reflected in Fig. 2.6 the SCIM method has a low sensitivity when it is used without spatial smoothing of the AUC maps and even with spatial smoothing the sensitivity of the reference methods is higher for p-values than 0.01. However, these differences are very small in comparison to the differences in specificity for p-values > 0.01, where the SCIM method outperform the reference methods.

For the experimental data, all three methods successfully identify informative regions, as reflected in low a-posteriori probabilities (p_{SCIM}) for the SCIM method, and low p_{rp}- and p_{bin}-values for the random permutation and binomial methods, respectively. Robustness of the spatial extent of informative region maps (IRMs) with respect to the threshold value applied during analysis, however, is found to be dependent on the choice of the analysis method. While the spatial map obtained with a binomial test

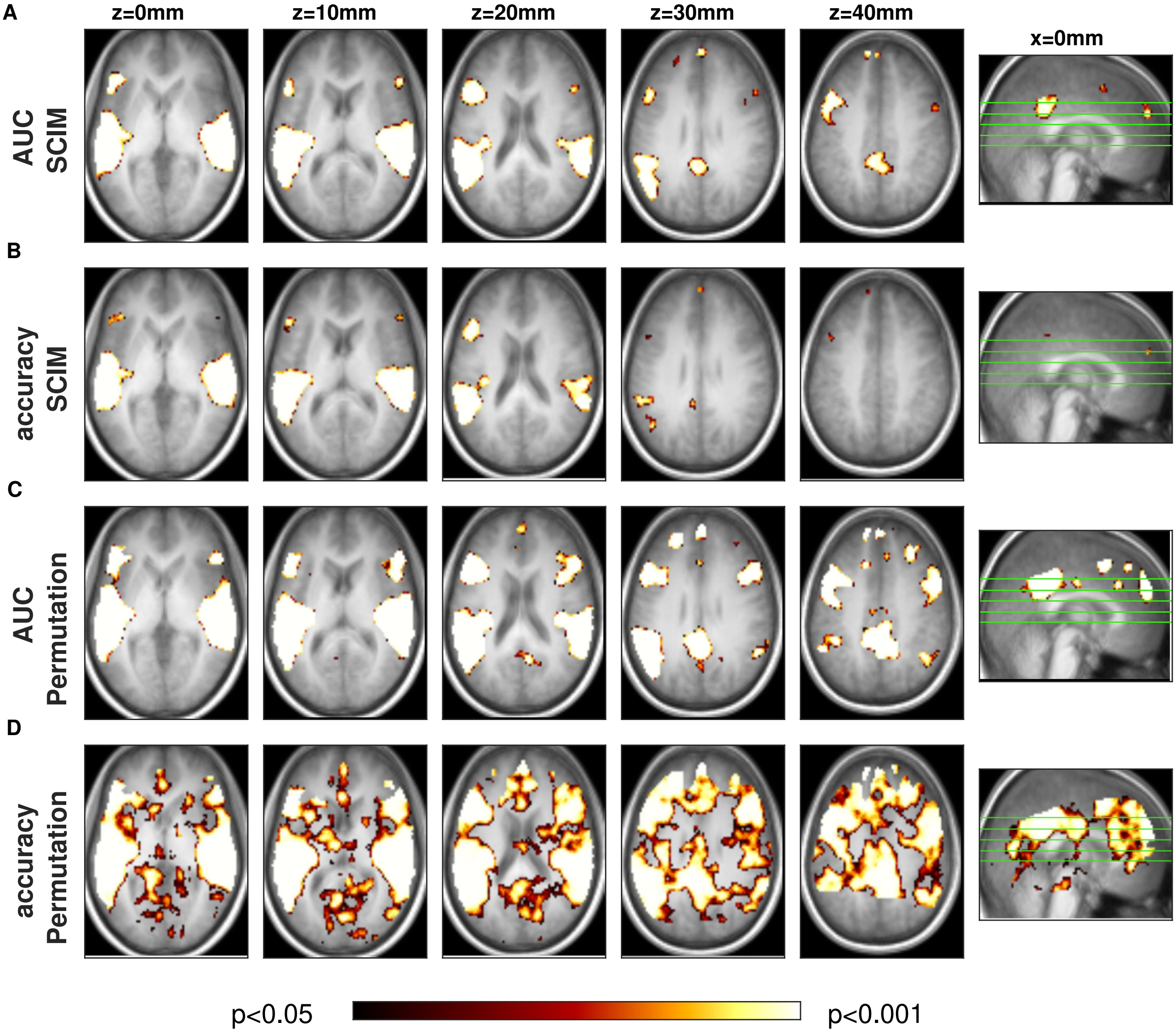

Figure 2.13.: Group result maps for the contrast semantic speech vs. non-semantic speech, with proposed SCIM method on (A) AUC maps, (B) accuracy maps, random permutation test on (C) AUC and (D) accuracy maps in five transversal slices and one saggital slice to display location of transversal slices. Informative regions for group results obtained with random permutation test and SCIM method on AUC maps are qualitatively consistent, however spatial extent of informative regions from random permutation test is slightly larger compared to those obtained from SCIM method analysis. While SCIM result maps based on accuracy measure show spatially smaller informative regions with less reliability, corresponding maps obtained from random permutation test seem to be too optimistic and lead to non-interpretable informative regions. For AUC measures informative regions are located in primary and secondary auditory cortex, namely in Heschl's gyrus (HG) and superior temporal gyrus (sts) as well as adjacent regions, Broca's area and Wernicke's area, that have been associated previously with speech processing. Additional informative regions can be found outside of temporal cortex, in anterior and posterior cingulate gyrus, previously being associated to semantic processing.

is highly dependent on the applied threshold and leads to non-plausible results for increased p-value thresholds (lower significance levels), the random permutation analysis test is more robust than the binomial test at the price of very high computational cost. IRMs obtained with the SCIM method are characterized by spatially smooth, low p_{SCIM}-values in informative regions that are clearly separated from non-informative, high p_{SCIM}-regions (cf. Fig. 2.7), and this partitioning is largely independent of the chosen threshold value. The same effect is visible for group results, as presented in Fig. 2.13 and Fig. 2.11, when comparing SCIM results with random permutation test. For AUC measure, a-posteriori probabilities obtained from group analysis with the SCIM method are comparable to those obtained from the random permutation test, though computationally of immensely higher efficiency. For the accuracy measure, the random permutation test is not only involved with high computational costs but leads also to non-interpretable results — in contrast to SCIM results, that are similar to those obtained with AUC measure, however with slightly smaller informative regions.

The dependence on applied thresholds was quantitatively illustrated in Fig. 2.9. The number of informative searchlights in IRMs is almost constant for a-posteriori probabilities smaller than 0.1. In binomial test result maps, the number of informative searchlights increases in the range of typically used thresholds between $p = 0.01$ and $p = 0.05$. The same effect can be observed for random permutation result maps based on the AUC measure without FDR correction. These results emphasize the need for false discovery correction in result maps arising from random permutation and binomial tests. Shifted thresholds have a small impact on resulting IRM informative areas. In most of the quantitative comparisons of the SCIM method with the reference methods in this paper a value range up to p=1 is depicted, even though values of p >0.05 have little relevance for the experiment data. This was done to illustrate the robustness of the SCIM method. Given the robustness of spatial patterns for large values of p >0.05, it is reasonable to expect highly robust results for lower choices of p-values (or, more specifically: for lower chosen a-posteriori probability levels). Therefore, the SCIM method might provide a tool for fMRI analysis that to some degree maintains sensitivity with increased specificity (Lieberman and Cunningham, 2009).

Across experiments and simulations presented here, our analyses failed to identify situations where use of SCIM would induce a considerable disadvantage compared to reference methods. Methodological differences, e.g., underlying assumptions of the methods and associated numerical effort, also did not negatively affect the range of situations where SCIM is applicable. Under the scenario of p-values larger than 0.01 and a simultaneous emphasis on high specificity, the permutation test might be preferred when its high computational cost is irrelevant. However, this scenario is of limited relevance for most studies. We expect that future studies will contribute to a broader understanding of the algorithm's qualities.

2.5.2. Smoothing of classification performance maps

Spatial smoothing influences the outcome of statistical tests and estimated posterior probabilities, since it alters the underlying AUC (and respectively accuracy) distributions in informative and non-informative brain regions in different ways. Regions of high spatial continuity, corresponding to comparably low standard deviation, are expected to coincide with informative regions of high mean AUC and accuracy, respectively. Smoothing further reduces their deviation even more and, thus, reduces the number of searchlights at the far right-hand tail of the distribution, as would be detected by a fixed threshold. Non-informative regions are characterized by lower spatial continuity, resulting in a comparatively larger reduction of the standard deviation being induced by spatial smoothing. The proposed SCIM method adaptively tracks changes in these underlying distributions since the two-component Gaussian mixture model adapts to the informative and non-informative distributions that are implied by the observed data, i.e., smoothed or unsmoothed AUC and accuracy. Spatial smoothing applied to the random permutation and binomial reference methods has predominantly the effect of reducing spatially "noisy" false positive searchlights (cf. Fig. 2.3 and Fig. 2.4, third and fourth rows), however, occasionally coinciding with a reduction in informative map extent. The higher sensitivity that can be achieved with all three methods is also reflected in the panels A and B of Fig. 2.6 and the ROC curves in panel F of the same figure. The SCIM method benefits from smoothing in particular through the inclusion of additional searchlight volumes into the informative region estimate, with an overall increase in physiological plausibility (cf. Fig. 2.3 and Fig. 2.4, second row).

2.5.3. Classification performance measure

Group result maps obtained from SCIM analysis in Fig. 2.13 (A and B) display larger informative regions resulting from the AUC measure than from the accuracy measure, with the former providing a better match across the spatial extent of physiologically task-relevant areas known from literature (cf. discussion of physiological results below). Robustness, evaluated as the dependency of the number of informative searchlights on applied thresholds, shown in Fig. 2.9, is comparable for both performance measures in the SCIM method. For random permutation tests, results based on accuracy are hard to interpret and inconsistent with informative regions known from literature for the task of semantic processing.

These observations likely reflect principled advantages of the AUC measure for classifier evaluation over the accuracy measure. AUC has been shown to provide a performance measure that is invariant to a priori class probabilities and exhibits increased sensitivity and decreased standard error (Green and Swets, 1966; Bradley, 1997; Spackman, 1989). In the context of the SCIM method, it provides us with a tool to perform reliable regularization and model selection in the SVM learning step and, thus, prevents over-fitting in the searchlight classification step. The informative and non-informative distributions, obtained in the GMM step of the SCIM method, are characterized by a small standard deviation due to effectively "decreased noise" in the AUC measure. This facilitates the decomposition into an informative and non-informative searchlight distribution in the SCIM method and increases the robustness of informative regions in IRMs.

2.5.4. Physiological results

The analyses presented above consistently found informative brain regions in auditory cortex areas that are associated with speech perception. Specifically, we identified superior temporal sulcus (STS) which was shown to play a role in speech processing in general (Markiewicz and Bohland, 2016; Osnes et al., 2011; Uppenkamp et al., 2006) and the processing of intelligible speech in particular (Abrams et al., 2012; Davis and Johnsrude, 2003; McGettigan et al., 2012). Heschl's gyrus, another brain region labeled as informative by the SCIM method, has previously been connected with different degrees of speech clarity (Wild et al., 2012), perception of vowels (Formisano et al., 2008), intelligible speech (McGettigan et al., 2012) and syllables (Markiewicz and Bohland, 2016). In auditory cortical areas associated with higher order auditory processing, regions in inferior frontal sulcus showed informative content for the semantic vs. non-semantic speech contrast. They had previously been reported to be relevant for semantic and phonological processing, word and syllable counting (Poldrack et al., 1999), as well as for hierarchical structures and sentence processing (Makuuchi et al., 2009), speech working memory (Friederici et al., 2006), and processing of intelligible speech (Abrams et al., 2012). In fronto-cortical areas, our group analyses showed reliable results in cingulate gyrus ($p_{\mathrm{SCIM}} < 0.001$), which is consistent with findings by Adank and Devlin (2010) for processing of auditory sentences, and for output-related vowel information by Markiewicz and Bohland (2016). In Rissmann et al. (2003), this region showed higher activation for words compared to non-words. Binder et al. (2009) described this area in their meta-analysis as interface between the semantic retrieval and episodic encoding systems.

2.6. Conclusion

This work explored a novel method to evaluate neurophysiological data that was tested on data obtained from an auditory fMRI study, investigating cognitive processes during the semantic processing of speech. The method is based on searchlight classification analysis with subsequent division of searchlight results into informative and non-informative searchlight regions and permits a more robust discrimination of informative vs. non-informative cortical regions than common evaluation methods like the random permutation tests or the binomial test. Informative regions obtained with the method are qualitatively consistent with those obtained from reference methods. Yet, a-posteriori probabilities resulting from the SCIM method dissociate into two distinctly separate distributions, whereas separation of significant from non-significant results in the reference methods are highly threshold-dependent. Since changes in applied thresholds change resulting informative region maps to a lesser degree, the SCIM method provides an evaluation tool that increases the specificity of multivariate fMRI analysis without degrading sensitivity in a considerable manner. The method is applicable to all fMRI studies that permit a classification of BOLD responses into distinct classes of tasks or conditions. It is beneficial in particular for fMRI studies in which sparse imaging is used and the data set is rather small. The example data presented in this study illustrate that the procedure allows for robust identification of plausible group effects that were not found with univariate statistical analysis.

3. Local BOLD Patterns - Disentanglement of the Neural Correlates of the Processing of Acoustic and Abstract Speech Features

3.1. Abstract

The understanding and interpretation of speech is both, indispensable for basic human interaction and a complex and abstract cognitive task for the human brain. Disentangling the neural substrates of processing different speech features requires a strategically designed stimulus set and a careful combination of individual results to group results.

In this chapter results from an auditory fMRI study encompassing stimuli that differ in the extent of semantic content, human voice pitch and amplitude modulation are presented. In addition to results from searchlight classification analysis, inter-contrast differences of fMRI data obtained from a correlation analysis approach are presented. Conclusions from the integration of quantitative single subject results differ partially from the investigation of group result maps and indicate a superior effect caused by modulation in known auditory regions like Heschl‘s gyrus, the superior and medial temporal gyrus and inferior frontal regions. Correlation effects for voice pitch content can be identified predominantly in belt and parabelt areas of the auditory cortex and in the inferior frontal gyrus, while regions in occipito-parietal areas and structures of the limbic system correlate more likely with semantic processing.

3.2. Introduction

A large part of human communication is relying on understanding spoken language, which is a complex cognitive operation including basic processes like pitch perception up to highly abstract tasks like semantic interpretation. To understand underlying neural substrates of this hierarchically structured process (Hickok and Poeppel, 2007), different stages of speech comprehension need to be investigated and compared across different stimuli. Furthermore, the investigation of rather complex cognitive processes requires careful methods to combine single subject results to group results. This is due to the increasing inter-individual variance of brain activity for abstract tasks that might lead to oversight of important effects in conservative group analysis approaches.

Human speech perception and comprehension shows an impressive robustness to everyday disturbance and variations. It is not possible to extract a single speech feature that predicts speech intelligibility (Arai and Greenberg, 1998) since different stages of speech perception are not fully separable. Their neural correlates show overlapping patterns and science has not been able to fully disentangle and understand these processes yet. However, much research has been conducted on the particular speech features and their representation in neural patterns during speech perception.

In acoustic speech, amplitude modulations are one important cue for information encoded in the signal. Giraud et al. (2000) showed that amplitude modulations contribute to neural activation from lower brainstem areas up to higher cortical stages, depending on the frequency and its correlate in speech perception. They found activation corresponding to higher frequencies in rather early stages of the auditory pathway, describing voicing and prosody. Lower modulation frequencies, correlating with articulation and syllabic rate, accompanied activation in areas around the primary and secondary auditory cortex. Even though amplitude modulations contribute to the semantic meaning of spoken utterances (Riecker et al., 2002), they are not sufficient for comprehension when they are imprinted as a speech envelope, e.g., on a simple noise carrier signal (Giraud, 2004). Another important cue for understanding the perception of human speech is the processing of formant structures. These have considerable impact on the perception as natural speech (Lattner et al., 2005; Heinrich et al., 2008). Results of a behavioral study by Johnsrude et al. (2013) indicate that the voice that is determined by the formant structure influences the intelligibility of spoken utterances, both when the speaker‘s voice or a disturbing speaker‘s voice are familiar to a listener. Areas that are voice sensitive have been identified in multiple studies, comparing speech vs. non-speech sounds (Belin et al., 2000; Uppenkamp et al., 2006), manipulation of speech-sounds (Jäncke et al., 2002; Meyer et al., 2004; Heinrich et al., 2008; Warren et al., 2006) or variation of speakers (Formisano et al., 2008; Lattner et al., 2005). These studies consistently reported activation in areas along the temporal lobe, overlapping with activation for other speech features, e.g., prosody as a percept resulting from variations in pitch, duration and loudness of spoken words (Gandour et al., 2002).

The extraction of meaning from spoken language can be examined on two different levels: (1) the lexical semantic stage that assigns specific meaning from perceived phoneme combinations to items in a lexical storage and (2) the conceptual semantic stage, where potential ambiguous meaning is clarified in context, typically on sentence level. Literature provides insights from studies investigating both stages of semantic processing. Lexical semantic processing has been subject of research in studies investigating neural activity during word- and pseudoword-processing (Binder et al., 2000a; Poldrack et al., 1999), word generation (Hwang et al., 2009; Birn et al., 2010; Jeon et al., 2009), word categorization (Akama et al., 2012; Birn et al., 2010; Handjaras et al., 2016) and, more particular, in contrast from ambiguous and unambiguous word-pairs (Bilenko et al., 2009).

On sentence level, the interpretability of speech can be modulated via syntactic and semantic violations. Humphries et al. (2006) examined the question, whether the cognitive load for processing sentences compared to randomly ordered words is due to the absence of syntactic structure or the absence of relations between content words. Contrasting effects of violations in syntactic vs. semantic structures of sentences has been also object in various other studies (Friederici et al., 2003; Rogalsky and Hickok, 2009; Tyler and Marslen-Wilson, 2007). Another aspect of semantic processing is the beneficial effect of semantic priming on speech intelligibility, which has been investigated by Obleser et al. (2007b), Rothermich and Kotz (2013), Schmidt and Seger (2009), Blank and Davis (2016), Rodd et al. (2012), and Davis et al. (2011); Devauchelle et al. (2008).

Cortical structures engaged in the extraction of semantic meaning from acoustic signals have been identified in areas associated with lower-level speech perception in the temporal lobe (Davis et al., 2011; Abrams et al., 2012; Devauchelle et al., 2008; Friederici et al., 2003; Humphries et al., 2006). However, more abstract tasks in speech processing correlate with activation more distant from the auditory cortex (Davis and Johnsrude, 2003), like the inferior frontal gyrus (Abrams et al., 2012; Devauchelle et al., 2008) or the angular gyrus (Abrams et al., 2012; Golestani et al., 2009; Humphries et al., 2006).

The aim of this study is the identification of cortical structures that are engaged in the extraction of semantic meaning from acoustically perceived sounds. For this purpose we presented subjects during a sparse imaging fMRI experiment with six different acoustic stimuli. The stimuli varied in their comparability to natural speech in semantic content and acoustic properties, in particular voice pitch and temporal modulation. The speech-like stimuli are of primary interest in this study, since their construction permits a comparison of cortical activation during the perception of normal spoken sentences and stimuli that are constructed from human speech recordings but do not hold semantic content without distortion of the physical stimulus properties.

Accounting for results from the previously mentioned studies, we derive the following two hypotheses:

H1: Cortical structures that are involved in the semantic processing of spoken language are located in a widely distributed network outside of the core auditory areas.

H2: The components of spoken language are processed hierarchically with an initial detection of temporal changes in early primary auditory areas, followed by the detection of a human voice percept in belt and parabelt areas and the analysis and interpretation of perceived stimuli in higher level cortical regions.

Results obtained from the analysis across subjects require a careful interpretation, especially for abstract cognitive tasks like the extraction of semantic meaning, since, e.g., individual strategies for specific tasks can cause differences in the localization of brain activity for these tasks (Seghier and Price, 2016). Seghier et al. (2004) showed high inter-subject differences in the localization of activation in the same anatomical area (prefrontal cortex) for semantic and phonological tasks in particular. High inter-subject variability in the prefrontal cortex has been also described in the literature (Poldrack, 2017; Laumann et al., 2015). Even after smoothing fMRI data in the pre-processing steps, the goal to normalize potential activation across subjects remains unfulfilled (Mikl et al., 2008). Local shifts of particular brain regions can be in the order of 1 cm (Thirion et al., 2007), which makes the comparison of effects in comparably small anatomical regions very challenging. Hence, group results suffer from an inter-subject variability that is in the same magnitude as the intra-subject variations (Wei et al., 2004).

The approaches presented in this study aim to compensate for this problem by analyzing the quantitative results in the particular anatomical brain regions on single subject basis and examine the resulting statistics across subjects. The BOLD activation is analyzed in the multivariate whole brain searchlight algorithm (Kriegeskorte et al., 2006) in combination with a support vector machine (SVM) classifier. The quantitative analysis of single subject result maps permits the observation of effects across contrasts that are not extractable from standard group result maps. They indicate the presence of patterns in the results that allow for the disentanglement of neural activity corresponding to the processing of different speech features.

Another challenge for the design of neuro-cognitive studies is the separation of tasks or stimulus features, while maintaining a high naturalness and comparability to real life tasks. The stimuli utilized in this study also differ in multiple dimensions that can be summarized essentially to the content of semantic meaning, voice pitch, and temporal modulation. To disentangle the neural correlates of speech-processing regarding these features, we propose to analyze data not only across subjects, but also across contrasts. A Spearman rank-correlation between classification analysis results and the ranking positions is calculated on a voxel basis. This provides whole brain maps, displaying the spatial distribution of implicitly correlated neural activity to the investigated task.

Correlated region maps (CRM) obtained from this analysis indicate a spatial separation of cortical structures that are engaged in the processing of acoustic features compared to the extraction from semantic meaning and a broadly distributed network for the extraction of semantic meaning from speech, in particular outside of the auditory temporal lobe.

3.3. Methods

3.3.1. Stimuli

The aim of this study is the investigation of different stages of speech understanding with cortical representations for the extraction of semantic meaning from spoken language in particular. The study design comprises six different acoustic conditions plus one control condition without any stimulus presentation via headphones to separate neural correlates of different speech processing stages.

To compare cortical activity during the processing of speech with semantic content in contrast to the processing of speech without semantic content, we utilized, amongst others, two speech stimuli: semantic speech and non-semantic speech. Non-semantic speech utterances were fragments extracted from the International Speech Test Signal (ISTS, Holube et al., 2010), a test signal designed for language

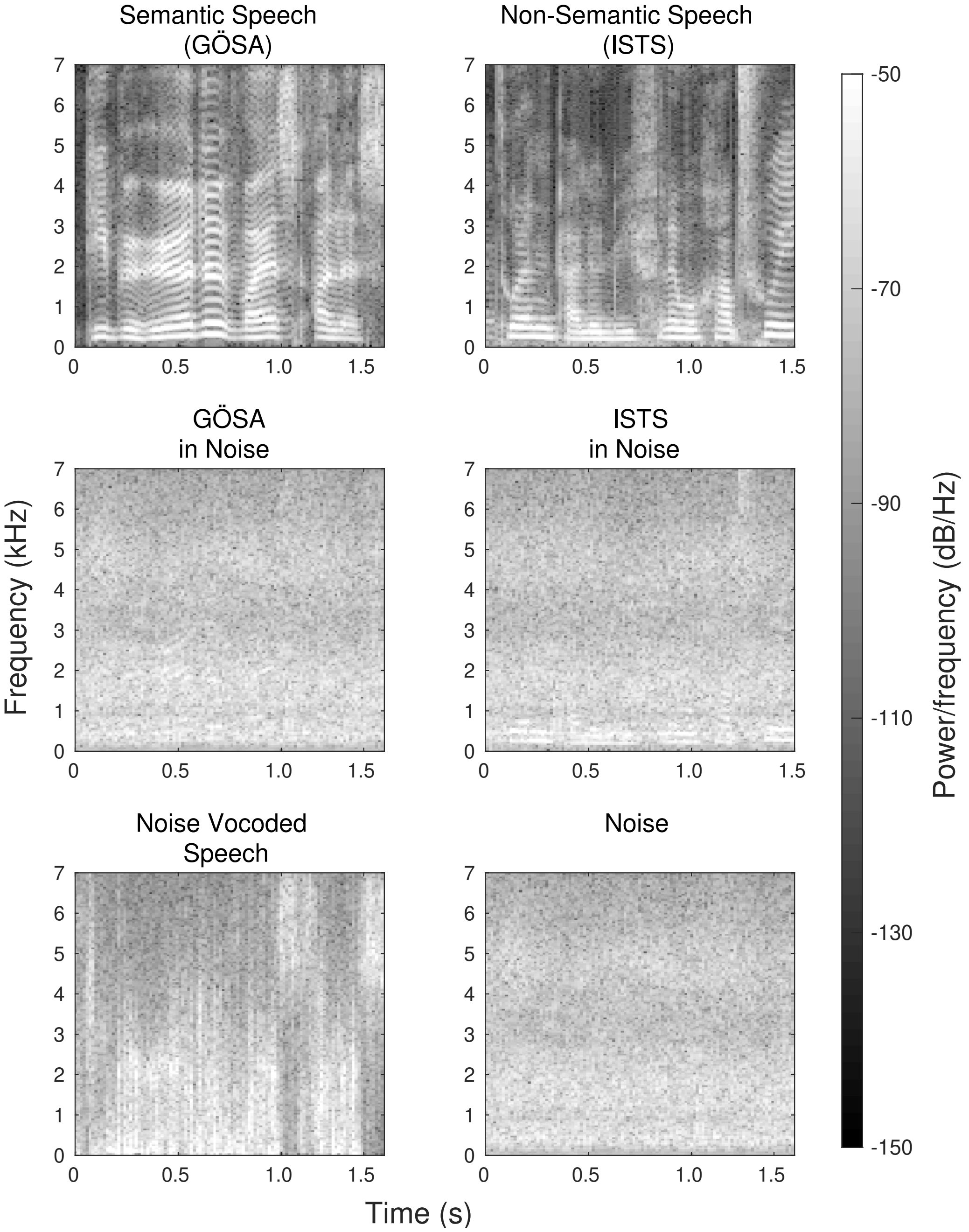

Figure 3.1.: Spectrograms of example stimuli from presented stimuli. Upper panels: Speech stimuli, semantic speech taken from the Göttingen Sentence Test (GÖSA) and non-semantic speech taken from International Speech Test Signal (ISTS) material. Both conditions show typical formant structure and are physically comparable. Middle panels: Semantic and non-semantic speech overlaid with speech-simulating noise from GÖSA material, respectively. In low frequency ranges formants are still observable. Lower panels: Noise vocoded speech (left panel) obtained from filtering white noise with LPC coefficients of GÖSA sentences and speech simulating noise (right panel) taken from GÖSA material.

independent hearing aid evaluation. It is comprised of speech recordings from six female speakers, whose particular native language is Arabic, English, French, German, Mandarin or Spanish. Sequences of 100 ms - 600 ms duration were rearranged in pseudo-random order under specific conditions that allow for the impression of a continuous natural speech stream. The resulting signal is controlled for physical properties of speech but does not contain any valid statements.

Each semantic speech condition sample contained two randomly chosen sentences from the Göttingen Sentence Test (GÖSA, Kollmeier and Wesselkamp, 1997) with an accumulated length of approximately 6 s, respectively. The GÖSA is a phonetically balanced speech intelligibility test in German language, containing short semantically meaningful statements. The perceptual comparability to the non-semantic condition was increased by a digital pitch shift towards higher frequencies and a digital vocal tract change, both under usage of *Tandem Straight* (Kawahara and Morise, 2011). This transformation of the original male voice allowed for the percept of a female voiced semantic speech stimulus.

Speech-simulating noise, generated from the GÖSA material, provided a noise signal without perceptable amplitude modulation but with long-term spectral similarity to speech stimuli. This stimulus was utilized as contrast stimulus for the speech stimuli and also as masker for the two speech stimuli, semantic speech (GÖSA) and non-semantic speech (ISTS) at a signal to noise ratio (SNR) of $-6\ dB$. To create noise vocoded speech (NVS) stimuli, white noise was filtered with the first twelve linear predictive coding (LPC) coefficients (Madisetti and Williams, 1998) of the original GÖSA material. All presented stimuli had a duration of approximately 6 s and have matching long-term, speech-like spectra. The stimuli utilized in this study mainly differ in the dimensions semantic content, voiced pitch and modulation.

3.3.2. Data acquisition

FMRI data were recorded with a 3 T Siemens MRI scanner. The group of participants included 19 volunteers (11 male, 8 female, 23,5 ± 2,6 years in age), 18 of whom were native speakers of German and were included in further analysis. In a passive listening paradigm, the subjects were presented with the previously described stimuli in a random order, that was fixed across participants. In the sparse imaging design with a trial duration of 9 s, a stimilus presentation of about 6 s was followed by an EPI data aquisition sequence of 2.9 s duration for a complete brain volume in 21 slices with a voxel size of $3.125\ \mathrm{mm} \times 3.125\ \mathrm{mm} \times 3.9\ \mathrm{mm}$, a field of view of 20 cm × 20 cm, a matrix size of 64×64 and an echo time (TE) of 55 ms. One complete session comprised 50 min, including 4 runs with 70 trials each (10 trials per condition), respectively, and the acquisition of a T1-weighted anatomical image for the exact localization of resulting active brain regions.

3.3.3. Data analysis

Raw fMRI data were preprocessed with the software SPM8 (Friston et al., 1995), including a spatial realignment to reduce artifacts caused by head movements during the data acquisition. Subsequently, all structural and functional images were normalized to MNI space to allow for inter-subject comparisons.

Preprocessed data were analyzed within the searchlight algorithm (Kriegeskorte et al., 2006). The BOLD response maps from each subject were spatially divided into overlapping sub-datasets. A spherically shaped "searchlight" with a three voxel radius surrounds each voxel, the corresponding center-voxel for the searchlight, containing 123 voxels in total. Sub-datasets are composed of 80 samples (two classes with 40 samples each) and 123 features (voxels in the "searchlight sphere") and are analyzed separately in a nested cross-validation procedure. Classification analysis was performed with a support vector machine (SVM) and classification performance results, expressed as the area under the ROC curve (AUC), are mapped to the center voxel, respectively. Choosing the AUC as classification performance measure has proven to result in more reliable informative region maps as compared to those results obtained with an accuracy measure (cf. Chapter 2). The described classification procedure

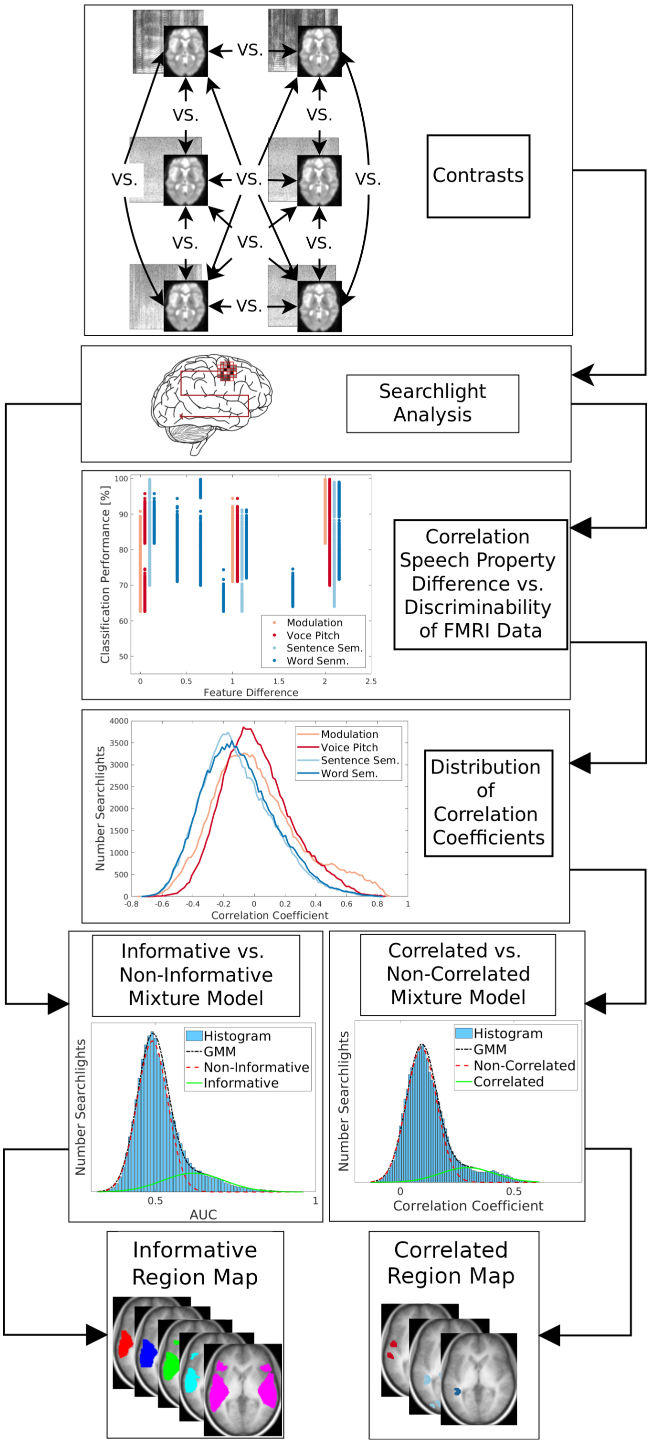

Figure 3.2.: Analysis algorithm procedure. FMRI data are classified in the searchlight algorithm with a linear kernel support vector machine. Resulting performance values are spatially smoothed and either decomposed into sub-distributions with a Gaussian mixture model (GMM), resulting in informative regions maps (IRM), or correlated with stimulus differences in semantic content, voice pitch, and modulation. Searchlights are, in turn, decomposed into correlated and non-correlated searchlights based on the distribution of the correlation coefficient from every searchlight with differences in the stimulus features, respectively, resulting in correlated region maps (CRM).

results in an “information-based map” (Kriegeskorte et al., 2006) that represents the informational content of a voxel and its surrounding neighborhoood. The spatial smoothing of information based maps has shown to be advantageous for the further processing. It accounts for inter-subject variations of the exact location of cortical structures and reduces the variance of classification performance results in regions without informative BOLD activation patterns (cf. Chapter 2). Therefore, AUC maps are spatially smoothed with a Gaussian kernel (full width half maximum = $3\ mm$). According to the SCIM algorithm, which has been presented in the previous chapter, searchlight regions, whose BOLD response patterns allowed for a reliable classification into one of two contrasted conditions, are separated from those searchlight regions, whose BOLD response patterns did not contain information regarding the contrasted conditions. A histogram of all smoothed searchlight AUCs is processed in an Expectation Maximization (EM) algorithm (Dempster et al., 1977) to determine a two-component Gaussian mixture model (GMM) that decomposes all searchlight results into an informative searchlight distribution and a non-informative searchlight distribution. This approach is of very low computational effort as compared to the random permutation test and has shown to be considerably more robust against the choice of significance levels (cf. Chapter 2). From the converged GMM, a-posteriori probabilities for the informative and non-informative searchlight distribution can be derived, the latter being similarly handled like a null-hypothesis’ probability. Thresholds for a-posteriori probabilities are set to values between $p_a = 0.01$ and $p_a = 0.05$. Voxels containing searchlight AUCs corresponding to below-threshold p_a-values appear in binary result masks.

3.3.4. Contrasts

In this study, participants were presented with six different acoustic stimuli, namely semantic speech (Sem), non-semantic speech (NSem), semantic speech in noise (Sem_N), non-semantic speech in noise (NSem_N), noise vocoded speech (*NVS*) and Noise (see section *Stimuli*). From the BOLD responses corresponding to these six acoustic conditions, thirteen contrasts between two stimuli in each case were selected for the analysis.

Based on the differences of the stimulus features (semantic content, voice pitch and modulation) between the two respective conditions, the contrasts were sorted with respect to each of these stimulus features. Since the correlation analysis is based on a Spearman rank-correlation (Spearman, 1904), only the order and not the absolute values are of interest, to avoid making assumptions about the distribution of feature peculiarities. Equidistant differences between contrasts do not reflect equidistant effects.

The assumptions that influenced the ranking are the following for modulation: (1) Clear speech conditions (Sem_N and NSem_N) are fully modulated signals with the highest ranking in modulation, followed in the ranking by NVS that provides a percept of a constant noise floor. (2) Contrasts between the noisy speech condition presented with an SNR of $-6\ dB$ and noise are ranked higher than clear speech conditions vs. noisy speech conditions, since the latter hold perceptable modulation effects in both conditions, yet, of different value. (3) Sem and NSem are both fully modulated, however, acoustic differences between Sem_N and NSem_N are even less perceptable.

Pitch: (1) Clear speech conditions (Sem and NSem) contain the highest portion of voiced pitch. However, since the semantic speech stimuli have been manipulated to construct a female voice percept, the NSem_N sounds slightly more natural than Sem_N and is provided with a higher voice pitch ranking in the contrasts. (2) Noisy speech conditions contrasted with noise get a higher ranking than noisy speech conditions contrasted with their clear versions, since the latter hold perceptable voice pitch components on both sides.

Semantic: (1) Non-Semantic speech contains isolated monosyllable words and, therefore, holds a minimal lexical semantic component. (2) Single words of the NVS could be understood when the clear version of these words was presented immediately before. Therefore, NVS holds a small semantic component.

The ranking of the contrasts is presented in Tab. 3.1. High values correspond to a high ranking and large differences between the corresponding conditions regarding the particular stimulus feature.

	Modulation	**Pitch**	**Semantic**
Sem vs. NSem	2	3	8
Sem_Nvs. NSem_N	1	2	6
Sem vs. Noise	8	7	9
Sem vs. Sem_N	5	4	6
Sem_N vs. Noise	6	5	7
NSem vs. Noise	8	8	5
NSem vs. NSem_N	5	4	4
NSem_N vs. Noise	6	6	4
NVS vs. Noise	7	1	1
NVS vs. Sem	3	7	8
NVS vs. Sem_N	4	5	6
NVS vs. NSem	3	8	3
NVS vs. NSem_N	4	6	2

Table 3.1.: The table displays the ranking of contrasts regarding the three different stimulus features modulation,voice pitch and semantic. The motivation for the ranking and underlying assumptions are discussed in the text. High values correspond to high differences between the conditions. Since the values are used in a Spearman rank-correlation analysis, only the order but not absolute values are of interest and equidistant rankings do not reflect equidistant feature differences.

3.3.5. Correlation analysis

As presented in Tab. 3.1, contrasts between conditions were ranked in their differences regarding their content of semantic meaning, voice pitch and temporal amplitude modulations to separate potential BOLD effects caused by these perceived properties in subsequent analyses. We hypothesize a correlation effect between the extent of BOLD differences in fMRI data, expressed as pattern differences and, accordingly, the area under the ROC curve (AUC) values obtained from searchlight classification analysis, and the differences between contrasted stimulus condition properties. Therefore, a correlation coefficient ρ_m according to Eq. 3.1 is calculated for each voxel m:

$$\rho_m = \frac{\mathrm{cov}(\mathrm{rg}_{\mathrm{AUC}},\mathrm{rg}_f)}{\sigma_{\mathrm{rg}_{\mathrm{AUC}}} \cdot \sigma_{\mathrm{rg}_f}} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{rg}(\mathrm{AUC}_{im}) - \overline{\mathrm{rg}(\mathrm{AUC}_m)}) \cdot (\mathrm{rg}(\mathrm{f}_i) - \overline{\mathrm{rg}(\mathrm{f})})}{\sigma_{\mathrm{rg}_{\mathrm{AUC}}} \cdot \sigma_{\mathrm{rg}_{\mathrm{f}}}} \tag{3.1}$$

with N the number of contrasts (13 in this study), AUC_{im} the searchlight AUC value in the i-th contrasts and the m-th voxel, f_i the feature difference in the i-th contrast and rg the respective rank across values obtained from all contrast.

These calculations result in one correlation map per stimulus-feature, containing a correlation coefficient for every voxel. The entirety of correlation coefficients is separable into correlated and non-correlated voxels, comparable to the separation into informative and non-informative searchlights from the SCIM method (cf. Chapter 2). However, a large portion of searchlight results showed a negative correlation with the feature difference and therefore also a three-component GMM was fitted to the results, including a third anti-correlated sub-distribution of voxels/searchlights. Results obtained from both approaches have been compared across subjects and for group results. A-posteriori probabilities for the assignment to the non-correlated (and anti-correlated/correlated) voxel distribution preserve a base to separate correlated regions and anti-correlated regions from non-correlated regions, and to create correlated region maps (CRM) for all three investigated stimulus features, respectively.

Regions obtained from these maps do not inevitably derive from the conjunction of informative regions regarding the investigated contrasts: While some searchlight classification results might be not robust enough for the detection as informative in single contrast analyses, the correlation of AUCs with property differences across contrasts might accentuate these subtle effects. Stable informative regions, however, might occur for multiple contrasts with diverse foci and show no strong correlation with specific features.

3.3.6. Group analysis

The statistical analyses of BOLD responses in this study result in informative region maps (IRM), displaying searchlight regions with distinguishable BOLD patterns for specific conditions and correlated region maps (CRM), displaying regions with high correlation between classification performance results and stimulus-feature differences. Group results obtained from theses analyses are presented in two manners within this chapter: (1) group result maps obtained from voxelwise averaging of classification performance results (AUCs) or correlation coefficients and subsequent evaluation of these average AUC/correlation maps, and (2) statistical evaluation of single subject effects in anatomical regions across subjects.

(1) Subsequent to the searchlight classification analysis of single subject data, the average classification performance map for every contrast is created for each voxel across subjects. Within the SCIM algorithm (cf. section *Data Analysis*), group informative searchlights are separated from non-informative searchlights based on every searchlight's a-posteriori probability for non-informative searchlight distribution in the converged GMM. For CRMs, the correlation of the voxelwise average correlation coefficients across subjects is calculated and correlating searchlights are separated from non-correlating and anti-correlating searchlights with a three-component GMM.

(2) Single subject result maps have been overlaid with the automated anatomical labeling template (AAL, Tzourio-Mazoyer et al., 2002) to determine the borders of the anatomical areas. For every anatomical area, the ratio of active (informative, correlating, anti-correlating) voxels to all voxels within the area is computed, giving the proportion of the area's involvement in the underlying cognitive process. In combined presentations, medians, quantiles and outliers of single subject results can be compared for different contrasts, properties and hemispheres. This permits a more differentiated evaluation of group results compared to conventional group results described in (1). This approach considers the large spatial inter-subject variance and allows for identification of weak but consistent effects.

3.4. Results

3.4.1. Informative region maps - IRMs

In the following, the expression *speech conditions* describes the entirety of the conditions that contain natural physical speech properties like human voice pitch and modulation. These include semantic speech (Sem), non-semantic speech ($NSem$), semantic speech in noise (Sem_N) and non-semantic speech in noise ($NSem_N$).

Group result maps obtained from the classification analysis of the 13 selected contrasts can be found in Fig. 3.3. All contrasts show similar patterns of informative regions for the different contrasts. The highest consistency across contrasts can be found in the superior and the middle temporal cortex. The contrast Sem vs. $NSem$ shows the largest left lateralized dilatation towards posterior parts in the temporal cortex extending into the angular gyrus (AnG). Inferior frontal regions, like the opercular part of the inferior frontal gyrus (OpIFG), the triangular part of the IFG (TrIFG) and the orbital part of the IFG (OrIFG) also contain information about several of the considered contrasts, as well as the precentral gyrus (PrG) and the postcentral (PoG) gyrus and surrounding regions. Systematic differences between contrasts are described in the following subsections.

Semantic vs. non-semantic speech

For both contrasts, Sem vs. $NSem$ and Sem_N vs. $NSem_N$, the medial part of the temporal cortex and the insular cortex (INS) are detected to be informative regions. For the contrast between clear conditions, additional informative regions in the anterior and the posterior parts of the temporal cortex can be observed. Informative regions could also be found in the PrG, the supramarginal gyrus (SMG), the AnG and the middle frontal gyrus (MFG).

NVS vs. speech conditions

The superior to middle temporal cortex regions consistently hold informative regions for contrasts between NVS and speech conditions with stronger dilatation toward anterior parts of the temporal cortex, in particular in the left hemisphere. For contrasts containing NVS and speech conditions, consistent informative regions can also be found for the IFG as well as the inferior part of the SMG and the INS. Informative regions in the PrG occur for all contrasts between NVS and speech conditions with decreased spatial extent for the contrast NVS vs. $NSem$. This contrast instead shows more prominent informative regions around the central sulcus. NVS contrasted with noisy speech conditions leads to additional informative regions in the PoG compared to NVS vs. clear speech conditions. The contrast NVS vs. $Noise$ shows comparable patterns to the NVS vs. speech contrasts. Informative regions outside of the temporal cortex are less prominent or absent in the right hemisphere for all contrasts between NVS and the other acoustic stimuli.

Noise vs. speech conditions

The contrasts between speech conditions and $Noise$ lead to consistent informative regions in the temporal cortex from anterior parts (extending into temporal pole regions) to posterior regions, in particular for contrasts between clear speech conditions and noise in the left hemisphere, including the INS. Informative regions in the OpIFG, the TrIFG and the PrG are of smaller spatial extent for contrasts between noisy speech conditions vs. $Noise$ compared to clear speech conditions vs. $Noise$. Informative regions in the OrIFG and the inferior part of the SMG are only present for clear speech conditions vs. $Noise$. In the PoG, informative regions are smaller for Sem vs. $Noise$ compared to the other speech conditions vs. $Noise$. A left lateralization can be observed across all contrasts between speech conditions and noise.

Speech clarity

Differences between BOLD responses for speech conditions with different stages of speech clarity can be found in the contrasts Sem vs. Sem_N and $NSem$ vs. $NSem_N$. Both contrasts show informative regions in the superior and middle temporal cortex of comparable spatial extent across both contrasts, but more prominent in the left hemisphere. Inferior frontal regions as well as the PrG and the PoG show larger informative regions for the non-semantic speech contrasts compared to the semantic speech contrasts.

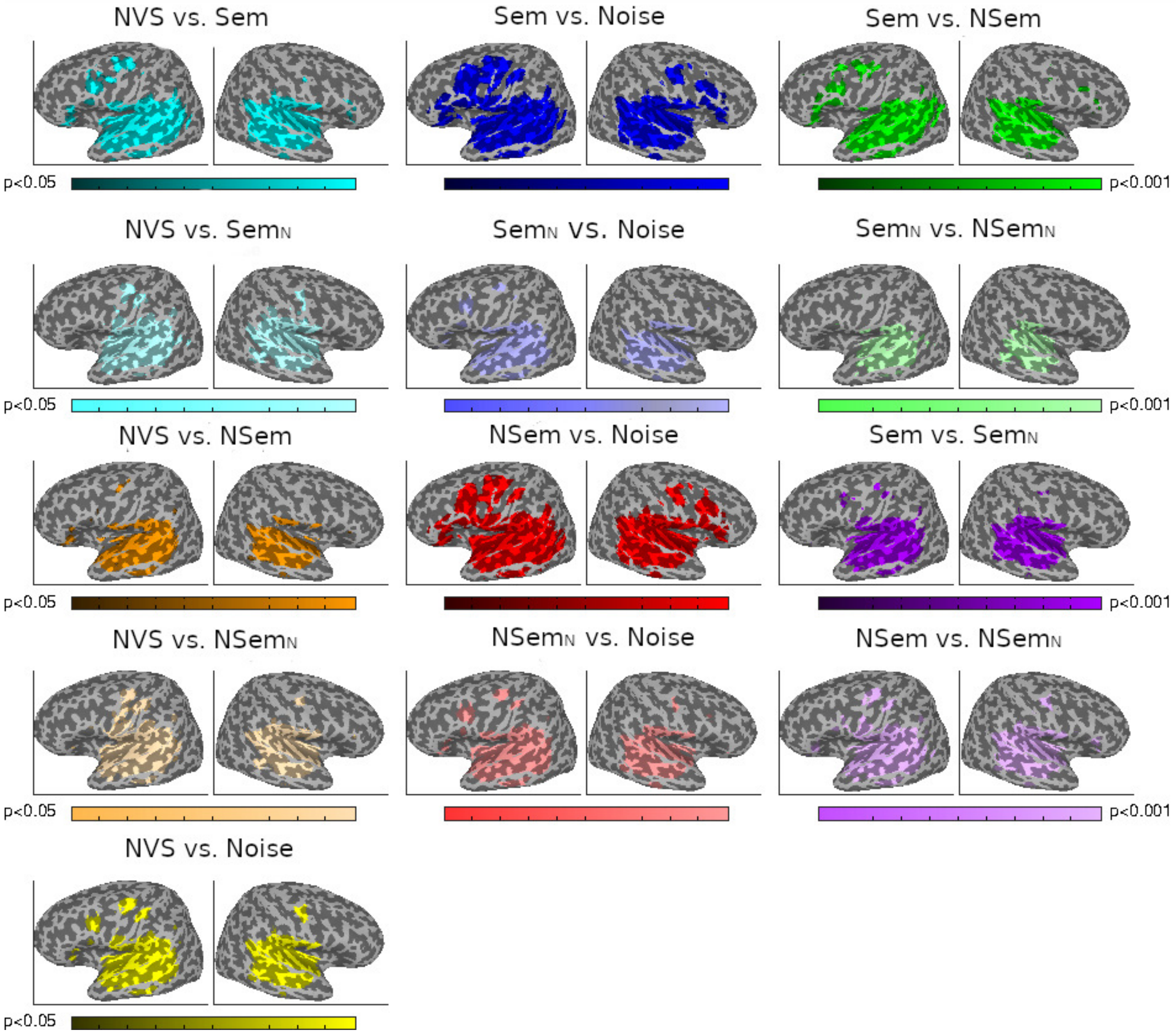

Figure 3.3.: Group results for 13 contrasts chosen from the combination of the six acoustic conditions semantic speech, semantic speech in noise, non-semantic speech, non-semantic speech in noise, noise vocoded speech (NVS) and noise. For all contrasts, informative regions can be found in the superior temporal gyrus (STG) and the middle temporal gyrus (MTG). In the frontal cortex informative regions can be identified for all contrasts except for Sem vs. Sem_N and Sem_N vs. $NSem_N$. Activation around the central sulcus is present for all contrasts except for Sem_N vs. $NSem_N$, with largest spatial extent in Sem vs. $Noise$, $NSem$ vs. $Noise$ and NVS vs. $Noise$. The results are presented at an FDR-corrected significance threshold of $p<0.05$.

Statistics across single subject results

Fig. 3.4 shows the statistical analysis of single subject results in comparison to group analysis results. Statistical analyses are shown for the contrasts between clear speech conditions, NVS and $Noise$ (panel A, blue colors), as well as noisy speech conditions, NVS and $Noise$ (panel B, pink colors). The location of the anatomical areas listed in panels A and B can be found in overlay images of one subject's anatomical image and the respective parts of the AAL template in two transversal slices (panel C).

The comparison of the BOLD responses for clear speech conditions and $Noise$ results in large informative regions in the superior temporal gyrus (STG) and Heschl's gyrus (HG), while contrasts of clear speech vs. NVS results in less informative searchlights in these areas with larger variance across participants. In the middle temporal gyrus (MTG), the superior temporal pole (StmP) and the posterior part of the INS, on average half of the searchlights are informative for contrasts $Noise$ vs. speech conditions, while for NVS vs. speech conditions the portion of informative searchlights in these areas ranges from 0% to 50%.

The corresponding contrasts containing noisy speech conditions (panel B) show comparable trends in the MTG, the StmP, the STG and the lateral HG, however, on average with a decreased size of informative regions compared to the corresponding contrasts with clear speech conditions. For the medial HG and the INS, the trend for more informative regions resulting from contrasts between $Noise$ and speech conditions compared to NVS with speech conditions is inverted. In these regions, $Noise$ vs. speech conditions contrasts result in less informative searchlights with higher variance across subjects, while NVS vs. speech condition contrasts show more informative searchlights with less variance across subjects.

3.4.2. Correlation analysis results

The correlation analysis leads to correlated region maps (CRMs) for the three investigated stimulus features semantic (SEM), voice pitch (VP) and modulation (MO). These display for every voxel/searchlight position the correlation coefficient obtained from the Spearman rank-correlation between the classification performance (AUC) and the corresponding feature difference between the conditions.

Statistical distribution of correlation coefficients

The distributions of correlation coefficients across all single subject result brain voxels and the voxelwise average across subjects are presented in Fig. 3.5. The CRMs for the MO condition on single subject level contain many voxels with negative correlation coefficients. However, voxelwise averaging results in less negative values on the left-sided tail of the data distribution, while the entirety of the single subject result distributions and the group result distribution are comparable on the right-hand side. This indicates that high correlation coefficients are spatially consistently located across group subjects. For VP results similar effects are observable, however, less distinct. The distribution of group average correlation coefficients for SEM shows a small variance with a small, not very pronounced deviation from a normal distribution towards negative correlation coefficients. Compared to MO and VP, the group SEM distribution is the only one that permits the identification of an anti-correlated searchlight distribution that is separable from the other distributions within a three-component Gaussian mixture model (GMM).

Yet, the comparison of results obtained from a two-component GMM (2C-GMM) and a three-component GMM (3C-GMM) fitted to the result data, depicted in Fig. 3.6, indicates for all features a better fit of the 3C-GMM. In the histograms in panel (A), a higher overlap of the original data (bar plot) and the fitted mixture model (solid black line) is observable for the 3C-GMM. The accumulated

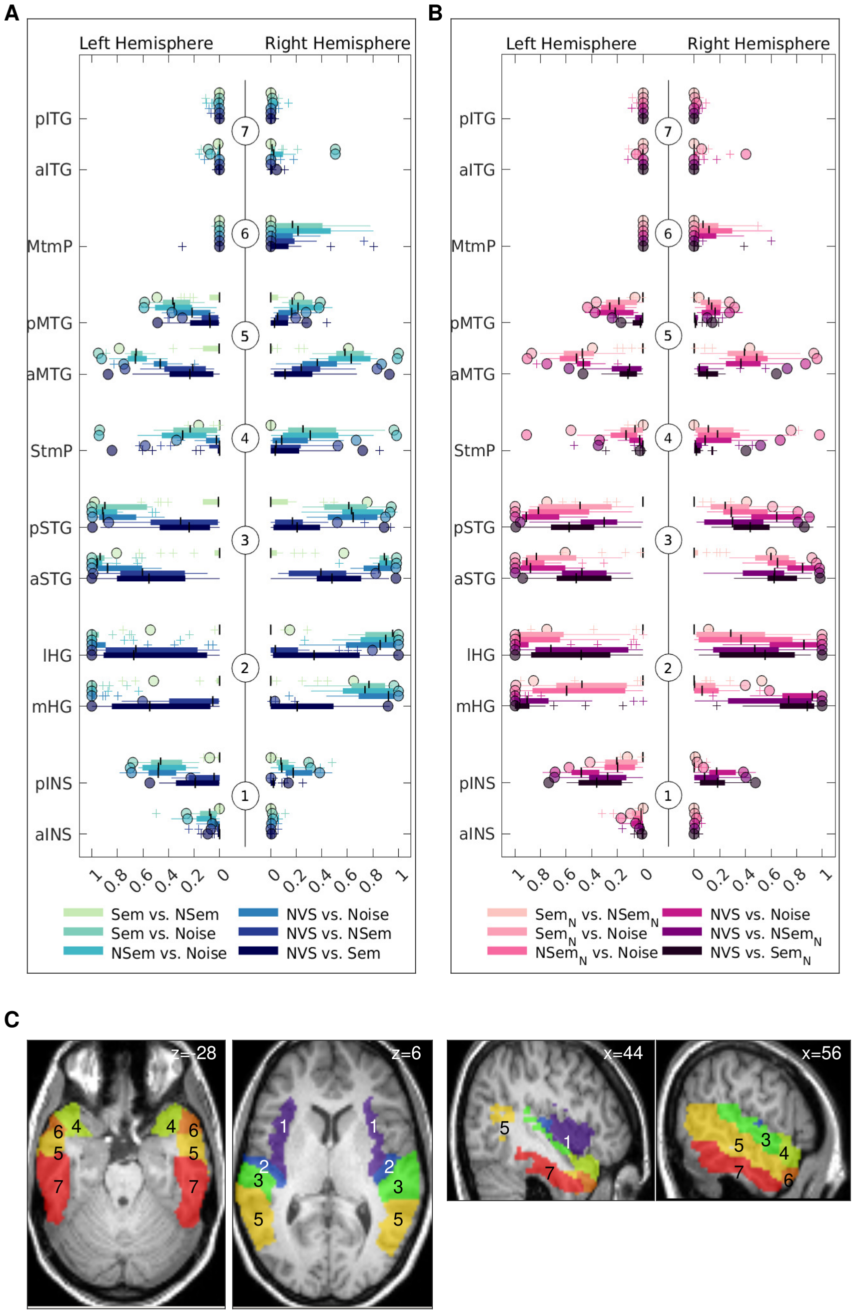

Figure 3.4.: Quantitative comparison of single subject results and group results for different contrasts. Panel A displays the quantitative statistics across all subjects for single subject IRMs for all contrasts that have been obtained from the combination of two conditions, respectively, from semantic speech (Sem), non-semantic ($NSem$) speech, noise vocoded speech (NVS) and noise at a significance threshold of $p<0.05$. Black lines display the median portion of informative searchlights from all searchlights within an anatomical region determined by the automated anatomical labeling (AAL) template. Inter-quartile ranges are displayed as colored bars for the corresponding contrasts, 5% and 95% percentiles as colored lines, and outliers are represented by colored crosses. Respective portions of informative searchlights in group result maps are displayed as filled circles. In panel B, the corresponding statistics are shown for combinations of conditions semantic speech in noise (Sem_N), non-semantic speech in noise ($NSem_N$), NVS and $Noise$. The location of the specified anatomical areas in panels A and B are displayed in two transversal slices in panel C.

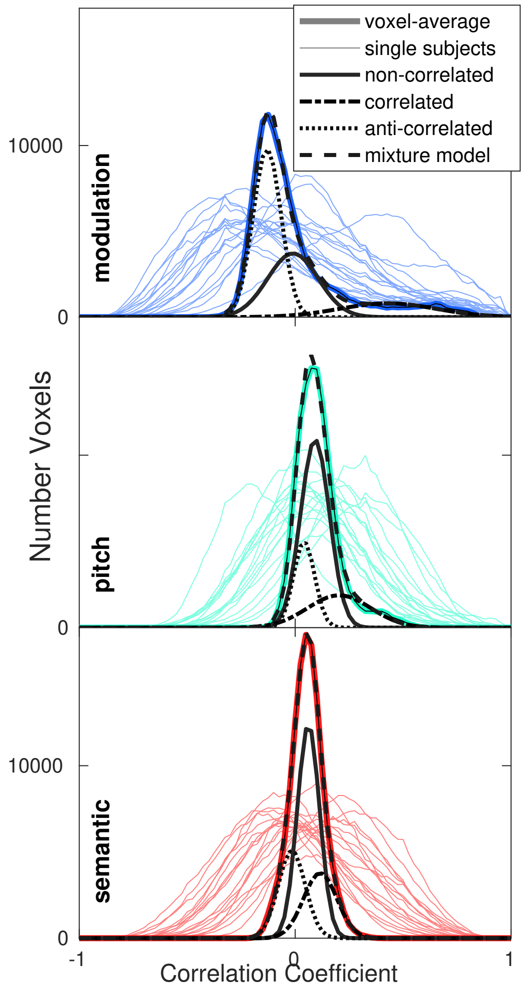

Figure 3.5.: Statistical distribution of correlation coefficients obtained from the correlation analysis across all brain voxels for single subject results (thin colored lines) and group results (thick colored lines). The modeled distributions obtained from a GMM decomposition are depicted in thick dashed lines for the mixture model, solid lines for the non-correlated sub-distribution, dashdot-lines for the correlated sub-distribution and dotted lines for the anti-correlated sub-distribution.

difference between the original data and the model is expressed as error in panel (B) for all single subject results. It is significantly smaller for the 3C-GMM in the MO condition across the group of subjects. For VP and SEM, the differences are not significant. For group SEM results (open circles), the 3C-GMM error decreases remarkably compared to the 2C-GMM error. The log-likelihood that expresses how well the model explains the underlying data, is significantly higher for all conditions across subjects for the 3C-GMM. In addition to the statistical advantages, the 3C-GMM approach results in more plausible sub-distributions. With the 2C-GMM approach the correlated distribution includes searchlights with negative correlation coefficients which appears somehow counterintuitive and the non-correlated distribution is centered around negative correlation coefficients. The 3C-GMM approach results in a model with a non-correlated distribution centered around zero for MO and only very few negative correlation coefficients included in the correlated sub-distribution for all conditions. The non-correlated distributions for VP and SEM are shifted slightly towards positive correlation coefficients, similar to the complete searchlight distribution.

Correlated region maps

The inflated brain images in Fig. 3.7 depict those regions that include voxels whose searchlight classification results correlate positively or negatively with stimulus feature differences of the corresponding conditions. The determination as correlated or anti-correlated voxels is based on a three-component GMM: correlated regions are those with an a-posteriori probability smaller than $p = 0.05$ for the uncorrelated and for the anti-correlated searchlight distribution, respectively. Anti-correlated regions are determined correspondingly from the quantity of searchlights with a smaller a-posteriori probability than $p = 0.05$ for the uncorrelated and the correlated searchlight distribution. The correlation coefficients are color-coded individually for every condition to permit a better visual separability of high and low values.

In the following, *correlating searchlights* describe those searchlights whose BOLD response classification analysis leads to AUC results that show high positive correlation with specific differences in stimulus properties between the investigated conditions. *Correlating regions* describe the according anatomical regions that contain many correlating searchlights or voxels. An equivalent terminology is applied to anti-correlation effects.

The quantitative statistical analyses for single subject CRMs are presented in Fig. 3.8 for regions of the temporal lobe, in Fig. 3.9 for regions of the frontal lobe and in Fig. 3.10 for regions in other cortical regions. Filled bars (single subject results) and circles (group results) depict positive correlation effects, while open circles depict anti-correlation effects in group results. Negative correlation or respective anti-correlated regions are only found rarely in single subject results, depicted by crosses.

The quantitative results are only shown for cortical structures with a minimum correlated portion of searchlights of 5% in at least five single subject CRMs. Since these criteria are fulfilled in very few structures in the parietal lobe, occipital lobe and structures of the limbic system, results from structures in these regions are presented collectively.

Acoustic features - modulation and voice pitch

Cortical structures that are engaged in the cognitive processing of modulation and voice pitch information from acoustic stimuli largely overlap. As shown in Fig. 3.7, differences between stimuli correlate primarily with the performance for classification analysis comparing BOLD patterns in the temporal lobe and small parts of the frontal and parietal lobe near the central sulcus. A comparison of correlation effects in the particular cortical structures determined by the AAL template in Fig. 3.8 shows that

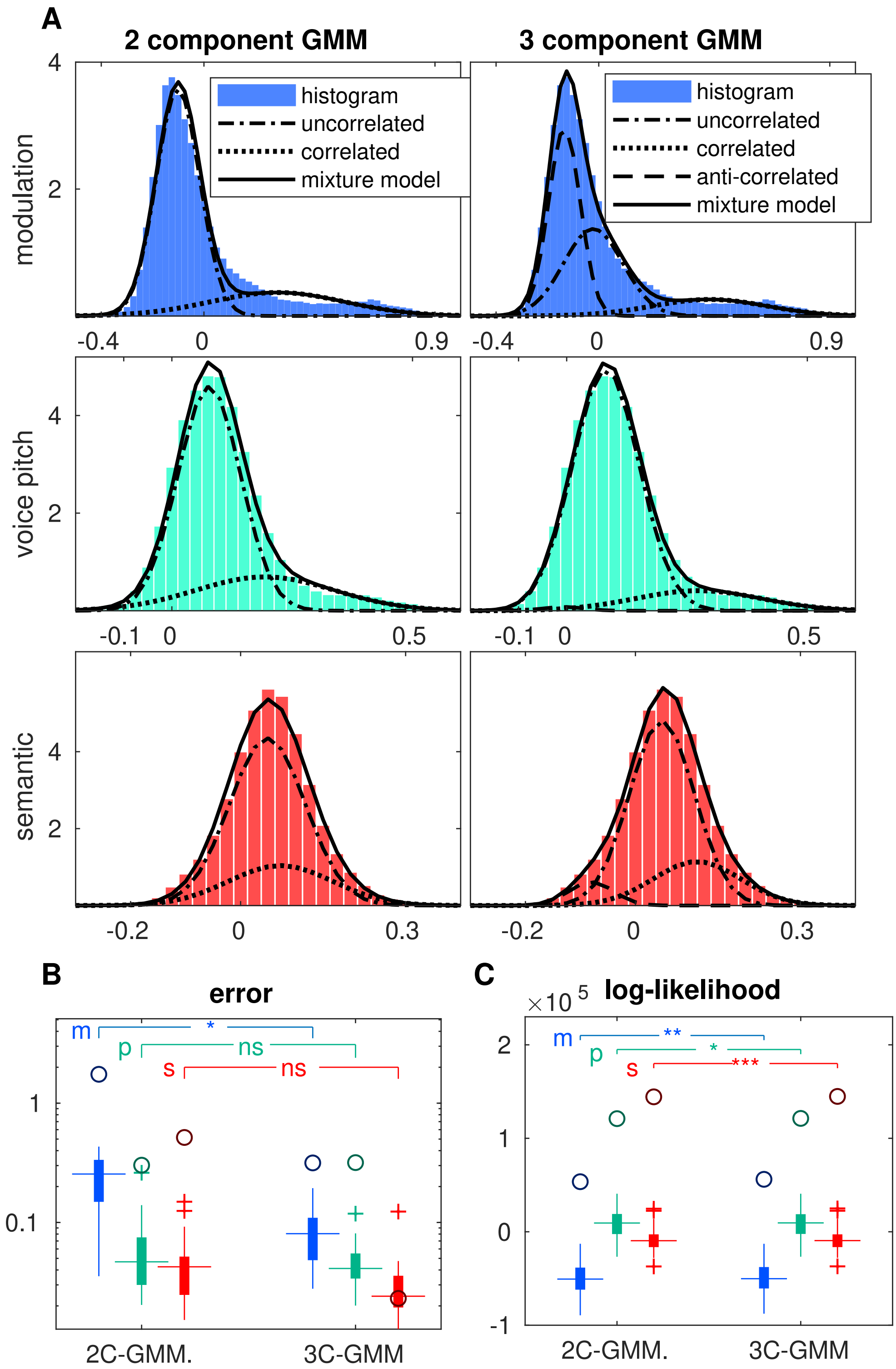

Figure 3.6.: Comparison of a two-component GMM (2C-GMM) vs. a three-component GMM (3C-GMM) for the distribution of correlation coefficients obtained from the Spearman-ranking correlation analysis. Panel (A) shows the distribution of group mean CRMs, overlain with the distribution obtained from the mixture models. For all the features the 3C-GMM, assuming the existence of an anti-correlated searchlight distribution, shows a better accordance with the data. The error, expressing the difference between mixture model and the data is shown for single subject results as box-plot in panel (B) for modulation (blue), voice pitch (green) and semantic (red). Horizontal lines designate the medians, filled boxes the inter-quartile ranges, vertical lines the 5% and 95% percentiles, crosses the outliers and circles the corresponding values for the group CRMs. The final log-likelihood on panel (C) expresses how well the model explains the data. The symbols are chosen equivalent to panel (B).

large parts of the HG and the STG correlate with MO on single subject and especially on group level. A slight left-lateralization can be observed, in particular in the medial HG. Anti-correlation effects can only be found in single subject outliers results in the ITG.

Condition differences for VP show very small correlation effects in the primary auditory cortex (medial HG), but almost the full lateral part of the HG consists of correlating searchlights. Correlated regions in the STG and the MTG are smaller than for MO, but consistently observable across single subject results and group results. Outside of the temporal lobe, correlation effects with MO can be found in the IFG and the rolandic operculum (RO) as well as the PrG, the SMG, and small effects in the posterior part of the middle cingulate gyrus (pMCgG). Effects for VP can only be found in the anterior part of the RO as well as in the OpIFG and the TrIFG.

Abstract feature - semantic

The CRMs in Fig. 3.7 show a main difference between correlated regions from SEM compared to the acoustic conditions: Group correlation analysis reveals considerably large anti-correlated regions in the temporal lobe, mainly located in the STG, the HG and the INS, depicted in Fig. 3.8. Correlated regions are located at the most anterior and posterior end of the temporal lobe, in the inferior temporal gyrus (ITG), the IFG as well as in the posterior part of the cingulate gyrus (CgG). Quantitative results across corresponding cortical structures in Fig. 3.8 to Fig. 3.10 indicate increasing correlation effects for SEM with increasing distance to core auditory areas, with no correlation in the STG, some effects in the MTG and slightly higher effects in the ITG. Single subject results show no large effects in the frontal lobe areas. However, in the TrIFG in the left-hemisphere correlated regions for SEM can be found. Effects on single subject level as well as on group level can be found in the precuneus (PCun), the AnG and the posterior half of the CgG. Yet, only in small parts of the particular regions high correlation coefficients could be found. Even though effects are small, a clear left lateralization can be observed. Positively correlated regions are very rare and small in the right hemisphere, and also anti-correlated regions are smaller as compared to the left hemisphere. Anti-correlated regions can only be located on group result level in the auditory cortex, in the RO and in the SMG - remarkably those regions that show the largest correlation effect for the acoustic conditions, in particular MO.

Specialization of cortical areas

Results that show areas with consistently larger portions of correlated regions for one stimulus feature compared to the other features indicate a specialization of these cortical structures for the processing of the particular stimulus feature. To quantify differences across stimulus features within the anatomical areas, differences between the size of correlated regions within anatomical areas are determined with a signed rank test across single subject results. The corresponding results are displayed in Fig. 3.11. It must be pointed out that the absence of large differences within an anatomical area does not at all exclude that this area is involved in the processing of one feature. It only indicates that this area is not specialized for one feature but involved in multiple stages of the stimulus processing.

Many regions in the temporal lobe and in the IFG seem to show higher correlation to the acoustic conditions (MO and VP) compared to SEM, as indicated by dark blue and green boxes, whereas the primary auditory cortex (mHG) shows stronger specialization to MO than to VP. In most areas outside of the temporal and the frontal lobe, only small differences between the sizes of correlated regions can be observed. However, the posterior part of the cingulate gyrus (pMCgG and pCgG), especially in the left hemisphere, shows significantly larger correlated regions for SEM compared to MO and VP. Single subject results, on which the results in Fig. 3.11 rely, do not show effect-differences for the anterior ITG, the superior PCun and AnG. However, for group results these regions show larger correlation effects for SEM compared to the acoustic conditions. In the TrIFG single subject results show

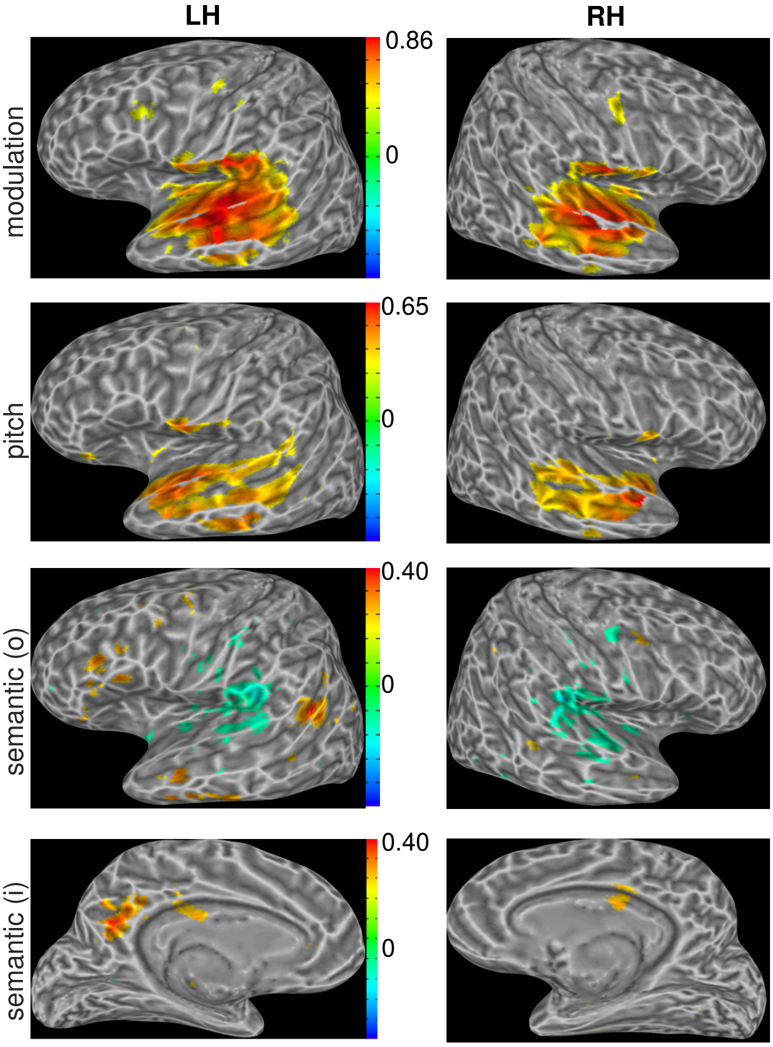

Figure 3.7.: Group correlation region maps (CRM). Regions which show high correlation between average classification performance results across subjects and differences between contrast-conditions are presented on inflated brain maps. Correlated regions can be located in superior (STG) and middle (MTG) and inferior temporal gyrus (ITG), from anterior to posterior regions, inferior frontal regions, rolandic operculum, bilaterally and precentral gyrus (PrG), postcentral gyrus (PoG), sub-parietal sulcus and small parts of middle cingulate gyrus, left lateralized. Correlated regions for pitch correlation analysis are located in regions identical to modulation correlation. The correlation with semantic differences results in correlated regions in the anterior part of middle temporal gyrus (MTG) in the left hemisphere (LH), inferior frontal sulcus (bilaterally), posterior cingulate gyrus (pCgG, LH) and occipito-temporal regions (LH). The results are presented at an FDR-corrected significance threshold of $p<0.05$

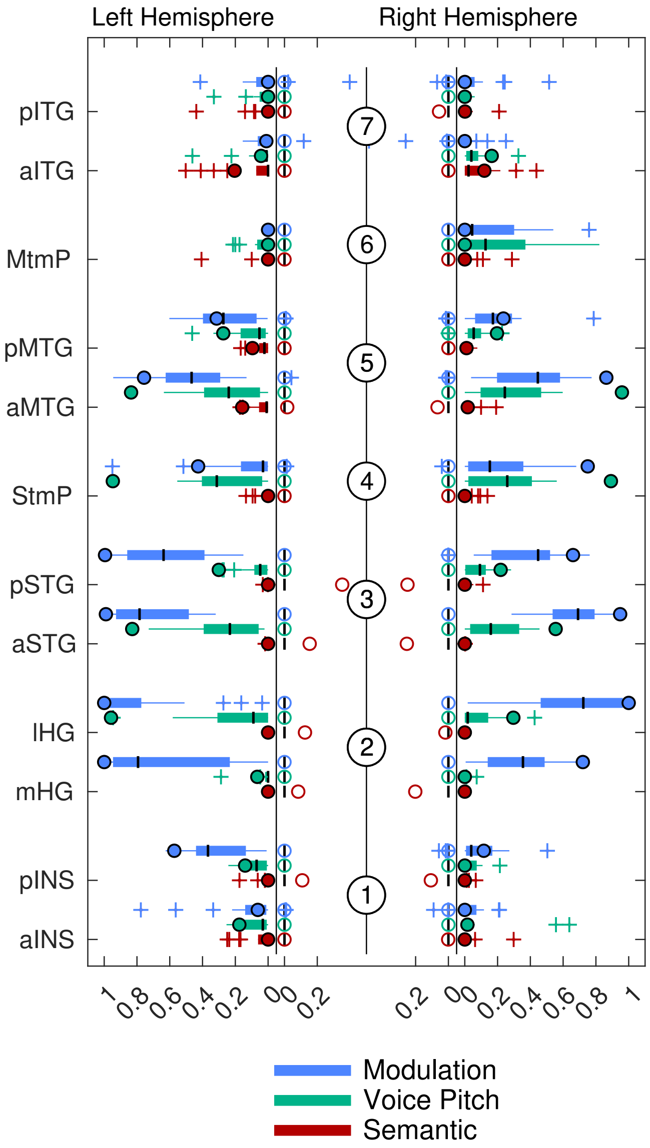

Figure 3.8.: Quantitative correlation analysis - Temporal lobe. The box-plots display the median portion of correlated searchlights from all searchlights within an anatomical area, determined by the AAL template across all subjects (black lines), with the inter-quartile range shown as colored bars and outliers plotted as crosses. The portion of correlated regions in the group CRMs are shown as filled circles. The location of anatomical regions can be found in Fig. 3.4 C and D. Correlation with modulation differences results in largest correlated regions in the primary auditory cortex (PAC) regions, showing also high variance across subjects. Correlation with pitch results in considerably smaller sizes of correlated regions in the superior and the middle temporal gyrus (STG and MTG) as well as superior temporal pole (StmP). Correlation with semantic differences can only be found in the anterior part of the inferior temporal gyrus (aITG). The group analysis shows correlated regions for modulation analysis across almost the full STG and MTG, and Heschl's Gyrus (HG). Group results for pitch analysis are comparable to those obtained from modulation analysis, however, correlated regions in the medial part of Heschl's Gyrus (mHG) are of decreased spatial extent. Correlation with semantic differences can only be found in th anterior part of ITG, left lateralized and anterior part of fusiform gyrus (aFuG), right lateralized. Single subject results are displayed at a significance threshold of $p<0.05$ with corrections for multiple comparison, group results are depicted at an FDR-corrected significance threshold of $p<0.05$

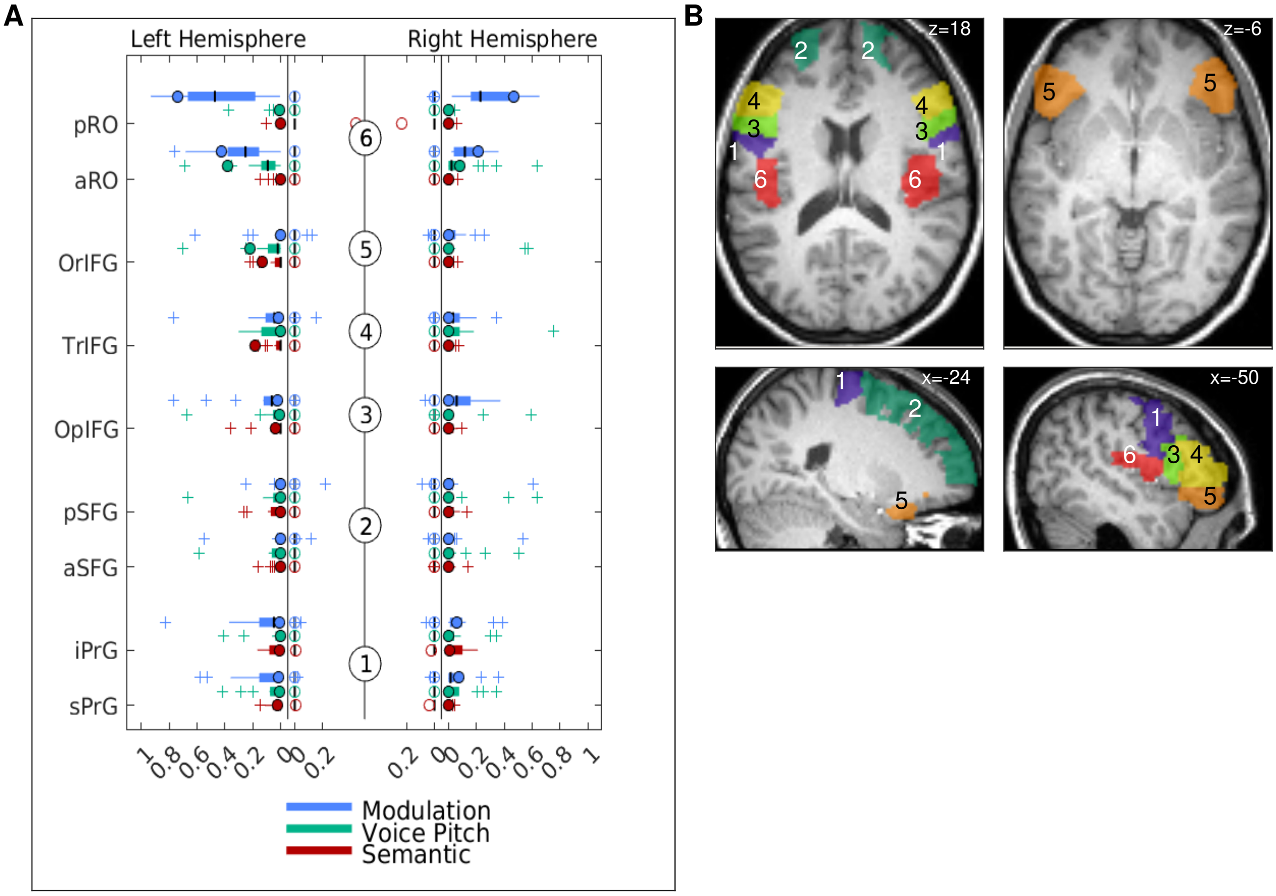

Figure 3.9.: Quantitative correlation analysis - Frontal lobe. The box plots in panel A display the median portion of correlated searchlights from all searchlights within an anatomical area determined by the AAL template across all subjects (black lines), with the inter-quartile range shown as colored bars and outliers plotted as crosses. The portion of correlated regions in the group maps are shown as filled circles. The location of anatomical regions mentioned in panel A can be found in panel B. Correlated regions obtained from modulation correlation analysis are most prominent in the posterior and anterior part on rolandic operculum (pRo, aRo) and in the aRo for pitch correlation analysis. In the inferior frontal gyrus (OrIFG, TriIFG, OpIFG) and the precentral gyrus (PrG) modulation correlation analysis leads to additional correlated regions. For single subjects, correlated regions obtained from semantic correlation analysis can be found in the medial frontal gyrus (MFG), thr superior frontal gyrus (SFG) and the supplementary motor area (SMA). Single subject results are displayed at a significance threshold of $p<0.05$ with corrections for multiple comparison, group results are depicted at an FDR-corrected significance threshold of $p<0.05$

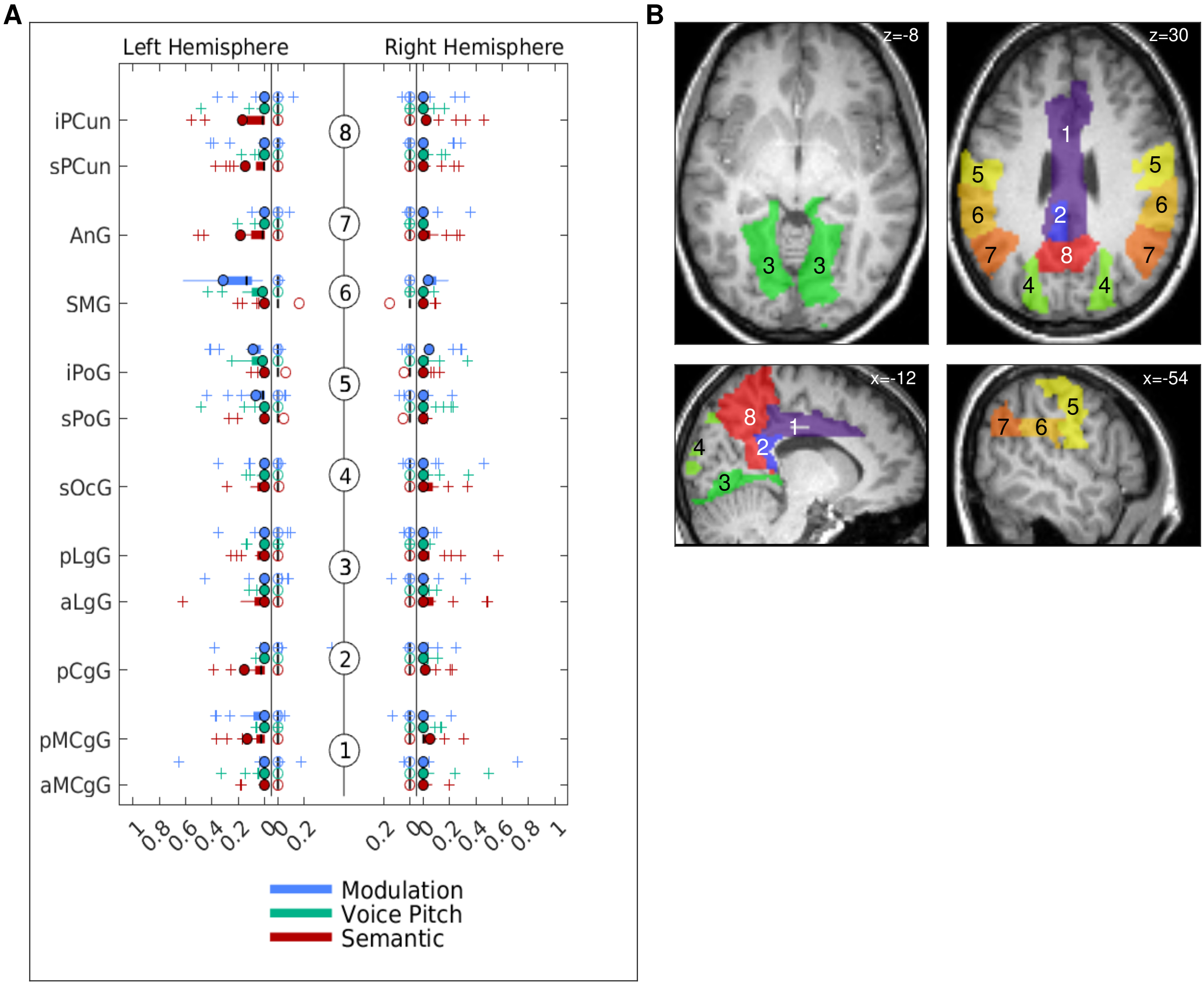

Figure 3.10.: Quantitative correlation analysis - regions outside of temporal and frontal cortex. The box plots in panel A display the median portion of correlated searchlights from all searchlights within an anatomical area, determined by the AAL template across all subjects (black lines), with the inter-quartile range shown as colored bars and outliers plotted as crosses. The portion of correlated regions in the group maps resulting from correlation of average classification performance across subjects and differences between contrast conditions are shown as filled circles. The location of anatomical regions mentioned in Panel A can be found in panel B. Modulation correlation analysis results in correlated regions in the supramarginal gyrus (SMG) and the inferior part of the postcentral gyrus (PoG). The SMG, the inferior parietal lobule (iPL) and iPoG show correlated regions for group analysis, however, single subjects results are below significance threshold in these areas. In the angular gyrus (AnG), superior parietal lobule (sPL) and posterior cingulate gyrus(pCgG) semantic correlation analysis effects can be observed. Single subject results are displayed at a significance threshold of $p<0.05$ with corrections for multiple comparison, group results are depicted at an FDR-corrected significance threshold of $p<0.05$

significantly larger correlated regions for *MO* compared to *SEM* in both hemispheres, while group results show a larger correlated region for *SEM* compared to the acoustic conditions. In conclusion temporal regions and large parts of the frontal lobe seem to be more correlated to acoustic stimulus processing, while a network of cortical structures in the parietal lobe and in the limbic system responds to semantic processing tasks.

3.5. Discussion

The aim of this study was the identification of cortical structures that are engaged in different aspects of speech processing, in particular the extraction of semantic content from acoustic stimuli. To disentangle these different processes, subjects were presented with six acoustic conditions, semantically valid sentences (clear and in noise), speech-like stimuli that are physically identical to normal speech but without semantic content (clear and in noise), noise vocoded speech and speech-simulating noise. Corresponding BOLD responses have been analyzed with the multivariate searchlight classification approach and results have been evaluated on single subject level and group level for 13 different contrasts between two different conditions respectively.

The contrasted conditions differ in multiple features whose corresponding effects on BOLD responses are not separable within common contrast investigation. The resulting group level maps show not very pronounced differences that permit the identification of cortical structures engaged in the particular speech feature processing. Therefore, these 13 contrasts were ranked regarding the contrasted condition's feature differences and a Spearman rank-correlation coefficient was calculated for the relation between classification performance results obtained from searchlight classification and the respective condition's speech feature difference rank.

The correlation analysis permits the identification of structures whose BOLD response differences for particular conditions show anti-correlation effects with stimuli feature differences. The anti-correlation indicates that for contrasts with large stimulus feature differences low classification performance results are found, while for small feature differences large classification performance results are found. This can be explained by the fact that contrasts with small differences between a specific feature might hold simultaneously large differences for another feature that permits a good separability of BOLD responses corresponding to the contrasted conditions. The second part of the explanation is based on the nature of BOLD responses that only represent indirectly the activation of cortical structures via the flow of oxygenic blood. If a cortical structure that is not engaged in a cognitive task is located close to structures that are active during the cognitive task, the inactive region might provide additional oxygenic blood for the active region, potentially leading to a lower recorded BOLD response compared to an active region and, particularly, other inactive regions not adjacent to an active region. Large feature differences lead to large cognitive demands in specific cortical structures which decreases the measured "activation" in neighboring inactive structures.

This motivates the separation of distributions of correlation coefficients across all brain voxels into three sub-distributions, namely a correlated, a non-correlated, and an anti-correlated distribution with a GMM in comparison to the two-component Gaussian mixture model, separating informative from non-informative searchlights. A direct comparison of the two-component GMM and the three-component GMM shows advantages for the three-component GMM in both dimensions, the difference between the model and the underlying data as well as the log-likelihood that serves as quality measure during the Equalization Maximazation (EM) algorithm.

With the correlation analysis, cortical structures were identified whose BOLD responses show correlation effects with specific stimulus features that could not be identified on this level with common

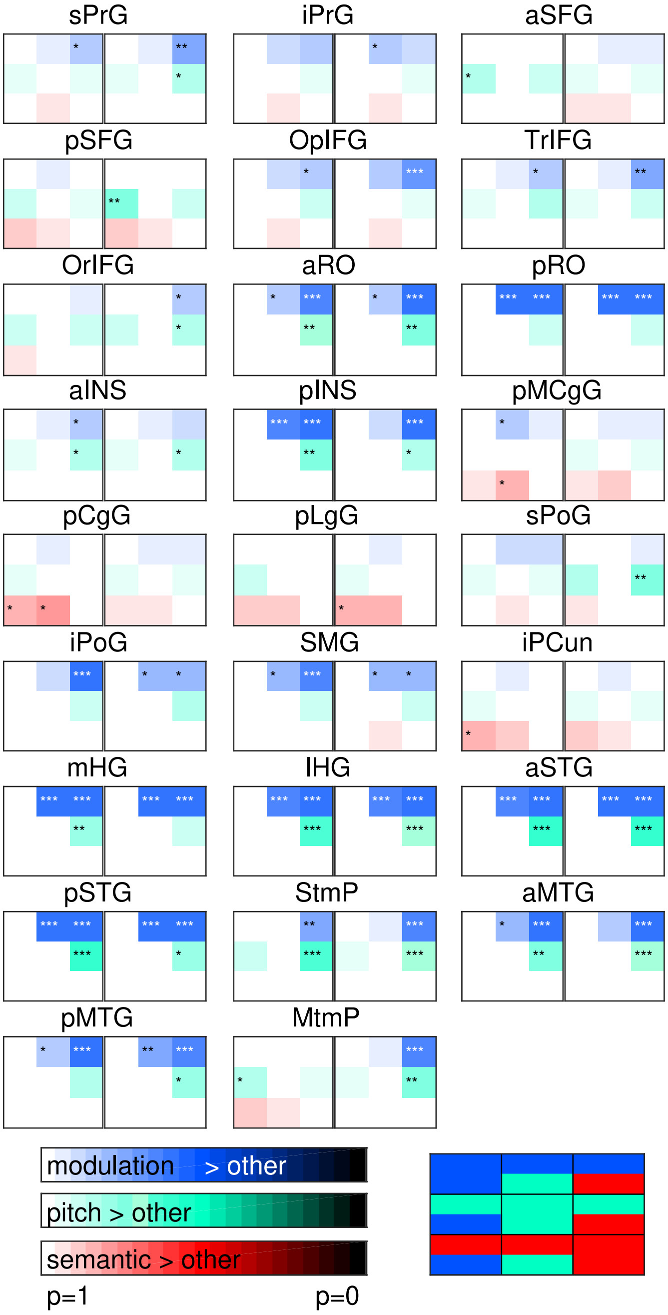

Figure 3.11.: Specialization of areas to the processing of stimulus features. The differences between sizes of correlated regions within determined anatomical areas across single subject results were calculated with the signed rank test. The left panels show the results for the left hemisphere, the right panels the respective right hemisphere results. Only areas, where significant differences between the sizes of correlated regions were identified, are shown. All panels are divided into nine sub-boxes. First row boxes (blue) display logarithmic color-coded p-value results obtained from a signed rank test across single subject results to test whether correlated regions for MO analysis are larger than those obtained from VP analysis (second column) or from SEM analysis (third column). The second row (green) shows corresponding results to test whether VP correlated regions are larger than MO correlated regions (first column) or SEM correlated regions (third column). Third row results show corresponding results to test whether SEM correlated regions are larger than MO correlated regions (first column) or VP correlated regions (second column). Significant differences are marked as * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

group result maps showing differences between presented conditions, e.g., in structures of the lymbic system. According to the second hypothesis H2, the present results indicate a hierarchical processing structure for acoustic features in the primary and secondary auditory cortex and in areas known to be involved in speech processing, while semantic processing involves a large network of structures across the different cortical lobes. Further effects are discussed below in the particular section for the investigated speech features.

3.5.1. Amplitude modulations

The stimuli used in this study exhibit either no amplitude modulations ($Noise$) or amplitude modulation with speech rate (semantic speech (in noise), non-semantic speech (in noise), noise vocoded speech). Therefore, results obtained from the correlation analysis described here are restricted to amplitude modulation in the speech rate or with speech rhythm. It has to be noted, additionally, that results obtained from this study do not permit the differentiation of cortical structures engaged in the processing from those who are involved only in the perception of modulation characteristics.

In the primary auditory cortex (PAC), located in the medial part of Heschl's gyrus (HG), correlation effects for modulation are smaller compared to those in the lateral part of the HG (cf. Fig. 3.8), but still very prominent. The high correlation effects in PAC are expected to be induced by the perception of amplitude modulation, a basic stimulus processing step, that is larger in the left hemisphere, comparable to BOLD responses found for processing of amplitude modulation in speech rate in the left HG and the superior temporal lobe (Giraud et al., 2000). Furthermore, in the belt and parabelt areas, MO correlation effects are larger in the left hemisphere compared to the right hemisphere. This might depict the speech-like nature of amplitude modulations present in the stimuli, whose processing has previously been found to induce cortical activation in superior temporal regions, e.g., by Kubanek et al. (2013), who found the STG to be activated during the processing of speech envelopes, or Hall et al. (2000), who found higher activation for modulated signals compared to static signals in the STG anterior to the HG. An engagement of the posterior STG in the cognitive processing of amplitude modulation has previously been found for, e.g., the comparison of flattened speech to normal speech (Meyer et al., 2004), isochronous vs. non-isochronous pseudosentence processing (Geiser et al., 2008), isochronous vs. non-isochronous syllable repetition (Riecker et al., 2002) or prosody, indicated by intonational phase boundaries (Ischebeck et al., 2008). Giraud et al. (2000) also reported the location of areas that are involved in the processing of amplitude modulations with speech rate in the posterior temporal lobe.

Other cortical structures that have previously been found to be involved in prosody processing of speech stimuli are the RO (Ischebeck et al., 2008; Hervé et al., 2012; Geiser et al., 2008) and the left temporo-parietal junction (Hervé et al., 2012; Geiser et al., 2008). The posterior part of the RO as well as the inferior PoG and the SMG, with the inferior part lying in the temporo-parietal junction, also show significantly higher correlation to modulation effects compared to correlation effects to voice pitch or semantic. This indicates a specialization of these areas to the temporal analysis of speech stimuli.

3.5.2. Voice pitch

The percept of a voice pitch and the recognition of a sound to be a human voice is a basic aspect of human communication. Information like the speaker's identity (Formisano et al., 2008) or the speaker's emotional state can be extracted from voice pitch information. However, the amount of cognitive resources required for this process is expected to be smaller than for, e.g., processing modulation that reflects the speech envelope and can, in combination with high spectral detail, carry information with

regard to the content of speech. Both, the number of involved cortical structures as well as the size of correlated regions for voice pitch perception, are smaller as compared to those identified for modulation processing, although also for voice pitch processing no differentiation between analysis and perception was made. The largest correlation effects were found in the left lateral part of the HG, which has previously been identified to be involved in the perception of human voice (Uppenkamp et al., 2006), as well as in the differentiation between female and male voices (Lattner et al., 2005) and even the identification of speaker's identities (Formisano et al., 2008). Large correlation effects were also found in the StmP bilaterally as the most anterior part of the superior temporal lobe. The anterior part of the temporal lobe also showed higher activation during the perception of human voices as compared to animal sounds (Fecteau et al., 2004) and female voices as compared to male voices (Lattner et al., 2005), while other studies identified anterior to posterior regions of the superior temporal lobe to be voice-selective (Moerel et al., 2012; Pernet et al., 2015; Liebenthal et al., 2005).
In addition, the MTG showed considerable correlation between differences in BOLD responses and differences in voice pitch content. Activation in this area has been found for contrasting normal sentences with speech envelope shaped noise (von Kriegstein et al., 2003), vocal and non-vocal sounds (Belin et al., 2000) or to be sensitive to the acoustic complexity of speech sounds controlled by the number of presented frequency channels (Warren et al., 2006). These findings support an involvement of the MTG in human voice recognition or perception, as indicated by the present results. Only small correlation effects can be found outside of the temporal lobe, in the anterior part of the RO and the IFG, left lateralized, consistent with the identification of increased activation in these areas during the processing of spectral and temporal cues of speech (Gandour et al., 2002), flattened speech compared to degraded speech (Meyer et al., 2004), and vocal sounds compared to non-vocel sounds (Pernet et al., 2015).

Since no differentiation between different hierarchical steps of voice processing has been considered in the correlation analysis, the results do not permit insights into hierarchical voice processing. However, those areas identified to be involved in human voice pitch processing accord with those found in previous studies, are of different cognitive demand and range from voice perception (Uppenkamp et al., 2006) to the identification of speaker's identity (Formisano et al., 2008) and still seem to share involved cortical networks.

3.5.3. Semantic

The main aim of this study was the identification of cortical structures that are engaged in the semantic processing of spoken language. A network of cortical structures, whose BOLD classification performance results correlate positively with differences in semantic content, could be identified across all cortical lobes. Temporal lobe regions that show correlation effects with semantic content are located at the most anterior and posterior end of the temporal gyri, predominantly in the ITG and the MTG, including the temporal pole. While the effects in the MTG are left lateralized with only small differences in the weighting between anterior and posterior regions, the observable bilateral effect in the ITG is more pronounced in anterior regions. Semantic processing on sentence level has been associated with cortical processing in both, the MTG (Mashal et al., 2009; Rogalsky and Hickok, 2009; Adank and Devlin, 2010; Devauchelle et al., 2008; Obleser and Kotz, 2010; von Kriegstein et al., 2003; Kuperberg et al., 2000) and the ITG (Sharp et al., 2009; Kuperberg et al., 2000; Rodd et al., 2005; Awad et al., 2007), while semantic processing on word level has been rather reported in the MTG (Davis and Gaskell, 2009; Diaz and McCarthy, 2009; Hwang et al., 2009). This could lead to the hypothesis that the MTG is predominantly involved in the semantic processing of word level semantics which supports the interpretation of heard sentences by identifying and integrating series of phonemes to words and their meaning. Even though the correlated regions obtained from group level analysis represent only about 20% of the MTG, this left lateralized process seems to be spatially consistent across participants, taking into account that the group level correlated region is larger than the highest individual correlated regions in this area. The bilateral effect of semantic correlation in the anterior

ITG, however, supports an involvement in sentence semantic processing on an on average higher level than in the MTG. This comes along with higher inter-individual differences induced by larger correlated regions obtained from individual single subject analysis compared to group level analysis. The bilateral nature of this effect also supports the hypothesis of the involvement of the corpus callossum in auditory sentence processing to support the inter-hemispheric interaction for prosody and syntactic processing (Friederici et al., 2007). This region also shows positive correlation effects with semantic differences as can be seen for group level correlated region maps in Fig. 3.7.

The different parts of the IFG have also been associated with higher order speech processing in several studies. Regions within the OrIFG have been identified to be involved in the recognition of words (Davis and Gaskell, 2009; Nosarti et al., 2010), sentence comprehension (Sharp et al., 2009; Tyler and Marslen-Wilson, 2007) and semantic retrieval in general (Price, 2010). In previous studies, the TrIFG has shown increased activation for words compared to pseudo-words (Kotz et al., 2009), reading metaphors compared to literal sentences (Schmidt and Seger, 2009) as well as word generation (Jeon et al., 2009; Whitney et al., 2009; Heim et al., 2009). In this study, the size of correlated regions in the OpIFG and the TrIFG are comparable, whereas the TrIFG shows a slightly larger correlated region obtained from group CRM. Additional correlated regions for the semantic processing to those discussed above could be identified in rather dorsal regions of the parietal and occipital lobe as well as in structures of the limbic system. Single subject correlated regions in these areas are larger as compared to group correlated regions, indicating a large spatial inter-individual variance in activation for semantic processing. The high inter-individual differences indicate the involvement of these areas in highly abstract tasks that demand individual solving-strategies and therefore different activation patterns across subjects (Seghier and Price, 2016).

On group level, correlated regions could be identified in the left PCun, while single subject correlated regions of individual subjects show bilateral correlation effects in this area. In previous studies, an engagement of the PCun in different semantic speech processing stages has been reported. These include the identification of existing words (Davis and Gaskell, 2009), the interpretation of metaphors (Mashal et al., 2009; Schmidt and Seger, 2009), the evaluation of sentences' plausibility (Mashal et al., 2009), and narrative language processing in general (Whitney et al., 2009). In other studies, deactivation for speech-specific tasks has been observed in the PCun (Abrams et al., 2012; Kouider et al., 2010). Correlated regions on group level were also identified in the AnG, having been associated to sentence processing (Sharp et al., 2009; Mashal et al., 2009; Obleser and Kotz, 2010; McGettigan et al., 2014) as well as semantic context of words (Golestani et al., 2009) or semantic relatedness between words (Sharp et al., 2009). The pCgG as well as the pMCgG have shown left lateralized group level correlation effects, consistent with previous findings for sentence processing (Mashal et al., 2009; Awad et al., 2007; Obleser et al., 2007a; Schmidt and Seger, 2009) and lexical processing of words (Newman et al., 2001). In the lingual gyrus (LgG), no correlation effect could be observed on group level. However, single subject results showed correlated regions in up to 50% of the LgG, in particular in the posterior part. The comparison of single subject correlated region sizes for the different speech features indicates a higher correlation in the right LgG for semantic features as compared to the acoustic features. Being associated with speech production or word generation (Hwang et al., 2009), it also has been shown to be involved in higher order sentence processing (Kuperberg et al., 2000; Schmidt and Seger, 2009; von Kriegstein et al., 2003) and word categorization (Bilenko et al., 2009). Negatively correlated regions for semantic processing could only be identified on group level in the posterior STG for a larger portion as compared to the anterior STG and in the HG and posterior INS. As described above, these negative correlations might arise from orthogonal feature differences between contrast for different features, and might indicate that these areas are not activated for semantic processing, while they probably are for other feature recognition processes.

In summary, from correlation analysis a wide network across all cortical lobes could be identified, with probable increasing distance of involved areas to the auditory core for increasing complexity of the cognitive processes of semantic processing. This supports the first hypothesis H1 of a largely distributed network for semantic processing of speech as compared to the processing network for acoustic speech features. Regions that seem to be specialized for the semantic processing of speech and not to the acoustic processing appear to be the posterior half of the cingulate gyrus, the lingual gyrus and the precuneus. In single subject CRMs, correlated regions in these areas for acoustic features can be found as well, which might indicate an integration of these acoustic features for the semantic interpretation in the particular areas. However, the size of correlated regions in the particular cortical structures is rather small, potentially supporting the hypothesis of a dynamic and individual network, which needs further systematic investigation to identify consistent processing strategies across humans.

3.6. Conclusion

In this chapter, results from an auditory fMRI study are presented, in which subjects were presented with six different acoustic stimulus conditions that differed in their content of semantic, voice pitch and amplitude modulation. Regions that are involved in different speech processing steps were initially identified within the searchlight algorithm in combination with an SVM classification and subsequent identification of significant results with the SCIM algorithm, from which informative region maps (IRMs) were obtained. However, informative regions across the 13 investigated contrasts did not permit a systematic identification of specialized regions for the processing of semantic content, voice pitch or amplitude modulation. The contrasts were ranked independently for each stimulus feature, serving as input for a Spearman rank-correlation. Based on this ranking, the relation between BOLD response classification performance and differences between conditions regarding the three named stimulus features was evaluated, resulting in correlated region maps (CRMs). The results show that the processing of abstract cognitive tasks can be realized via individual strategies across subjects and pooling results into group level maps might smear observable effects. Additionally, these abstract tasks are difficult to reproduce with stimulus conditions that differ only in the processing task of interest. It is, therefore, meaningful to combine different conditions and analyze cognitive responses to these conditions across contrasts.

Within the correlation analysis, separate networks for the processing of acoustic stimulus features and for the semantic processing of auditory perceived speech could be identified. While the temporal cortex and areas located in the inferior frontal gyrus seem to be more sensitive to acoustic properties of speech, regions in a network ranging from the far anterior and posterior middle and inferior temporal gyrus, over motor regions, to structures in the limbic system seem to be involved in the semantic processing of speech.

4. Global Activation Patterns - Neural Networks for the Processing of Spoken Language

4.1. Abstract

Humans cope with the interpretation of spoken language every day remarkably well even in very distorted acoustic situations. However, it has not been possible yet to fully connect the different findings from various studies in which the neural correlates of different aspects of this complex task have been investigated. While there are well established models for the acoustic processing of speech, the extraction of meaning from spoken language is much less investigated. A large obstacle is the choice of stimuli that permit the disentanglement of physical from abstract characteristics. Therefore, it is often only possible to examine one specific feature of spoken language interpretation. In this chapter, results from an auditory fMRI study are presented, in which neural responses to stimuli are compared that are both physically similar to natural speech. While one condition comprises valid spoken sentences, the contrast condition consists of randomized speech fragments in six different languages, that are controlled for acoustic properties of speech but do not have perceptible semantic content or syntactic structure. To account for global patterns of activation across the cortex, the psychophysiological interactions in fMRI data have been investigated in addition to the statistical parametric maps. For intelligible speech, a largely left lateralized network has been identified between the temporal lobe, the frontal lobe, the temporo-parietal junction and the cingulate gyri. The same areas, although with different connections, are part of an excitatory and an inhibitory network which have been identified for the processing of natural sounding speech that is unintelligible.

4.2. Introduction

Spoken language is a fundamental part of human social interaction and remarkably robust against various distortions. We can not only understand language in very distorted versions but also the integration from phonemes to an abstract meaning can be achieved by human listeners for different languages that might contain different phoneme groups and patterns. The cognitive processes that reflect the interpretation of spoken language have been the object of many studies. Still, the neural interactions in the cortex have not been fully understood yet.

Cortical structures that are frequently described to be part of a language processing network are the frontal cortex, the temporal lobe structures, the supramarginal gyri, the angular gyri and the cingulate gyri. Functional or effective connectivity between these structures for semantic processing is mostly investigated for visual modalities (Snijders et al., 2009; Smirnov et al., 2014; Burianova et al., 2010; Davey et al., 2016; Canini et al., 2016). However, Graessner et al. (2019) presented subjects with auditory two-word phrases, differing in their plausibility, and they identified connectivity between the left inferior frontal gyrus (IFG), the bilateral pre-supramarginal gyri (SMG) and the angular gyrus (AnG) during a task of plausibility assessment.

Depending on the research question and the stimuli that are presented, the previously mentioned structures are coactivated in different subgroups. The processing of pseudo-words compared to normal speech has been associated with activation in the left IFG (Newman and Twieg, 2001; Davis and Gaskell, 2009; Baumgaertner et al., 2002; Kotz, 2002), in the left inferior parietal lobe (IPL) (Davis and Gaskell, 2009; Baumgaertner et al., 2002; Humphries et al., 2006), bilateral temporal lobe structures (Humphries et al., 2006; Davis and Gaskell, 2009) and the posterior cingulate gyrus (CgG, Davis and Gaskell, 2009). Semantic priming and the corresponding increased intelligibility of spoken language lead to increased activation in the same subgroup. Left IFG activation has been reported for priming effects in various studies (Clos et al., 2014; Smirnov et al., 2014; Sohoglu et al., 2012; Baumgaertner et al., 2002; Graessner et al., 2019). While Sharp et al. (2009) found deactivation effects in the superior temporal lobe bilaterally, these areas showed activation in studies by Obleser et al. (2007b) and Obleser and Kotz (2010). However, some studies did only identify corresponding activation in the left superior temporal lobe (Devauchelle et al., 2008; Sohoglu et al., 2012). Activation in the IPL or AnG is influenced by priming as identified by Sharp et al. (2009), Baumgaertner et al. (2002), Obleser and Kotz (2010), and Graessner et al. (2019).

Obleser et al. (2007b) found additional activation for priming in the SMG, middle frontal gyrus (MFG) and middle temporal gyrus (MTG) of the left hemisphere. The latter area showed bilateral effects in studies by Devauchelle et al. (2008) and Baumgaertner et al. (2002). The ability to differentiate between words and pseudo-words, as well as the advantage of priming effects for the intelligibility of spoken language, leads to cognitive memory demands : short-term memory for priming effects and intermediate-term to long-term memory for the identification of pseudo-words.

Dissolving ambiguity or the evaluation of plausibility requires an assessment of perceived spoken language and addresses only a few of the previously mentioned structures. In particular, the right hemisphere seems to be less included in this cognitive task. Rodd et al. (2005) and Bilenko et al. (2009) found bilateral activation in the IFG in addition to a left lateralized inferior temporal gyrus (ITG) and MTG activation for the processing of ambiguous words. However, Rogers and Davis (2017), Rodd et al. (2012) and Snijders et al. (2009) could only identify corresponding activation in the ITG, the MTG and the CgG of the left hemisphere. A similar pattern becomes apparent for the estimation of the plausibility of spoken language. An involvement of the left IFG and superior frontal gyrus (SFG) could be identified by Ye and Zhou (2009), while Mashal et al. (2009) found corresponding activation in the left MTG and posterior CgG. Both studies reported activation in the left AnG for plausibility estimation. These findings support the proposed model of increasing lateralization for increasing abstractness and cognitive demands during spoken language processing by Peelle (2012).

To investigate the processing of intelligible speech in general, a very popular method is the presentation of noise vocoded speech, for which the intelligibility can be controlled by the number of channels for the generation of the signal. While some studies use different numbers of channels (Obleser et al., 2007b; Obleser and Kotz, 2010; Blank and Davis, 2016), others compared intelligible noise vocoded speech to their spectrally rotated, unintelligible version (Scott, 2000; Narain, 2003; Eisner et al., 2010; Okada et al., 2010).

The left frontal cortex, in particular the IFG, seems to play a crucial role for the understanding of speech compared to the processing of degraded speech (Sharp et al., 2009; Clos et al., 2014; Obleser et al., 2007b; Obleser and Kotz, 2010; Eisner et al., 2010; McGettigan et al., 2012; Sohoglu et al., 2012). The left AnG consistently shows activation across studies with a study design incorporating noise vocoded speech (Evans and Davis, 2015; Awad et al., 2007; Obleser et al., 2007b; Obleser and Kotz, 2010; Eisner et al., 2010), too. Activation across the left temporal lobe, in the MTG (Clos et al., 2014), the STG (Scott, 2000; Narain, 2003; Obleser et al., 2007b; Obleser and Kotz, 2010; Okada et al., 2010; McGettigan et al., 2012; Sohoglu et al., 2012), and the fusiform gyrus (FuG) (McGettigan et al., 2012) has also been found to be associated with the perception of intelligible speech contrasted with degraded, non-intelligible speech.

Even though it is possible to learn understanding noise vocoded speech, there is a considerable perceptual difference to natural human speech. To avoid a potential interaction between the extraction of meaning from spoken language and acoustic deviations from natural speech, the stimuli we used consist of natural human speech recordings without physical speech degradation. We presented subjects during fMRI acquisition with spoken everyday sentences and compared the neural responses with those corresponding to the presentation with a speech-like stimulus of real human speech recordings, without syntactic structure or semantic meaning. The stimuli are a pseudo-randomized composition of short speech fragments, obtained from speech recordings in six different languages, that are controlled for acoustic properties of speech. Effective connectivity, indicated as psychophysiological interactions, is investigated to disentangle overlapping cortical activations during the acoustic and the abstract processing of spoken language. The regions of interest (ROIs) comprise the temporal lobes, the superior, middle and inferior frontal gyri, the cingulate gyri, the supramarginal gyri and the angular gyri. Three networks, one for the processing of acoustic features of spoken language, one for the interpretation of valid spoken language and one for the processing of non-semantic speech could be identified, that are in line with results from previous studies and relate those findings to a considerable degree.

4.3. Materials and methods

4.3.1. Stimuli

The stimuli used in this study and the data analysis strategy have been described in the methods section of the previous chapter. Six auditory conditions and one silent condition are included in the design. Each example of the semantic speech condition is composed of two randomly chosen sentences from the Göttingen Sentence Test (Kollmeier and Wesselkamp, 1997, GÖSA) corpus with meaningful everyday content, respectively. Non-semantic speech stimuli were fragments of 6 s duration from the International Speech Test Signal (ISTS, Holube et al., 2010). This is a signal that permits the impression of natural speech without containing any valid content. Speech-simulating noise obtained from the GÖSA material served both as a separate noise condition and as a masker for noisy semantic speech and noisy non-semantic speech. The noisy speech conditions were presented at an SNR of -6 dB. Noise vocoded speech (NVS) has been obtained by filtering a white noise carrier with the first twelve linear predictive coding (LPC) components of the GÖSA trials, resulting in a stimulus that cannot be understood without presentation of the natural stimulus immediately before. The silence condition served as baseline condition for the subsequent generalized psychophysiological interaction (gPPI) analyses.

4.3.2. Subjects and data-acquisition

The study was conducted with 19 voluteers (11 male, 8 female, 23.5 ± 2.6 years in age) with a 3 T Siemens MRI scanner. 18 of the 19 participants were native speakers of German and their results were included for further analyses. The previously described stimuli were presented in a randomized order for a passive listening task. For the temporal separation of the auditory stimuli and the noise during the EPI sequence, we used sparse imaging with a TR of 9 s including 2.9 s for the EPI data acquisition of a complete brain volume.

4.3.3. Preprocessing

The data preprocessing was conducted with the SPM software SPM12 (Friston et al., 1995) and included coregistration, spatial realignment, normalization to MNI space and Gaussian smoothing with a full width at half maximum of 5 mm. Subsequently, the data passed through the default denoising pipeline implemented in the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012).

4.3.4. Generalized psychophysiological interaction — gPPI

Psychophysiological interaction analysis is a method to analyze integrated brain activity and to identify networks between distal cortical areas for specific cognitive tasks. The approach is based on the well-known general linear model (GLM, Friston et al., 1995), where the neural activity $\mathbf{x}_{ij}$ in a voxel j at the time point i is modeled as:

$$\mathbf{x}_{ij} = \mathbf{g}_{i,1}\beta_{1j} + \mathbf{g}_{i2}\beta_{2j} + \ldots + \mathbf{g}_{iK}\beta_{iK} + e_{ij} \tag{4.1}$$

with $\mathbf{g}_{ik}$ being explanatory variables that describe covariates of no interest, e.g., the global cerebral blood flow (CRB), time, or indicator variables, indicating levels or factors regarding the experimental conditions under which the observation (response CRB) was made. The error e_{ij} is independent for each measure $\mathbf{x}_{ij}$. The model from Eq. 4.1 can also be written in matrix form as:

$$\mathbf{X} = \mathbf{G}\beta + e \tag{4.2}$$

with X being a response CRB matrix, containing elements $\mathbf{x}_{ij}$, one row per scan i and one column per voxel j. The matrix $\mathbf{G}$ represents the design matrix which comprises the elements $\mathbf{g}_{ik}$ and contains one row per scan i and one column per effect (covariate or factor) k. Column vectors β_j for every voxel j build a parameter matrix β. Later, Gitelman et al. (2003) added a step, where the factorial column is convolved with the hemodynamic response function (HRF), to account for the temporal nature of the BOLD signal and therefore provides a and better model of the neural activity. This allows for the application of this approach in studies with an event related design.

Assuming that there are two factors of interest relating to condition A and condition B, the design matrix $\mathbf{G}$ is composed of the column vectors $\mathbf{g}_A$ and $\mathbf{g}_B$ and a matrix consisting of column vectors representing confound variables $\mathbf{G}_c$, so that $\mathbf{G} = [\mathbf{g}_A\mathbf{g}_B\mathbf{G}_c]$, leading to:

$$\mathbf{X} = [\mathbf{g}_A\mathbf{g}_B\mathbf{G}_c]\beta + e \tag{4.3}$$

as described in Friston et al. (1997).

Assuming further that in a factorial design the interaction between condition A and B is of interest, the design matrix is expanded with the column vector $\mathbf{g}_A$ x $\mathbf{g}_B$ which is an element-wise multiplication of the two vectors $\mathbf{g}_A$ and $\mathbf{g}_B$.

$$\mathbf{x}_j = \mathbf{g}_A\mathrm{x}\mathbf{g}_B\beta_1 + [\mathbf{g}_A\mathbf{g}_B\mathbf{G}_c]\beta_G \tag{4.4}$$

Comparable to the interaction of two factors that influence the activity in a voxel j, a factor can modulate the activity $\mathbf{x}_n$ in another voxel or region n. This interaction may influence the activity in voxel or region i, which leads to the model in Eq. 4.5, where β_1 represents the contribution of the activity in voxel or region n that was modulated by the experimental factor A to the activity in voxel j and all other β-weights modeling confound parameters.

$$\mathbf{x}_j = \mathbf{g}_A\mathrm{x}\mathbf{g}_n\beta_1 + \mathbf{g}_A\beta_2 + \ldots + \mathbf{g}_K\beta_K + e_j \tag{4.5}$$

This approach supposes that the activity in a voxel or region j can also be modeled as the sum of influences from all other regions n, leading to:

$$\mathbf{x}_{ij} = \sum_n C_{jn}\mathbf{x}_{in} \tag{4.6}$$

where C_{jn} represents the effective connectivity between voxel or region j and voxel or region n, when all regions or voxels are included. There is no self-connectivity in this model (C_{jj}=0) and the matrix C is not symmetric ($C_{jn} \neq C_{nj}$).

The parameter β_1 in Eq. 4.5 only represents a contribution to the activity in voxel j but does not represent the effective connectivity, since only the activation in this one region was considered in the model. The more regions are included, the more comparable are the corresponding β-values to the effective connectivity weights that represent the psychophysiological interactions (PPI).

Considering the potential influence of all other brain regions leads to the following model for a single effect gPPI for N regions or voxels contributing to the activity in voxel j with a matrix $\mathbf{G}^*$, including the design matrix for region or voxel j and the time-series from regions n :

$$\mathbf{x}_j = \sum_n^N \mathbf{g}_A\mathrm{x}\mathbf{g}_n + \mathbf{G}^*\beta_{\mathbf{G}^*} + ej. \tag{4.7}$$

McLaren et al. (2012) introduced a generalized form of the gPPI to model psychophysiological interactions for more than one condition A or a difference vector $A - B$ between two conditions. In this model, the column vector $\mathbf{g}_A$ is replaced by a matrix with one column vector for each condition and permits the identification of single gPPI effects or similar effects across conditions.

4.3.5. Network identification

The connections between spatially distal cortical regions have been identified with the ROI-to-ROI analyses which shows the extent of connectivity between pre-selected ROIs. The seeds have been determined based on the automated anatomical template (AAL Tzourio-Mazoyer et al., 2002). Large regions within this template have been divided into two parts based on their direction of the largest spatial extent along the anterior-posterior axis, the superior-inferior axis or the medial-lateral axis, respectively. The ROIs' time-series have been summarized from all voxel time-series within the respective ROI in the first component of a singular value decomposition (SVD). A previous z-scoring of the data has explicitly not been performed to emphasize time-series from voxels that show the largest effects.

The choice of the set of ROIs is based on previous results from correlation analysis (cf. chapter 3) and results from studies described in the introduction. They incorporate the bilateral superior, middle and inferior temporal gyri (STG, MTG, ITG), Heschl's gyri (HG), the angular gyri (AnG), the supramarginal gyri (SMG), the fusiform gyri (FuG), the cingulate gyri (CgG) and the inferior, middle and superior frontal gyri (IFG, MFG, SFG). The areas were chosen to be selected bilaterally to investigate, if a left lateralization effect can be observed in the current data. Additionally, the results obtained from the correlation analyses described in the previous chapter, showed effects in the right hemisphere, even though they have been smaller than left-hemisphere effects.

According to the network-based statistics (NBS) presented by Zalesky et al. (2010) the connectivity or psychophysiological interaction between each pair of ROIs has been calculated as depicted in Fig. 4.1. The corresponding non-thresholded (upper row) and thresholded (lower row) non-symmetric connectivity matrices contain the F-values obtained from group level analyses for four conditions, respectively contrasts. The order of the ROIs in this matrix, which influences each ROI's neighborhood within the matrix, is determined by temporal correlations between time-series. To test the significance of gPPI-links between ROIs, clusters of supra-thresholded links in the connectivity matrix are identified, as depicted in the lower panels of Fig. 4.1, and the number of supra-threshold links within this cluster (the cluster's size) is stored. In a subsequent permutation test, condition labels are permuted and the same procedure is repeated M (in this study 1000) times, each time storing each cluster's size as part of the underlying null-distribution. Each cluster's p-value is then determined by the number of permutations that leads to a higher or equal cluster size as compared to the cluster size observed in the original data, normalized by M.

4.4. Results

In this study we compare the activation maps obtained from the GLM analyses with the connections graphs obtained from generalized psychophysiological interaction (gPPI) analyses for BOLD responses to different acoustic conditions. These conditions comprise valid everyday content sentences and speech-like stimuli that are physically comparable to speech but do not contain any semantic content or a syntactic structure. Both conditions have been presented to participants in a clear version and overlaid with speech-simulating noise at -6 dB SNR. The effective connectivity, expressed as psychophysiological interactions, is based on ROI-to-ROI analyses, with ROIs being determined by the automated anatomical labeling (AAL) template. To decrease the spatial extent of the ROIs and allow for a differentiation between sub-parts of anatomical areas, the ROIs have been subdivided into two parts along the axis of the largest spatial extent, resulting in an anterior and a posterior, an inferior

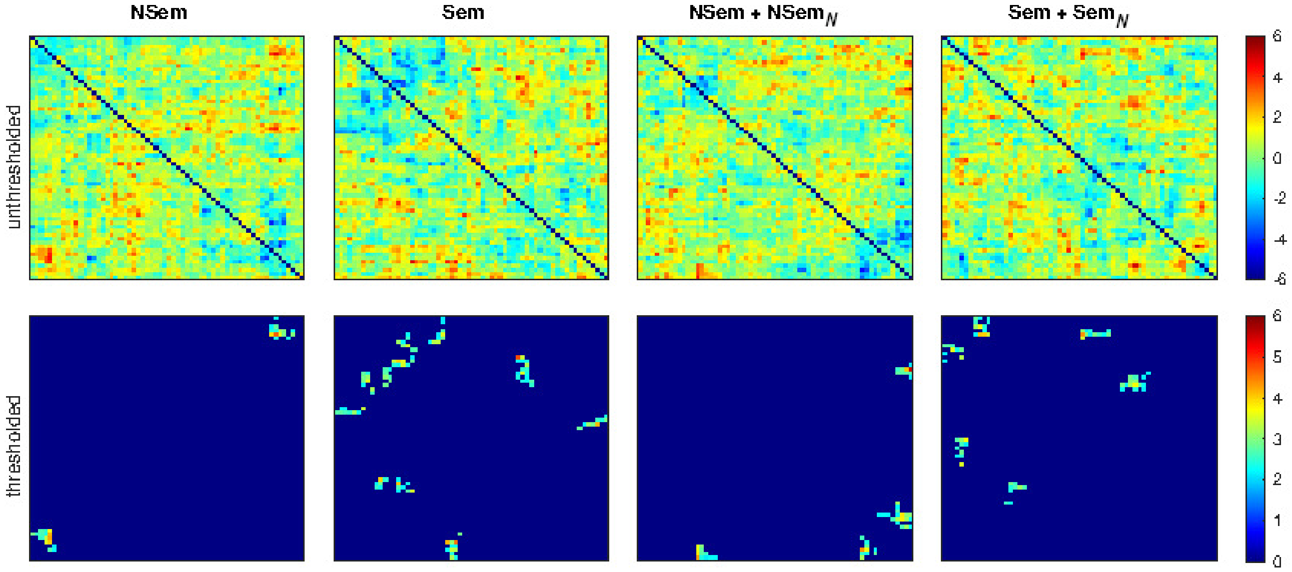

Figure 4.1.: gPPI connectivity matrices for the conditions semantic speech (Sem) and non-semantic speech ($NSem$) as well as semantic speech and semantic speech in noise (Sem_N) (weighted 1 and 0.5, respectively) and non-semantic speech with non-semantic speech in noise ($Nsem_N$) (weighted 1 and 0.5, respectively). The upper panels represent the F-values obtained from the group level F-test and the lower panels depict the absolute values of the supra-threshold ($p<0.05$, uncorrected) links between ROIs, that have a minimum of n spatially connected neighbors (24-neighbors criterion). Correction for multiple comparison was subsequently done using the network-based statistics. Note that the order of ROIs has been determined by temporal correlation of ROI time-series for the respective conditions and is not comparable across conditions.

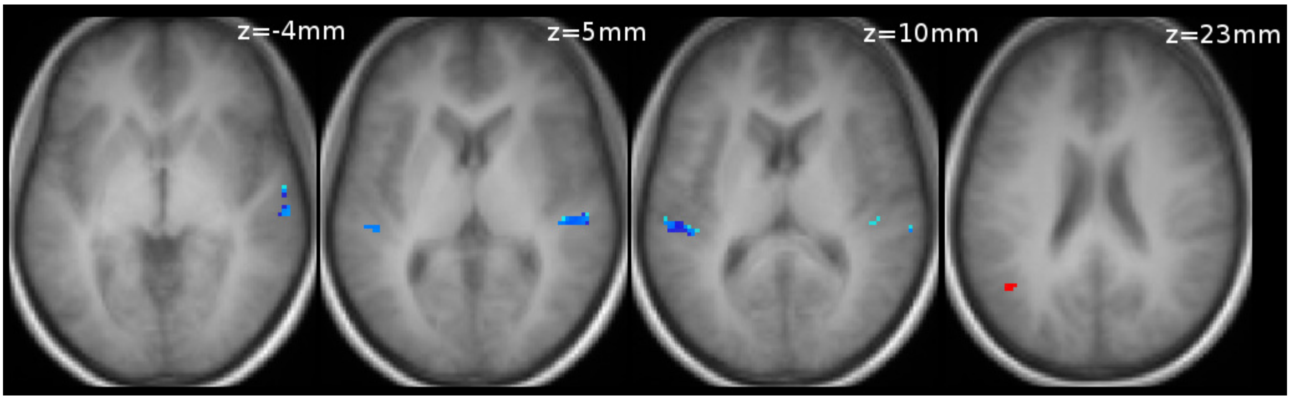

Figure 4.2.: Result maps obtained from univariate GLM, $Sem > NSem$ red, $Nsem > Sem$ blue, $Sem + Sem_N > NSem + NSem_N$ magenta, $NSem + NSem_N > Sem + Sem_N$ cyan. Activation could be identified in the Heschl Gyrus and in the posterior temporal lobe in Wernicke's area.

and a superior or a lateral and a medial part, respectively.

4.4.1. General linear model contrasts

In this section, the GLM results are presented to identify regions that show different activations for semantic and non-semantic speech and regions that show different activations for clear and noisy stimuli.

To shorten the names of contrasts that have been investigated, the following abbreviations are introduced: Sem (semantic speech), Sem_N (semantic speech in noise), $NSem$ (non-semantic speech) and $NSem_N$ (non-semantic speech in noise). For the identification of cortical areas that show different activations for semantic vs. non-semantic speech we utilized four contrasts, namely Sem vs. $NSem$ in both directions and Sem Sem_N vs. $NSem$ $NSem_N$ in both directions.

In Fig. 4.2, the corresponding activation maps are depicted for an FWE-corrected p-value threshold of $p < 0.05$. For the contrast $Sem > NSem$ (red), significantly higher activation in the left posterior superior temporal gyrus (STG) could be identified, while no significant activation could be found for the contrast $Sem + Sem_N > NSem + NSem_N$. Both contrasts, $NSem > Sem$ (blue) and $NSem + NSem_N > Sem + Sem_N$ (cyan), show similar activation patterns with higher activation in the left middle STG and bilateral Heschl's Gyrus (HG). For the contrasts $Sem_N > NSem_N$, no supra-threshold activations were found after FWE correction.

To show trends in the data and to determine regions to be investigated with the gPPI analyses, the uncorrected maps for the contrasts $Sem > NSem$ (red), $NSem > Sem$ (blue), $Sem_N > NSem_N$ (magenta), and $NSem_N > Sem_N$ (cyan) are depicted in Fig. 4.3. For $Sem > NSem$, increased activation is located from the left anterior to the posterior STG, the left posterior middle frontal gyrus (MFG) and the left anterior superior frontal gyrus (SFG) as well as in the left paracingulate sulcus (PCGS) and in the right central sulcus. For $Sem_N > NSem_N$ there is only increased activation in the left lateral fissure. In the auditory cortex areas, bilaterally in the anterior STG and HG both $NSem > Sem$ and $NSem_N > Sem_N$ show increased activation. For the contrast $NSem_N > Sem_N$ additional regions with increased activation are located in the left inferior, postcentral gyrus (PoG), the right parietal gyrus and the right SFG. The sagittal view of the left hemisphere (x=-50 mm) shows the increased activation for the non-semantic condition as compared to the semantic conditions in the left auditory cortex, with increased activation for semantic speech as compared to non-semantic speech

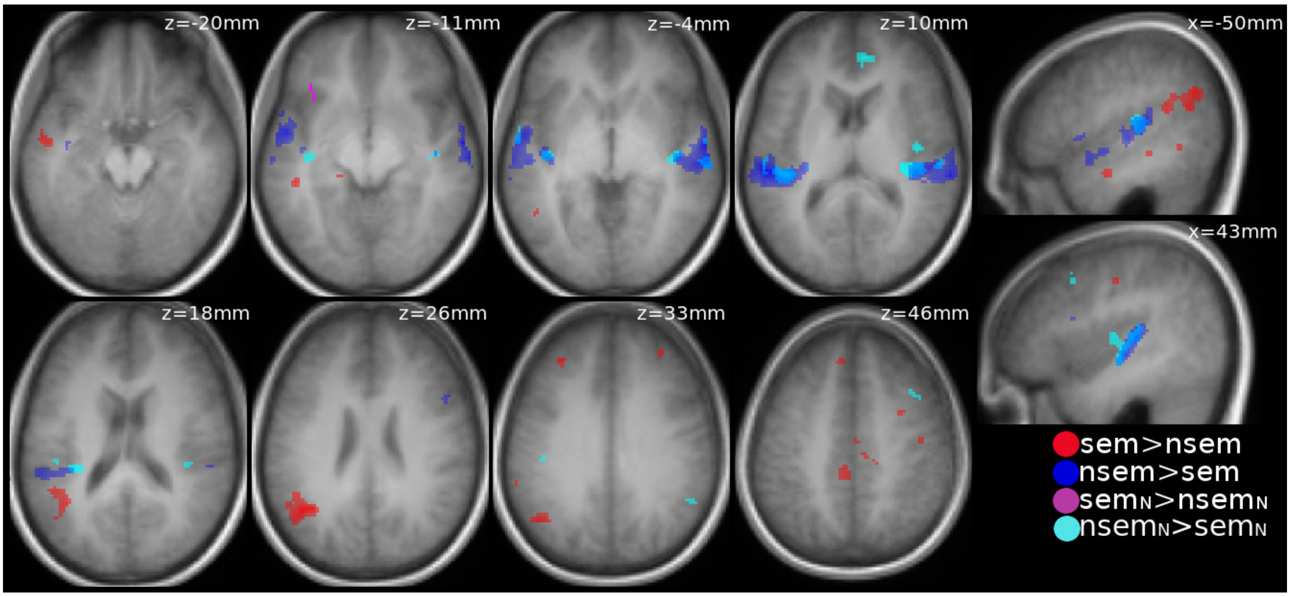

Figure 4.3.: Trends for the differences between semantic and non-semantic speech. The activation map for $Sem > NSem$ (red), $NSem > Sem$ (blue), $Sem_{\mathrm{N}} > NSem_{\mathrm{N}}$ (magenta) and $NSem_{\mathrm{N}} > Sem_{\mathrm{N}}$ (cyan) are shown without correction for multiple comparison to present trends in the data.

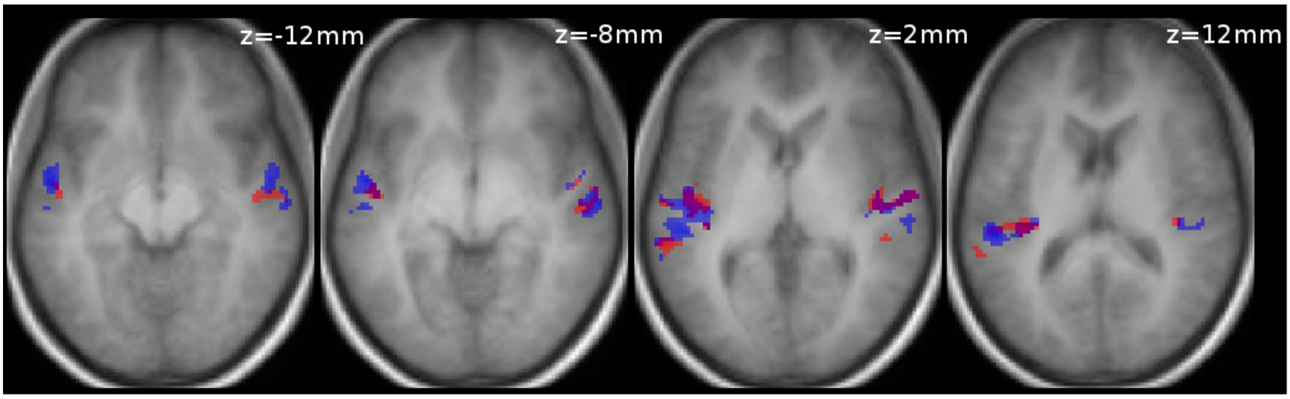

Figure 4.4.: Effect of clarity: Speech>Speech in Noise (red) and ISTS>ISTS in noise (blue)

areas that are located around the auditory cortex.

The result maps for the contrasts $Sem > Sem_N$ (red) and $NSem > NSem_N$ (blue) are depicted in Fig. 4.4. For both contrasts, the clear conditions lead to increased activation across the STG, the HG and the lateral fissure (LF) bilaterally for the clear versions as compared to the noisy versions. However, for the semantic speech, this effect can be rather observed in the anterior and posterior end of the STG, while for the non-semantic conditions the effect can be observed in a connected cluster along the STG. For the inverse direction, noisy stimuli > clear stimuli, no effect can be found for the semantic speech version and a small cluster in the inferior frontal gyrus (x=-28 mm, y=24mm, z=36mm) for $NSem_N > NSem$.

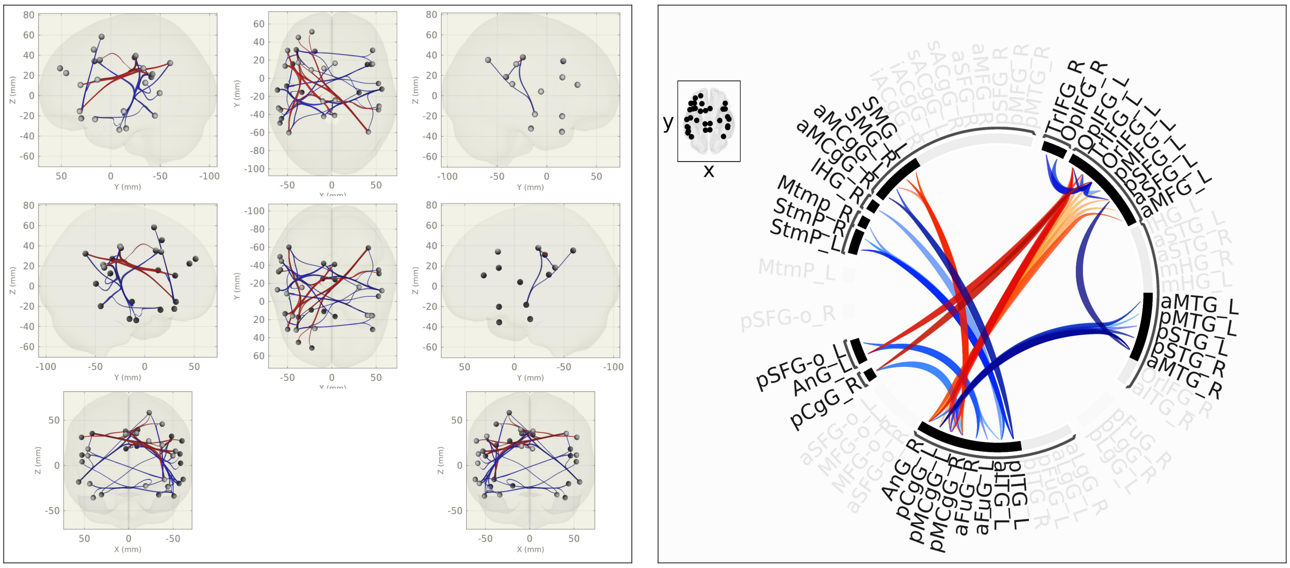

Figure 4.5.: Contrast of gPPI: $Sem(1)$ and Sem_N (0.5) vs. $NSem$ (-1) and $NSem_N$(-0.5). Orange and red colors indicate a stronger connection during the semantic speech conditions, blue colors indicate a stronger connectivity during non-semantic conditions. Darker colors depict stronger connections in both cases.

4.4.2. ROI-to-ROI analyses

In this section, the results obtained from ROI-to-ROI analyses are presented for the contrast between semantic speech conditions and non-semantic speech conditions, as well as for both classes, semantic speech and non-semantic speech in separate analyses. To account for the different extent of perceivable speech, both semantic or non-semantic, the clear versions are weighted with 1 while the noisy versions are weighted with 0.5. The contrast between speech conditions (semantic speech (1), semantic speech in noise (0.5)) and non-semantic speech conditions (non-semantic speech (-1) and non-semantic speech in noise (-0.5)) is presented in Fig. 4.5. Red colors depict connections that are significantly stronger for semantic speech conditions than for non-semantic conditions, while blue colors depict the opposite effect. Connections between the posterior part of the cingulate gyrus (CgG) bilaterally and the left inferior frontal gyrus (IFG) are stronger for semantic speech conditions, while non-semantic stimuli lead to higher interconnection effects within the left auditory cortex and between the left auditory cortex and the CgG. The presentation of the connection on the glass brain shows that for both directions effects can be found primary in the left hemisphere. While connections that are stronger for the semantic speech conditions are located rather in a transversal plane, connections that are stronger for the non-semantic conditions are spread in all directions.

The gPPI ROI-to-ROI effects obtained for the semantic speech (clear weighted with 1, and in noise weighted with 0.5), depicted in Fig. 4.6, indicate three clusters of connections. All connection-values are positive, meaning that only positive regression slopes were found for the influence between distant ROIs on another in the GLM. Regions in the left IFG and the right posterior inferior temporal gyrus (ITG) as well as the right angular gyrus (AnG) are interconnected, while an “orthogonal” cluster of connections between the right IFG/MFG region and the left AnG can be identified. A third cluster shows a more symmetrical pattern of inter-hemispheric connections, including cingulate gyrus regions being connected to the bilateral posterior superior temporal gyrus (pSTG), supramarginal gyrus (SMG) and the orbital part of the right SFG.

Fig. 4.7 depicts clusters of significant connections identified for the non-semantic speech conditions (clear weighted with 1, noisy weighted with 0.5). One cluster of positive connections could be identi-

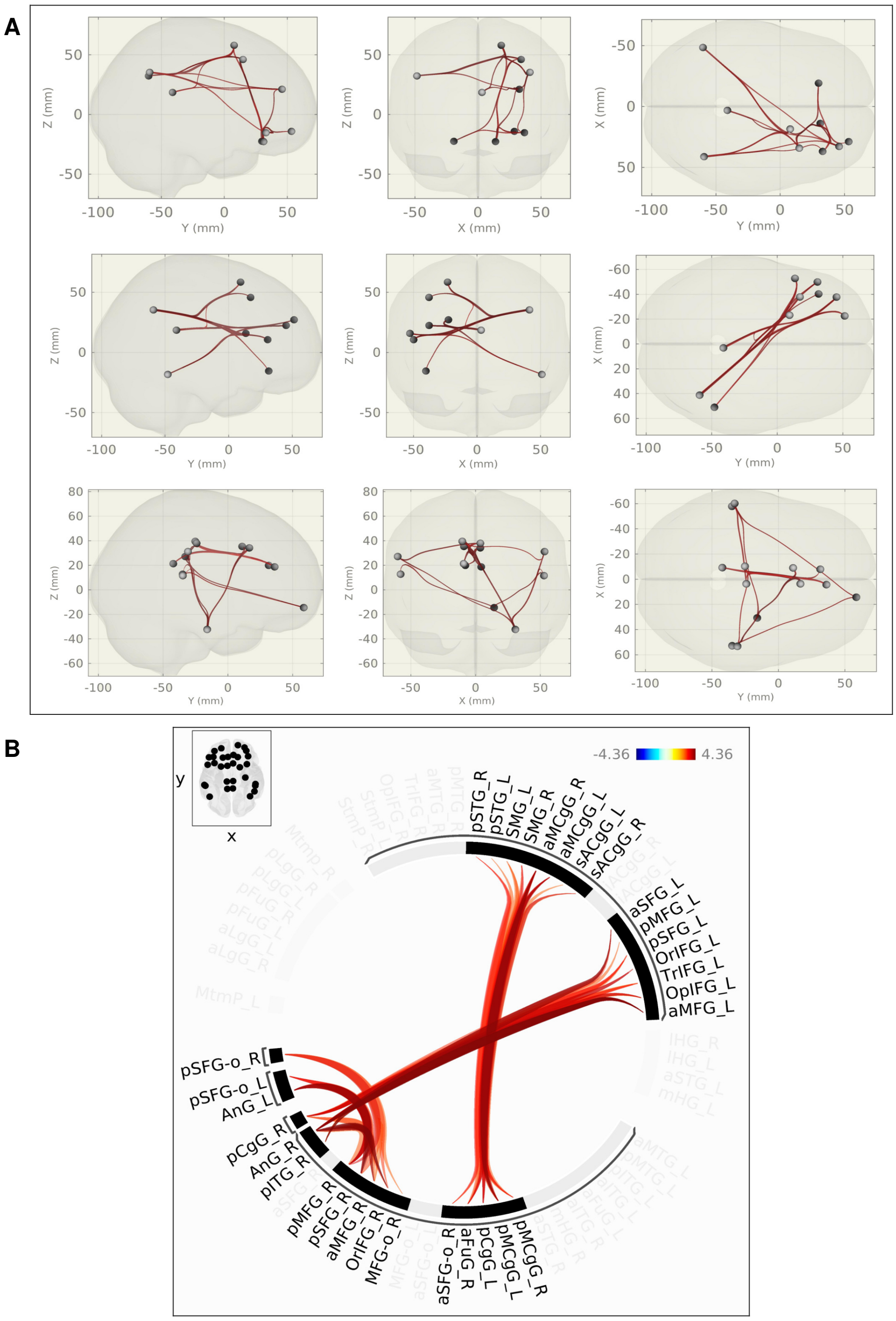

Figure 4.6.: Generalized PPIs for the conditions speech (weighted 1) and speech in noise (weighted 0.5), including temporal cortex regions, angular gyrus, supramarginal gyrus, fusiform gyrus, cingulate cortex, and inferior, middle and superior frontal gyrus. A: Spatial distribution of the three observed clusters (one per row) obtained from the gPPI analysis. B: Graph of significant connections. Light red lines depict low, positive values as compared to dark lines depicting large, positive values. Positive values represent positive slopes for the regression slope obtained from the GLM that includes the respective ROI time series as regressors.

fied, representing positive regression slopes obtained from the GLM with BOLD-time series being used as regressors. Negative connections, respectively, indicate negative regression slopes. The regions in the right frontal lobe (IFG/SFG and MFG) are negatively interconnected and connected to the right posterior MTG and bilateral SMG as well as to the left auditory cortex (STG and lHG). Positive connections can be found bilaterally between the SMG with the lingual gyrus (LgG), the FuG and the right posterior ITG, as well as between the middle part of the CgG and the LgG.

The effective connections, including their directions, are summarized in Fig. 4.8 and show that the processing of semantic speech and non-semantic speech involve three partly overlapping networks, with the cingulate gyri and the angular gyri bilaterally playing an intersection role in these networks. While these structures only get input from the temporal lobes (with some left lateralization) for non-semantic speech, they are largely interconnected with the left inferior frontal gyrus as well as with the middle and superior frontal gyrus for semantic speech. The supramarginal gyrus, however, communicates bidirectionally with the temporal lobe (pITGr, FuG) and with the IFG for non-semantic speech. For semantic speech, there are only bidirectional connections to the posterior middle cingulate gyrus. A third, inhibitory network for non-semantic speech processing is observable between the left temporal lobe, the right frontal lobe and the anterior cingulate gyrus, bilaterally.

4.5. Discussion

In the previous sections the results obtained from the analysis of differences between the BOLD responses to the perception of meaningful sentences (semantic speech) and a natural speech stream without any semantic meaning or syntactic structure (non-semantic speech) in clear and noisy versions have been presented. Based on the activation maps and the psychophysiological interaction graphs that have been depicted, the previous neural network models for the processing of meaningful spoken language can be confirmed. The stimuli have not been controlled for differences between semantic meaning or syntactic structure. Both classes of stimuli, semantic and non-semantic speech, contain naturally spoken language recordings but differ in their content of interpretable abstract features.

4.5.1. Extraction of meaning from semantic speech

For the processing of syntactically correct and semantically meaningful spoken language — named semantic speech in this study — a processing network could be identified, which includes areas that have been associated with a more complex and abstract task than acoustical processing. This network involves left hemispheric frontal lobe structures, the middle to posterior cingulate gyrus (CgG) and structures of the temporo-parietal junctions, namely the the supramarginal gyrus (SMG) and the angular gyrus (AnG). The SMG in both hemispheres is bilaterally and bidirectionally connected to the posterior to middle CgG. This sub-network can be observed in the present analysis both for the contrast semantic speech and semantic speech in noise > non-semantic speech and non-semantic speech in noise and for the condition analysis semantic speech and semantic speech in noise vs. baseline. The posterior CgG has been described to be involved in the processing of sentences (plausible vs. implausible sentences (Mashal et al., 2009), high predictable sentences (Obleser et al., 2007b), or to the narrative language comprehension (Smirnov et al., 2014)) and also the semantic processing of words (pseudo-words>words, (Newman and Twieg, 2001)). The SMG has also been associated with sentence processing (semantically challenging context (Obleser and Kotz, 2010), sentence comprehension (Richardson and Price, 2009; Raettig et al., 2010; Price, 2010) and cross-modal intelligibility (Akama et al., 2012)). Some studies associated this area in particular with the syntactic processing (syntactic errors, grammatical errors (Raettig et al., 2010), syntactically complex sentences (Friederici et al., 2003), sentences where meaning depends on the order of words (Richardson and Price, 2009),

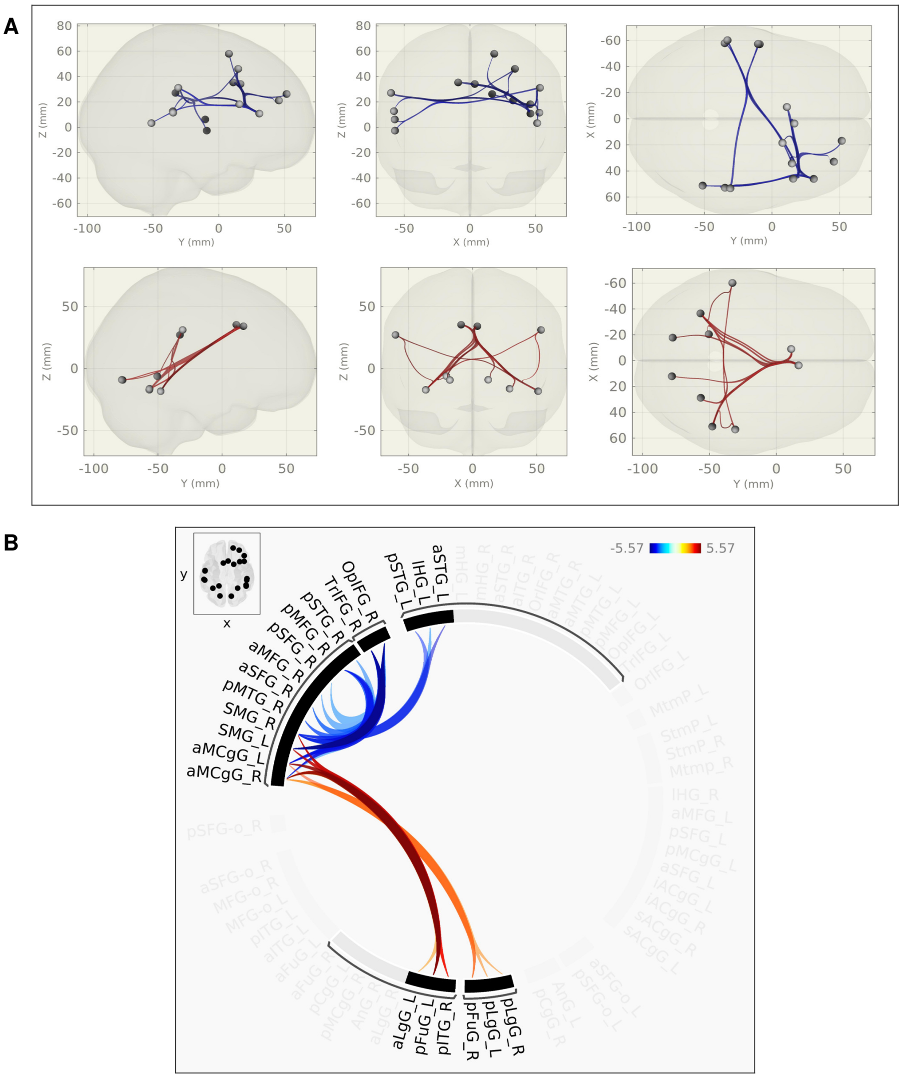

Figure 4.7.: Generalized PPIs for conditions $NSem$ (weighted 1) and $NSem_N$ (weighted 0.5), including temporal cortex regions, angular gyrus, supramarginal gyrus, fusiform gyrus, cingulate cortex, and inferior, middle and superior frontal gyrus. A: Spatial distribution of the three observed clusters (one per row) obtained from the gPPI analysis. B: Graph of significant connections obtained from the gPPI analyses. Light red lines depict low, positive values as compared to dark lines depicting large, positive values. Blue lines represent negative values, respectively. Positive values represent positive slopes for the regression slope obtained from the GLM that includes the respective ROI time series as regressors, while negative values represent negative regression slopes.

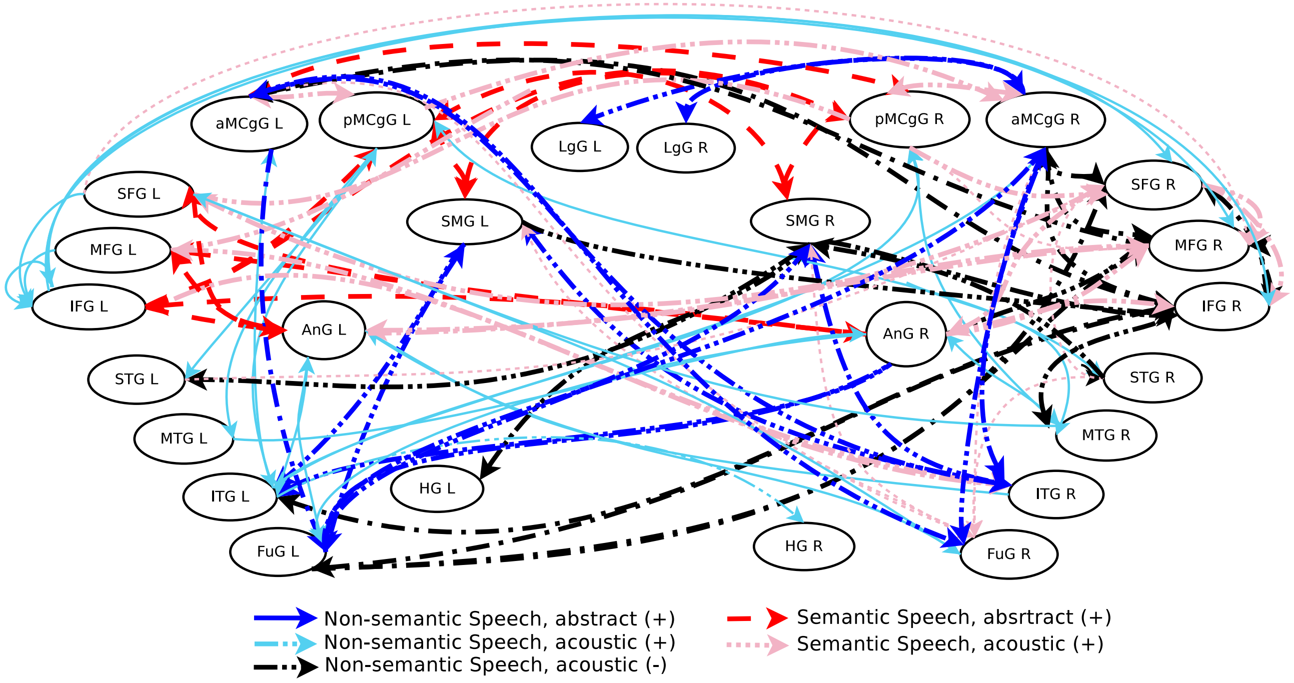

Figure 4.8.: Network for the processing of spoken language. Red arrows depict connections that are present during the processing of semantic speech and that are stronger for semantic speech than for non-semantic speech and which are therefore interpreted to be specific for valid spoken language processing. Pink arrows depict conections that can be identified for semantic speech and semantic speech in noise, but which are not significantly stronger than for the processing of non-semantic speech. They reflect spoken language processing on an acoustic level. Dark blue arrows represent connections that are present for the processing of non-semantic speech and not for the processing of semantic speech, while black arrows depict inhibitory connections for the same task. These connections seem to be part of an error detection network. Light blue arrows represent those connections, which are interpreted as part of a network to process the acoustic characterictics of spoken language and which are increased for increased task demands, when spoken language cannot be understood.

syntactically correct but semantically incorrect sentences (Hickok, 2012), or word order constraints (Price, 2010)). The interaction between the SMG and the posterior cingulate gyrus might therefore reflect the transition from the syntactic to the semantic analysis of spoken language.

Connections from the left inferior frontal gyrus (IFG), superior frontal gyrus (SFG) and the middle frontal gyrus (MFG) to the bilateral middle to posterior cingulate gyrus and the angular gyrus (AnG) bilaterally are found for the contrast semantic speech > non-semantic speech and for the condition semantic speech and semantic speech in noise. They seem to be involved in the semantic processing of speech. For non-semantic speech > semantic speech, the posterior cingulate gyrus and the AnG are bidirectionally connected to the anterior middle temporal gyrus (MTG) and the posterior superior temporal gyrus (STG) bilaterally. The left AnG is connected in both directions with the left inferior temporal gyrus (ITG) and the left fusiform gyrus (FuG). It must be noted that the connections described for the analysis of the conditions semantic speech vs. non-semantic speech are obtained without correction for multiple comparison. Connections that are stronger for non-semantic speech than for semantic speech only show a trend. They might be on average slightly stronger for non-semantic speech as compared to semantic speech due to increased task demands, but might be not specific for stimuli that contain acoustic features of natural speech but no valid meaning. However, those results that describe connections that are stronger for semantic speech than for non-semantic speech, can also be identified for the investigation of the conditions semantic speech and semantic speech in noise (with FDR-correction). This can be interpreted as an effect that is specific for the interpretation of meaningful spoken language.

These results imply a coordination function of the AnG and the posterior CgG. Acoustic stimuli that are identified to contain speech are processed in the AnG and the pCgG. In a pre-analysis, e.g., a syntactic structure analysis, the activation in the frontal gyrus, in particular in the left IFG, might increase for the analysis regarding interpretable content. However, if no language structure beyond the acoustic features can be identified, the temporal lobe structures closely located to the auditory cortex might be increased in activation via top-down processes due to the apparently increased task demand. The AnG has been associated with both, sentence processing and word processing. It has shown increased activation for plausible vs. implausible sentence (Sharp et al., 2009), and for non-degraded speech as compared to degraded speech with respect to the processing of the sentence meaning (Clos et al., 2014). Its activation also showed correlation with the predictability of sentences (Obleser and Kotz, 2010). In studies by Golestani et al. (2009) and Sharp et al. (2009) a relation between activation in the AnG and the semantic relation between words has been described. From these observations, an integration function from single word interpretation to meaning can be concluded, that requires neuronal communication with the left hemispheric frontal cortex.

4.5.2. Lateralization of spoken language processing

The left IFG, in particular the opercular part and the orbital part, have been reported to be activated during the processing of plausibility of sentences (Ye and Zhou, 2009), the ambiguity of sentences (Rodd et al., 2005), differences between semantic and syntactic processing of sentences (Schafer and Constable, 2009), or the predictability of sentences (Obleser and Kotz, 2010). Gonzalez-Castillo and Talavage (2011) reported a rather bilateral network that is involved when listening to sentences. The lateralization of the neural activation during the processing of spoken language on a word level is also a matter of debate. Sharp et al. (2009) observed a dominance of left hemisphere structures like IFG, MFG and SFG during a semantic relation task for words, and the comparison of words and pseudowords was discovered to be represented in a left lateralized network (Kotz et al., 2009). On the other hand, Newman and Twieg (2001) found considerable involvement of right hemispheric structures for the processing of pseudo-words as compared to words. Also Bilenko et al. (2009) discovered a rather bilateral network of structures that are involved in resolving ambiguity. Some studies found evidence for a lexico-semantic processing in the MFG and SFG (Sharp et al., 2009) and word-production in the MFG (Whitney et al., 2009; Jeon et al., 2009; Heim et al., 2009). In these studies language processing was investigated on a word-level and not on sentence level. All identified a rather left lateralized network for semantic processing. Mashal et al. (2009) and colleagues discuss in their work that the inconsistencies of lateralization effects could possibly be explained by task differences.

The present data support a model presented in Peelle (2012) that assumes an increasing left lateralization with increasing task complexity. While acoustic features like amplitude modulations and a part of the phonetic analysis are located in bilateral auditory cortex structures, the processing of words and sentences requires cognitive resources outside of the auditory cortex, predominantly in the left hemisphere. This is consistent with the present observation of strong connections between the left frontal cortex with the AnG and the posterior CgG for semantic speech conditions against baseline. The effective connectivity for probably more basic speech feature stages can be identified between bilateral frontal cortex structures and bilateral temporal structures (bilateral STG, right FuG and right ITG). These connections could be identified in the network for semantic speech and semantic speech in noise, but not significantly stronger for the semantic speech conditions than for the non-semantic speech conditions. The activation maps in Fig. 4.2 to Fig. 4.4 are in line with this assumption. Acoustical differences, like those between clear and noisy stimuli, are reflected in activation differences in the auditory cortex. The presentation with non-semantic speech that might demand more resources in the auditory cortices to increase putative decreased signal quality, leads to higher activation in the auditory cortex. Simultaneously, the automated processing of normal spoken language processing is reflected in increased activation in the left AnG, the bilateral frontal cortices and the CgG. Bilateral connections can also be found for semantic speech and semantic speech in noise between the anterior and posterior parts of the CgG as well as inter-hemispheric connections between these structures.

The anterior CgG has been associated with both speech production (Ali et al., 2010; Haupt et al., 2009; Birn et al., 2010) and speech perception on a rather acoustic level, e.g., increased activation for masked speech as compared to clear speech (Evans and Davis, 2015) or the emotional prosody of sentences (Wittfoth et al., 2010), while the posterior CgG has been associated to sentence processing, as described before. Since the anterior and posterior part of the CgG are effectively connected to temporal lobe structures for both semantic and non-semantic speech processing, an integration of information from the anterior to the posterior part of the CgG with top-down connections leading back to the auditory cortex is concluded.

4.5.3. Processing of invalid spoken language

Another strongly lateralized effect that can be observed for the processing of non-semantic speech is the inhibitory effective connectivity of the right frontal cortex to the left temporal cortex and bilaterally to the anterior CgG. This network appears to represent a network orthogonal to the acoustic processing of valid spoken language, with connections between the right inferior temporal lobe and the left inferior and middle frontal cortex. The negative connection values might reflect inhibitory connection effects between the temporal lobe and the right frontal cortex for incongruent speech as it has been described by Dick et al. (2009) to be reflected in right inferior frontal lobe activations.

Positive connections for the conditions non-semantic speech and non-semantic speech in noise can be found bilaterally between the inferior temporal lobe (ITG and FuG) with the bilateral SMG and the anterior CgG. Studies which cover the cortical activation during the processing of speech have found increased activation in the ITG for plausible vs. implausible sentences (Price, 2010) or normal vs. anomalous sentences (Kuperberg et al., 2000), for cross-modal intelligibility (Akama et al., 2012), ambiguity of sentences (Rodd et al., 2005) or speech comprehension in general (Awad et al., 2007). In the FuG increased activation has been identified for words compared to pseudo-words (Davis and

Gaskell, 2009), or spoken sentence processing with the attention directed to the verbal content (von Kriegstein et al., 2003). The SMG and the anterior CgG are both inhibitorily influenced by the right inferior frontal cortex for non-semantic speech processing. They might be nodes in a feedback loop that contribute to the identification of structures that characterize speech. Depending on the similarity of the perceived stimulus to spoken language on an acoustic and semantic or syntactic level, they might coordinate the activation and cognitive load in the auditory cortex.

4.6. Conclusion

In this study we compared the neuronal processing of natural spoken sentences to the processing of stimuli that are acoustically similar to speech but do not have any meaningful content on the activation level and on the level of effective connectivity between spatially distant cortical structures. We found that some effects are not reflected in the absolute activation level, but rather in condition specific connectivity patterns between different cortical structures. This includes a network for the processing of natural spoken language that is bilaterally distributed for the more basic, acoustic processing steps with increasing left lateralization for increasing abstractness of the task. The cingulate gyrus and the bilateral angular gyri and supramarginal gyri seem to play a crucial role in the integration of information to extract meaning from spoken language. For the processing of invalid spoken language, the top-down connections to the auditory cortex and the inferior temporal lobe are more pronounced than during the processing of valid sentences, and a control feedback loop could be identified that controls for the validity of the perceived stimuli. The identified network is very complex and further investigations are required to verify the different processes that have been described. However, the results are consistent in their structure and are in line with results obtained by previous studies. Simultaneously, they build a quite complex framework that summarizes different aspects of the extraction of meaning from spoken language.

5. General Discussion

5.1. Analysis methods

In the previous chapters the results from an auditory fMRI study investigating the cortical activation corresponding to spoken language processing, in particular on the semantic level, have been presented and discussed. In addition to the presentation of physiological results, different analysis methods have been applied to the data set and their applicability and results have been compared to each other as well as to results obtained from previous studies.

The stimulus set used for the fMRI study was composed of six different acoustic conditions and one silence condition as baseline. The conditions of the highest interest are the semantic speech condition and the non-semantic speech condition. Stimuli from both conditions are acoustically similar to normal speech. However, while the semantic speech condition is composed of valid sentences, the stimuli from the non-semantic speech condition do not hold any meaningful semantic content. Both conditions have been overlaid with speech-simulating noise to create two additional conditions. For the comparison with non-speech stimuli with comparable spectral characteristics, the subjects have been additionally presented with the speech-simulating noise itself and with noise vocoded speech, derived from the filtering of white noise with the LPC coefficients of the material from the semantic speech condition. The analysis of the BOLD responses corresponding to these stimuli has been realized with different methods that are summarized, compared and discussed in the following.

In Chapter 2, the Searchlight Classification Informative Region Mixture Model (SCIM) method, a novel method for the evaluation of classification performance results obtained from the searchlight classification with a support vector machine, has been presented. The approach is based on the assumption that the distribution of classification performance values obtained from a whole brain searchlight classification can be decomposed into informative searchlights, which reflect differences between BOLD patterns corresponding to two contrasted conditions in high classification performance values, and, respectively, non-informative searchlights. The non-informative searchlight distribution can be interpreted as a Null-distribution which has to be separated from those searchlights, whose BOLD patterns contain information about the investigated conditions. This separation is realized with a Gaussian mixture model (GMM), which is fitted to the joint distribution of all classification performance results from a whole brain classification analysis. From the converged model, an a-posteriori probability for each searchlight can be extracted that reflects the probability of the searchlight's classification performance to be drawn from the Null-distribution or non-informative distribution. The SCIM method has two major advantages: 1) it does not assume a specific characteristic of the Null-distribution, as it is the case for the binomial test, and 2) since the Null-distribution is represented by the junction of non-informative searchlights in the brain, the numerically expensive and time-consuming repetition of the analysis as done in the random permutation test is not necessary. The differences of the results obtained from the analysis of the contrast semantic speech vs. non-semantic speech from the fMRI dataset and from the analysis of a simulation data set with the SCIM method and with the reference methods can be summarized as follows: Compared to the reference methods, in particular the binomial test, the results obtained with the SCIM method are found to be highly robust with respect to changing p-value thresholds. This high dependence on those thresholds is more pronounced for the binomial test than for the random permutation test. However, the result maps obtained from the random permutation test also show informative regions that are not in line with findings from previous research on this topic.

The spatial smoothing of classification performance results was found to have a beneficial impact on the quality of result maps. In combination with the SCIM method, the sensitivity was increased due to a better separability of the informative and non-informative searchlight distribution. For the reference methods, spatial smoothing increases the specificity of the analysis, which is reflected by fewer informative regions in areas which are physiologically non-plausible and across the different methods inconsistent.

The evaluation of the classification performance with the AUC has some advantages over the accuracy measure. It benefits from an increased sensitivity with a decreased standard error (Green and Swets, 1966; Bradley, 1997; Spackman, 1989), which allows for a dependable regularization and prevents over-fitting in the training phase. The decreased standard error of the AUC measure, as compared to the accuracy measure, facilitates the decomposition of the informative and non-informative searchlight distribution and supports the robustness of the SCIM method. The SCIM method, with an AUC measure as performance quantification and spatial smoothing of resulting AUC maps, provides an evaluation tool that reliably identifies informative regions based on differences of local BOLD patterns.

The results obtained from the analysis of the full data set with the SCIM method have been presented in Chapter 3. These include - in addition to the contrast semantic speech vs. non-semantic speech - the results of the analysis of twelve further contrasts on single subject level and group level. The different contrasts between the conditions are very similar, even though they reflect the discriminability of local BOLD patterns corresponding to acoustic stimuli that vary on different speech features, respectively. The quantitative analysis of single subject result maps, as inferred by the portion of informative voxels in anatomical regions determined by the automated anatomical labels (AAL, Tzourio-Mazoyer et al., 2002) atlas, however, showed some systematic differences in the superior temporal gyrus (STG), Heschl's gyrus (HG) and the insulae (INS) between the contrasts, that can be assorted to groups of contrasts.

The six acoustic conditions differed in their content of semantic structure, human voice pitch and temporal modulation. In a subsequent analysis, the thirteen contrasts have, therefore, been sorted due to the degree of differences between the contrasted conditions with respect to these three speech features. With a Spearman rank correlation (Spearman, 1904) the voxelwise correlation between the rank of feature difference and the searchlight classification performance has been evaluated to create correlation-based maps, equivalent to the information-based maps (Kriegeskorte et al., 2006) obtained from the searchlight classification itself. Similar to the decomposition of informative searchlights from non-informative searchlights, the distribution of correlated searchlights has been separated from the non-correlated searchlights and anti-correlated searchlights, which allows for the identification of those regions whose discriminability of activation patterns correlates with the difference of specific speech features. For the correlation analysis, a three-component GMM has shown to be advantageous in comparison to a two-component GMM, which leads to an additional distribution of anti-correlated regions.

5.2. Cortical structures and neural networks

Differences in the depth or presence of temporal modulation showed the largest correlation effects in the temporal lobe, in particular in the primary auditory cortex, which has previously been associated with temporal modulations in the frequency range of the speech rate (4Hz, Giraud, 2004). The superior temporal lobe has also been associated with the processing of speech envelopes (Kubanek et al., 2013) and speech rhythm (Geiser et al., 2008; Ischebeck et al., 2008). The activation corresponding to speech rhythm has been identified in the rolandic operculum (RO, Ischebeck et al., 2008; Heinrich et al., 2008; Geiser et al., 2008) and the temporo-parietal junction (Hervé et al., 2012; Geiser et al., 2008). Those regions also showed correlation effects with modulation differences in the present study.

Those regions which show correlation effects with voice pitch differences are located within the lateral HG, the temporal pole, the superior and middle temporal gyri (STG and MTG) and in the left inferior frontal lobe (IFG). Activation in the HG has been associated with voice pitch perception on different stages. The perception of vowels in contrast to non-speech sounds has found to induce activation in the HG (Uppenkamp et al., 2006), but also interpretational tasks like the identification of a speaker's sex (Lattner et al., 2005) or identity (Formisano et al., 2008). Voice selective regions have also been identified along the superior temporal lobe (Moerel et al., 2012; Pernet et al., 2015; Liebenthal et al., 2005) and in the MTG (von Kriegstein et al., 2003; Belin et al., 2000; Warren et al., 2006). Activation in the frontal lobe has found to be associated with the processing of spectral details of speech stimuli (Gandour et al., 2002; Meyer et al., 2004). The cortical structures, whose BOLD patterns were found to contain information about voice pitch difference again seem to be hierarchically structured. The initial recognition of speech is reflected in cortical activation in the auditory cortex, while the interpretation of voice pitch information is reflected in both, frontal lobe areas and in the auditory cortex. Further insights to neural networks for the processing of voice pitch information might be obtained from connectivity analyses of data sets that are focused upon this stage of spoken language processing.

With the correlation analysis only small regions showing high correlation between classification performance values and stimulus feature differences could be identified for the semantic processing of spoken language. These are located in the bilateral inferior temporal gyrus (ITG), the left MTG, the inferior frontal gyrus (IFG), the precuneus (Pcun), the angular gyrus (AnG), the lingual gyrus (LgG) and the posterior cingulate gyrus (pCgG). While activation in the MTG has previously been found to be related to semantic processing on word level (Davis and Gaskell, 2009; Diaz and McCarthy, 2009), the ITG has also been found to be related to the more complex sentence processing (Rodd et al., 2005; Price, 2010). Activation in the inferior frontal lobe has previously been reported for both, the processing of words (Kotz et al., 2009; Nosarti et al., 2010) and the processing of sentences (Sharp et al., 2009; Tyler and Marslen-Wilson, 2007). This is also the case for the PCun, which was found to be involved in the identification of existing words (Davis and Gaskell, 2009) but also in the interpretation of spoken sentences (Mashal et al., 2009) and narrative spoken language processing in general (Whitney et al., 2009). Activation in the AnG has been particularly associated with the integration of words to a meaning (Golestani et al., 2009; Sharp et al., 2009), while the LgG has been identified to be involved in word categorization (Bilenko et al., 2009) and higher order sentence processing (Kuperberg et al., 2000; von Kriegstein et al., 2003). Negative correlation effects have been identified in the HG, the posterior STG, the RO and the SMG.

The cortical regions, whose involvement in the semantic processing has been identified with the correlation analysis, have also been repeatedly identified in previous studies on spoken language processing on a semantic level. These regions include the temporal lobe regions, frontal lobe regions, the temporo-parietal junction (SMG, AnG), the lingual gyri and the cingulate gyri. They have been involved in the effective connectivity gPPI analysis, which is presented in Chapter 4. The results obtained from the gPPI analysis indicate separate neural networks for the acoustic processing of spoken language, the interpretation of valid sentences and semantic content, and for the processing of natural speech without any meaningful content.

For the extraction of meaning from spoken language, a neural network has been identified which includes left frontal lobe structures, regions of the temporo-parietal junction, namely the AnG and the SMG, and the posterior CgG. While the posterior cingulate gyrus has been associated with a lexico-semantic processing on both, sentence level (Mashal et al., 2009; Obleser and Kotz, 2010; Smirnov et al., 2014) and word level (Newman et al., 2001), the SMG has found to be involved in sentence processing on a conceptual level (Obleser and Kotz, 2010; Richardson and Price, 2009; Price, 2010), which explicitly involves the processing of syntactic structure (Raettig et al., 2010; Richardson and Price, 2009; Hickok, 2012). We interpret the connections between the SMG and the posterior CgG, therefore, as part of a processing stage to integrate information from syntactic structure to semantic meaning. Activation in the AnG has been reported to be related to an interpretational level of spoken language processing to evaluate the plausibility (Sharp et al., 2009), the predictability (Obleser and Kotz, 2010) or the relatedness between words (Golestani et al., 2009; Sharp et al., 2009). The AnG and the pCgG have been found to be connected for non-semantic speech with bilateral temporal lobe structures, while they are connected with the left frontal lobe for semantic speech. With respect to their role identified in previous studies, we assume a coordinating function in spoken language processing for these areas.

For non-semantic speech a bilateral network between the inferior temporal lobe, the bilateral SMG and the anterior CgG could be identified, which might reflect feedback processes, induced by the increased task difficulty to extract meaning from a non-intelligible speech stimulus. The SMG and the anterior CgG, which has shown to be involved in both speech production (Ali et al., 2010; Haupt et al., 2009; Birn et al., 2010) and speech perception (Evans and Davis, 2015; Wittfoth et al., 2010), detect inconsistencies in the non-semantic speech and increase via top-down processing the activation in the inferior temporal lobe. An inhibitory network between the left inferior temporal lobe, the right frontal lobe and the left anterior CgG as well as the bilateral SMG is consistent with activation in the right IFG for incongruent speech (Dick et al., 2009) and higher activation during the perception of pseudo-words compared to words (Newman and Twieg, 2001). It might reflect a control network to economize cognitive resources, when stimuli are classified as unintelligible.

5.3. Lateralization of spoken language processing

The lateralization of spoken language processing was observed in the data from this study both in local activation patterns and in connection to spatially distinct patterns. Even though the correlated regions for semantic processing obtained from the correlation analysis in Chapter 3 are small, they show a clear left lateralization, while a lateralization for the processing of acoustic features is much less pronounced. The activation maps for the contrasts between semantic speech and non-semantic speech in Chapter 4 (cf. Fig. 4.2) also support the assumption that the acoustic processing is less lateralized than the abstract processing. The bilateral temporal lobes show increasing activation during the processing of non-semantic speech as compared to semantic speech, which might be driven by the increased task demand. Semantic speech processing, on the other hand, leads to increased activation in the left temporo-parietal lobe. The effective connectivity, presented in Chapter 4, also shows a lateralization effect, which seems to correlate with the abstractness of the task (cf. Fig. 4.8). The lateralization of cortical activation during spoken language processing is still a matter of debate. In particular, cortical regions that are associated with higher levels of task demands, are reported to show lateralization effects. The left IFG has found to be involved in sentence processing (Ye and Zhou, 2009; Rodd et al., 2005; Schafer and Constable, 2009), while Gonzalez-Castillo and Talavage (2011) found rather bilateral effects for sentence processing. Lateralization effects for language processing on a word level have been reported by Kotz et al. (2009) and Sharp et al. (2009), while Bilenko et al. (2009) reported rather bilateral activation during an ambiguity resolving task. These inconsistencies are explained by Mashal et al. (2009) to arise from differences in the experimental task. The data presented in this thesis support a model by Peelle (2012), in which an increasing lateralization with increasing task complexity is predicted. For local BOLD patterns and connectivity measures, bilateral effects for the acoustic processing of spoken language can be observed, while the activation corresponding to interpretation on an abstract level is strongly left lateralized.

5.4. Conclusion and outlook

This work demonstrates that different analysis approaches on the same data set can lead to remarkably different results. The different methods are based on different assumptions, which have to be controlled carefully. The multivariate analysis methods have shown to be more sensitive than univariate methods (Norman et al., 2006; Kriegeskorte et al., 2006).

The general linear model (GLM, Friston et al., 1995), on the other hand, has the big advantage that it can account for overlapping BOLD responses since it is easier to convolve the cognitive condition regressors with the hemodynamic response function than deconvolving the BOLD responses to separate the responses from each other. Additionally, different regressors, that might explain variance triggered by the cognitive demands to solve an attention task, can be included. This is not possible in classification analysis. In a sparse imaging design with passive listening, these problems are not very pronounced, however they should be considered for other studies with a smaller TR and with tasks to control the attention of participants. The SCIM method is therefore particularly beneficial for passive listening tasks in a sparse imaging design, which permits the creation of condition contrasts. However, future studies will certainly contribute to a more detailed evaluation of the method's advantages and disadvantages.

The correlation analysis has shown to facilitate the disentanglement of BOLD responses corresponding to different stimulus features. In this study, a Spearman rank correlation was calculated to quantify the relation between the separability of BOLD responses and the stimulus feature differences, since it was difficult to quantify the stimulus features directly. For future studies it might be more beneficial to create a data set which permits a parametric variation of the stimulus features across the conditions. It can be assumed that measurable differences increase the beneficial effect of this method.

The results obtained from this work have supplemented the current knowledge with respect to the neural correlates of spoken language processing. While many studies have investigated very specific aspects of this complex task, the data which have been presented here permit a rough disentanglement of acoustic from abstract processing stages for meaningful and nonsense speech stimuli. Since the inter-individual solving strategies for cognitive tasks increase with increasing complexity (Seghier and Price, 2016), it is very difficult to find robust group results with respect to very specific aspects of spoken language. The neural network models presented here can be interpreted as a framework which supports and merges current established models. However, the different parts of this framework require further investigation with studies that focus on one stage, respectively.

Bibliography

D. A. Abrams, S. Ryali, T. Chen, E. Balaban, D. J. Levitin, and V. Menon. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's and Geschwind's area. *Cerebral Cortex*, 23:1703–1714, 2012.

P. Adank and J. T. Devlin. On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, 49:1124–1132, 2010. doi: 10.1016/j.neuroimage.2009.07.032.

A. M. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm. Dynamics of neuronal firing correlation: modulation of "effective connectivity". *Journal of Neurophysiology*, 61(5):900–917, 1989. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.1989.61.5.900. URL `https://www.physiology.org/doi/10.1152/jn.1989.61.5.900`.

H. Akama, B. Murphy, L. Na, Y. Shimizu, and M. Poesio. Decoding semantics across fMRI sessions with different stimulus modalities: A practical MVPA study. *Frontiers in Neuroinformatics*, 6:1–10, 2012.

H. Akama, B. Murphy, M. M. Lei, and M. Poesio. Cross-participant modelling based on joint or disjoint feature selection: an fMRI conceptual decoding study. *Applied Informatics*, 1(1):1, Dec. 2014. ISSN 2196-0089. doi: 10.1186/2196-0089-1-1. URL `https://applied-informatics-j.springeropen.com/articles/10.1186/2196-0089-1-1`.

N. Ali, D. W. Green, F. Kherif, J. T. Devlin, and C. J. Price. The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience*, 22(10):2369–2386, 2010. ISSN 0898-929X, 1530-8898. doi: 10.1162/jocn.2009.21352. URL `http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21352`.

C. Allefeld and J.-D. Haynes. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage*, 89:345–357, 2014. doi: 10.1016/j.neuroimage.2013.11.043.

C. Allefeld, K. Görgen, and J.-D. Haynes. Valid population inference for information-based imaging: From the second-level t -test to prevalence inference. *NeuroImage*, 141:378–392, 2016. ISSN 10538119. doi: 10.1016/j.neuroimage.2016.07.040. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811916303470`.

T. Arai and S. Greenberg. Speech intelligibility in the presence of cross-channel spectral asynchrony. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, volume 2, pages 933–936, Seattle, WA, USA, 1998. IEEE. ISBN 978-0-7803-4428-0. doi: 10.1109/ICASSP.1998.675419. URL `http://ieeexplore.ieee.org/document/675419/`.

J. S. Arsenault and B. R. Buchsbaum. Distributed neural representations of phonological features during speech perception. *Journal of Neuroscience*, 35:634–642, 2015. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.2454-14.2015. URL `http://www.jneurosci.org/content/35/2/634`.

M. Awad, J. E. Warren, S. K. Scott, F. E. Turkheimer, and R. J. S. Wise. A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, 27(43):11455–11464, 2007. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5257-06.2007. URL `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5257-06.2007`.

A. Baumgaertner, C. Weiller, and C. Büchel. Event-related fMRI reveals cortical sites involved in contextual sentence integration. *NeuroImage*, 16(3):736–745, 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1134. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811902911341`.

P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike. Voice-selective areas in human auditory cortex. *Nature*, 403:4, 2000.

D. Ben Shalom and D. Poeppel. Functional anatomic models of language: Assembling the pieces. *The Neuroscientist*, 14(1):119–127, 2008. ISSN 1073-8584, 1089-4098. doi: 10.1177/1073858407305726. URL `http://journals.sagepub.com/doi/10.1177/1073858407305726`.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, Jan. 1995. ISSN 00359246. doi: 10.1111/j.2517-6161.1995.tb02031.x. URL `http://doi.wiley.com/10.1111/j.2517-6161.1995.tb02031.x`.

N. Y. Bilenko, C. M. Grindrod, E. B. Myers, and S. E. Blumstein. Neural correlates of semantic competition during processing of ambiguous words. *Journal of Cognitive Neuroscience*, 21(5):960–975, 2009. ISSN 0898-929X, 1530-8898. doi: 10.1162/jocn.2009.21073. URL `http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21073`.

J. Binder, J. Frost, T. Hammeke, P. Bellgowan, J. K. J.A. Springer, and E. Possing. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10:512–528, 2000a.

J. R. Binder, J. A. Frost, T. A. Hammeke, P. S. F. Bellgowan, J. K. J.A. Springer, and E. T. Possing. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 2000b.

J. R. Binder, R. H. Desai, W. W. Graves, and L. L. Conant. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19:2767–2796, 2009. doi: 10.1093/cercor/bhp055.

R. M. Birn, L. Kenworthy, L. Case, R. Caravella, T. B. Jones, P. A. Bandetti, and A. Martin. Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency. *NeuroImage*, 49:1099 – 1107, 2010.

H. Blank and M. H. Davis. Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology*, 14(11):e1002577, Nov. 2016. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002577. URL `https://dx.plos.org/10.1371/journal.pbio.1002577`.

S. Bode and J.-D. Haynes. Decoding sequential stages of task preparation in the human brain. *NeuroImage*, 45(2):606–613, Apr. 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.11.031. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811908012226`.

A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.

M. P. Broca. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de la Société Anatomique*, page 18, 1861.

K. H. Brodersen, C. Mathys, J. R. Chumbley, J. Daunizeau, C. S. Ong, J. M. Buhmann, and K. E. Stephan. Bayesian Mixed-Effects Inference on Classification Performance in Hierarchical Data Sets. *NeuroImage*, page 44, 2013. doi: 10.1016/j.neuroimage.2013.03.008.

H. Burianova, A. R. McIntosh, and C. L. Grady. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *NeuroImage*, 49(1):865–874, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.08.066. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909009781`.

M. Canini, P. A. Della Rosa, E. Catricalà, K. Strijkers, F. M. Branzi, A. Costa, and J. Abutalebi. Semantic interference and its control: A functional neuroimaging and connectivity study: Semantic interference and its control. *Human Brain Mapping*, 37(11):4179–4196, 2016. ISSN 10659471. doi: 10.1002/hbm.23304. URL `http://doi.wiley.com/10.1002/hbm.23304`.

J. D. Carlin, J. B. Rowe, N. Kriegeskorte, R. Thompson, and A. J. Calder. Direction-Sensitive Codes for Observed Head Turns in Human Superior Temporal Sulcus. *Cerebral Cortex*, 22(4):735–744, Apr. 2012. ISSN 1460-2199, 1047-3211. doi: 10.1093/cercor/bhr061. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr061`.

M. Clos, R. Langner, M. Meyer, M. S. Oechslin, K. Zilles, and S. B. E. 1. Effects of prior information on decoding degraded speech: An fmri study. *Human Brain Mapping*, 35:61–74, 2014.

J. Davey, H. E. Thompson, G. Hallam, T. Karapanagiotidis, C. Murphy, I. De Caso, K. Krieger-Redwood, B. C. Bernhardt, J. Smallwood, and E. Jefferies. Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage*, 137:165–177, 2016. ISSN 10538119. doi: 10.1016/j.neuroimage.2016.05.051. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811916301719`.

M. H. Davis and M. G. Gaskell. A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London*, 00:1–28, 2009.

M. H. Davis and I. S. Johnsrude. Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23:3423–3431, 2003. doi: 10.1016/j.heares.2007.01.014.

M. H. Davis, M. A. Ford, F. Kherif, and I. S. Johnsrude. Does semantic context benefit speech understanding through "Top – Down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, 23:3914–3932, 2011.

G. I. de Zubicaray and K. L. McMahon. Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscienc*, 9(3):260–269, 2009. ISSN 1530-7026, 1531-135X. doi: 10.3758/CABN.9.3.260. URL `http://link.springer.com/10.3758/CABN.9.3.260`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 39:1–38, 1977.

A.-D. Devauchelle, C. Oppenheim, L. Rizzi, S. Dehaene, and C. Pallier. Sentence Syntax and Content in the Human Temporal Lobe: An fMRI Adaptation Study in Auditory and Visual Modalities. *Journal of Cognitive Neuroscience*, 25:1000–1012, 2008.

M. T. Diaz and G. McCarthy. A comparison of brain activity evoked by single content and function words: An fMRI investigation of implicit word processing. *Brain Research*, 1282:38–49, 2009. ISSN 00068993. doi: 10.1016/j.brainres.2009.05.043. URL `https://linkinghub.elsevier.com/retrieve/pii/S0006899309009937`.

A. S. Dick, S. Goldin-Meadow, U. Hasson, J. I. Skipper, and S. L. Small. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, 30(11):3509–3526, nov 2009. ISSN 10659471, 10970193. doi: 10.1002/hbm.20774. URL `http://doi.wiley.com/10.1002/hbm.20774`.

W. B. Edmister, T. M. Talavage, P. J. Ledden, and R. M. Weisskoff. Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, 7:89–97, 1999.

F. Eisner, C. McGettigan, A. Faulkner, S. Rosen, and S. K. Scott. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, 30(21):7179–7186, 2010. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4040-09.2010. URL `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4040-09.2010`.

J. A. Etzel, J. M. Zacks, and T. S. Braver. Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, 78:261–269, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.03.041. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811913002917`.

S. Evans and M. H. Davis. Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex*, 25(12):4772–4788, 2015. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhv136. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhv136`.

B. S. Everitt and E. T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 1999. doi: 10.1002/(SICI)1097-0193(1999)7:1<1::AID-HBM1>3.0.CO;2-H.

S. Fecteau, J. L. Armony, Y. Joanette, and P. Belin. Is voice processing species-specific in human auditory cortex? an fMRI study. *NeuroImage*, 23(3):840–848, 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.09.019. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811904005440`.

E. Formisano, F. De Martino, M. Bonte, and R. Goebel. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, 322:970–973, 2008. ISSN 0036-8075. doi: %10.1126/science.1164318.

A. D. Friederici. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2):78–84, 2002. ISSN 13646613. doi: 10.1016/S1364-6613(00)01839-8. URL `https://linkinghub.elsevier.com/retrieve/pii/S1364661300018398`.

A. D. Friederici. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91:1357–1392, 2011.

A. D. Friederici, M. Meyer, and D. Y. v. Cramon. Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, 74:289–300, 2000. doi: 10.1006/brln.2000.2313.

A. D. Friederici, S.-A. Rüschemeyer, A. Hahne, and C. J. Fiebach. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex*, 13:170–177, 2003.

A. D. Friederici, C. J. Fiebach, M. Schlesewsky, I. D. Bornkessel, and D. Y. v. Cramon. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, 16:1709–1717, 2006. doi: 10.1093/cercor/bhj106.

A. D. Friederici, D. Y. von Cramon, and S. A. Kotz. Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron*, 53(1):135–145, 2007. ISSN 08966273. doi: 10.1016/j.neuron.2006.11.020. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627306009408`.

A. D. Friederici, S. A. Kotz, S. K. Scott, and J. Obleser. Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, 31:448–457, 2010. ISSN 10659471, 10970193. doi: 10.1002/hbm.20878. URL `http://doi.wiley.com/10.1002/hbm.20878`.

K. Friston, C. Buechel, G. Fink, J. Morris, E. Rolls, and R. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6(3):218–229, 1997. ISSN 10538119. doi: 10.1006/nimg.1997.0291. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811997902913`.

K. J. Friston and W. L. Harrison. Dynamic causal modelling. *NeuroImage*, 19:1273–1302, 2003.

K. J. Friston, A. P. H. a. K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2: 189–210, 1995. doi: 10.1002/hbm.460020402.

J. Gandour, D. Wong, M. Lowe, M. Dzemidzic, N. Satthamnuwong, Y. Tong, and X. Li. A Cross-Linguistic fMRI Study of Spectral and Temporal Cues Underlying Phonological Processing. *Journal of Cognitive Neuroscience*, 14(7):1076–1087, Oct. 2002. ISSN 0898-929X, 1530-8898. doi: 10.1162/089892902320474526. URL `http://www.mitpressjournals.org/doi/10.1162/089892902320474526`.

E. Geiser, T. Zaehle, L. Jancke, and M. Meyer. The Neural Correlate of Speech Rhythm as Evidenced by Metrical Speech Processing. *Journal of Cognitive Neuroscience*, 20(3):541–552, 2008.

A. Giraud. Contributions of Sensory Input, Auditory Search and Verbal Comprehension to Cortical Activity during Speech Processing. *Cerebral Cortex*, 14(3):247–255, Mar. 2004. ISSN 1460-2199. doi: 10.1093/cercor/bhg124. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhg124`.

A.-L. Giraud, C. Lorenzi, J. Ashburner, J. Wable, I. Johnsrude, R. Frackowiak, and A. Kleinschmidt. Representation of the Temporal Envelope of Sounds in the Human Brain. *Journal of Neurophysiology*, 84(3):1588–1598, Sept. 2000. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.2000.84.3.1588. URL `https://www.physiology.org/doi/10.1152/jn.2000.84.3.1588`.

D. R. Gitelman, W. D. Penny, J. Ashburner, and K. J. Friston. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage*, 19(1):200–207, 2003. ISSN 10538119. doi: 10.1016/S1053-8119(03)00058-2. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811903000582`.

N. Golestani, S. Rosen, and S. K. Scott. Native-language benefit for understanding speech-in-noise: The contribution of semantics*. *Bilingualism: Language and Cognition*, 12(03):385, July 2009. ISSN 1366-7289, 1469-1841. doi: 10.1017/S1366728909990150. URL `http://www.journals.cambridge.org/abstract_S1366728909990150`.

J. Gonzalez-Castillo and T. M. Talavage. Reproducibility of fMRI activations associated with auditory sentence comprehension. *NeuroImage*, 54(3):2138–2155, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.09.082. URL `https://linkinghub.elsevier.com/retrieve/pii/S105381191001284X`.

A. Graessner, E. Zaccarella, and G. Hartwigsen. Differential contributions of left-hemispheric language regions to basic semantic composition. *Neuroscience*, 2019. doi: 10.1101/2019.12.11.872457. URL `http://biorxiv.org/lookup/doi/10.1101/2019.12.11.872457`.

D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.

P. Hagoort. Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, 28:136–141, 2014. ISSN 09594388. doi: 10.1016/j.conb.2014.07.013. URL `https://linkinghub.elsevier.com/retrieve/pii/S095943881400141X`.

D. A. Hall, M. P. Haggard, M. A. Akeroyd, A. R. Palmer, A. Q. Summerfield, M. R. Elliott, E. M. Gurney, and R. W. Bowtell. "sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, 7:213–223, 1999.

D. A. Hall, M. P. Haggard, M. A. Akeroyd, A. Q. Summerfield, A. R. Palmer, M. R. Elliott, and R. W. Bowtell. Modulation and task effects in auditory processing measured using fMRI. *Human Brain Mapping*, 10:107–119, 2000.

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

N. V. Hartvig and J. L. Jensen. Spatial mixture modeling of fMRI data. *Human Brain Mapping*, 11: 233–248, 2000. doi: 10.1002/1097-0193(200012)11:4<233::aid-hbm10>3.0.co;2-f.

O. Hauk, I. Johnsrude, and F. Pulvermüller. Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2):301–307, 2004. ISSN 08966273. doi: 10.1016/S0896-6273(03)00838-9. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627303008389`.

S. Haupt, N. Axmacher, M. X. Cohen, C. E. Elger, and J. Fell. Activation of the caudal anterior cingulate cortex due to task-related interference in an auditory stroop paradigm. *Human Brain Mapping*, 30(9):3043–3056, 2009. ISSN 10659471, 10970193. doi: 10.1002/hbm.20731. URL `http://doi.wiley.com/10.1002/hbm.20731`.

L. Hausfeld, G. Valente, and E. Formisano. Multiclass fMRI data decoding and visualization using supervised self-organizing maps. *NeuroImage*, 96:54–66, Aug. 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2014.02.006.

S. Heim, S. B. Eickhoff, and K. Amunts. Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *NeuroImage*, 48(3):616–624, 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.06.044. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909006806`.

A. Heinrich, R. P. Carlyon, M. H. Davis, and I. S. Johnsrude. Illusory Vowels Resulting from Perceptual Continuity: A Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, 20(10):1737–1752, Oct. 2008. ISSN 0898-929X, 1530-8898. doi: 10.1162/jocn.2008.20069. URL `http://www.mitpressjournals.org/doi/10.1162/jocn.2008.20069`.

P.-Y. Hervé, A. Razafimandimby, M. Vigneau, B. Mazoyer, and N. Tzourio-Mazoyer. Disentangling the brain networks supporting affective speech comprehension. *NeuroImage*, 61(4):1255–1267, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.03.073. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811912003540`.

G. Hickok. The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6):393–402, 2012. ISSN 00219924. doi: 10.1016/j.jcomdis.2012.06.004. URL `https://linkinghub.elsevier.com/retrieve/pii/S0021992412000688`.

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

I. Holube, S. Fredelake, M. Vlaming, and B. Kollmeier. Development and analysis of an international speech test signal (ISTS). *International Journal of Audiology*, 49:891–903, 2010.

C. Humphries, J. R. Binder, D. A. Medler, and E. Liebenthal. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18: 665–679, 2006.

K. Hwang, E. D. Palmer, S. Basho, J. R. Zadra, and R.-A. Müller. Category-specific activations during word generation reflect experiential sensorimotor modalities. *NeuroImage*, 48(4):717–725, Dec. 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.06.042. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909006788`.

A. K. Ischebeck, A. D. Friederici, and K. Alter. Processing prosodic boundaries in natural and hummed speech: An fMRI study. *Cerebral Cortex*, 18(3):541–552, 2008. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhm083. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhm083`.

H.-A. Jeon, K.-M. Lee, Y.-B. Kim, and Z.-H. Cho. Neural substrates of semantic relationships: Common and distinct left-frontal activities for generation of synonyms vs. antonyms. *NeuroImage*, 48(2):449–457, Nov. 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.06.049. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909006843`.

I. S. Johnsrude, A. Mackey, H. Hakyemez, E. Alexander, H. P. Trang, and R. P. Carlyon. Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, 24(10):1995–2004, Oct. 2013. ISSN 0956-7976, 1467-9280. doi: 10.1177/0956797613482467. URL `http://journals.sagepub.com/doi/10.1177/0956797613482467`.

L. Jäncke, T. Wüstenberg, H. Scheich, and H.-J. Heinze. Phonetic Perception and the Temporal Cortex. *NeuroImage*, 15(4):733–746, Apr. 2002. ISSN 10538119. doi: 10.1006/nimg.2001.1027. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811901910274`.

T. Kahnt, J. Heinzle, S. Q. Park, and J.-D. Haynes. The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 107(13):6010–6015, Mar. 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0912838107.

H. Kawahara and M. Morise. Technical foundations of Tandem-STRAIGHT, a speech analysis, modification and synthesis framework. *SADHANA - Academy Proceedings in Engeneering Sciences*, 36: 713–722, 2011.

S. Kim, P. Smyth, and H. Stern. A bayesian mixture approach to modeling spatial activation patterns in multisite fMRI data. *IEEE Transaction on Medical Imaging*, 29:1260–1274, 2010. doi: 10.1109/TMI.2010.2044045.

B. Kollmeier and M. Wesselkamp. Development and evaluation of a German sentence test for objective and subsective speech intelligibility assessment. *Journal of Acoustical Society of America*, 102:2412–2421, 1997.

S. Kotz. Modulation of the lexical–semantic network by auditory semantic priming: An event-related functional MRI study. *NeuroImage*, 17(4):1761–1772, 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1316. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811902913169`.

S. Kotz, A. D'Ausilio, T. Raettig, C. Begliomini, L. Craighero, M. Fabbri-Destro, C. Zingales, P. Haggard, and L. Fadiga. Lexicality drives audio-motor transformations in broca's area. *Brain and Language*, 112(1):3–11, 2009. ISSN 0093934X. doi: 10.1016/j.bandl.2009.07.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S0093934X0900087X`.

S. Kouider, V. de Gardelle, S. Dehaene, E. Dupoux, and C. Pallier. Cerebral bases of subliminal speech priming. *NeuroImage*, 49(1):922–929, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.08.043. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909009537`.

N. Kriegeskorte, R. Goebel, and P. Bandettti. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the U.S.A.*, 103:3863–3868, 2006.

J. Kubanek, P. Brunner, A. Gunduz, D. Poeppel, and G. Schalk. The tracking of speech envelope in the human cortex. *PLoS ONE*, 8(1):e53398, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0053398. URL `https://dx.plos.org/10.1371/journal.pone.0053398`.

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

G. R. Kuperberg, P. K. McGuire, E. T. Bullmore, M. J. Brammer, S. Rabe-Hesketh, I. C. Wright, D. J. Lythgoe, S. C. R. Williams, and A. S. David. Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *Journal of Cognitive Neuroscience*, 12(2):321–341, 2000. ISSN 0898-929X, 1530-8898. doi: 10.1162/089892900562138. URL `http://www.mitpressjournals.org/doi/10.1162/089892900562138`.

S. Lattner, M. E. Meyer, and A. D. Friederici. Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, 24(1):11–20, Jan. 2005. ISSN 1065-9471, 1097-0193. doi: 10.1002/hbm.20065. URL `http://doi.wiley.com/10.1002/hbm.20065`.

T. O. Laumann, E. M. Gordon, B. Adeyemo, A. Z. Snyder, S. J. Joo, M.-Y. Chen, A. W. Gilmore, K. B. McDermott, S. M. Nelson, N. U. Dosenbach, B. L. Schlaggar, J. A. Mumford, R. A. Poldrack, and S. E. Petersen. Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, 87(3):657–670, Aug. 2015. ISSN 08966273. doi: 10.1016/j.neuron.2015.06.037. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627315006005`.

E. Liebenthal, J. R. Binder, S. M. Spitzer, E. T. Possing, and D. A. Medler. Neural substrates of phonemic perception. *Cerebral Cortex*, 15(10):1621–1631, 2005. doi: 10.1093/cercor/bhi040. URL `+http://dx.doi.org/10.1093/cercor/bhi040`.

E. Liebenthal, R. Desai, M. M. Ellingson, B. Ramachandran, A. Desai, and J. R. Binder. Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, 20(12):2958–2970, 2010. ISSN 1460-2199, 1047-3211. doi: 10.1093/cercor/bhq045. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhq045`.

M. D. Lieberman and W. A. Cunningham. Type I and type II error concerns in fMRI research: Rebalancing the scale. *Social Cognitive and Affective Neuroscience*, 4:423–428, 2009. doi: 10.1093/scan/nsp052.

P. C. LoCasto, D. Krebs-Noble, R. P. Gullapalli, and M. W. Burton. An fMRI investigation of speech and tone segmentation. *Journal of Cognitive Neuroscience*, 16:1612–1624, 2004. doi: 10.1162/0898929042568433.

V. Madisetti and D. B. Williams, editors. *The digital signal processing handbook*. The electrical engineering handbook series. CRC Press ; IEEE Press, Boca Raton, Fla. : [New York], 1998. ISBN 978-0-8493-8572-8.

A. Mahmoudi, S. Takerkart, F. Regragui, D. Boussaoud, and A. Brovelli. Multivoxel pattern analysis for fMRI data: A review. *Computational and Mathematical Methods in Medicine*, 2012:1–14, 2012. ISSN 1748-670X, 1748-6718. doi: 10.1155/2012/961257. URL `http://www.hindawi.com/journals/cmmm/2012/961257/`.

M. Makuuchi, J. Bahlmann, A. Anwander, and A. D. Friederici. Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the U.S.A.*, 106:8362–8367, 2009. doi: 10.1073/pnas.0810928106.

C. J. Markiewicz and J. W. Bohland. Mapping the cortical representation of speech sounds in a syllable repetition task. *NeuroImage*, 141:174–190, 2016. doi: 10.1016/j.neuroimage.2016.07.023.

N. Mashal, M. Faust, T. Hendler, and M. Jung-Beeman. An fMRI study of processing novel metaphoric sentences. *Laterality: Asymmetries of Body, Brain and Cognition*, 14(1):30–54, Jan. 2009. ISSN 1357-650X, 1464-0678. doi: 10.1080/13576500802049433. URL `http://www.tandfonline.com/doi/abs/10.1080/13576500802049433`.

C. McGettigan, S. Evans, S. Rosen, Z. K. Agnew, P. Shah, and S. Scott. An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience*, 24:636–652, 2012. doi: 10.1162/jocn_a_00161.

C. McGettigan, A. Faulkner, I. Altarelli, J. Obleser, H. Baverstock, and S. K. Scott. Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, 50(5):762–776, 2012. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2012.01.010. URL `https://linkinghub.elsevier.com/retrieve/pii/S0028393212000267`.

C. McGettigan, S. Rosen, and S. K. Scott. Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: implications for cochlear implantation. *Frontiers in Systems Neuroscience*, 8, 2014. ISSN 1662-5137. doi: 10.3389/fnsys.2014.00018. URL `http://journal.frontiersin.org/article/10.3389/fnsys.2014.00018/abstract`.

D. G. McLaren, M. L. Ries, G. Xu, and S. C. Johnson. A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, 61 (4):1277–1286, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.03.068. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811912003497`.

M. Meyer, K. Steinhauer, K. Alter, A. D. Friederici, and D. von Cramon. Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89(2):277–289, May 2004. ISSN 0093934X. doi: 10.1016/S0093-934X(03)00350-X. URL `https://linkinghub.elsevier.com/retrieve/pii/S0093934X0300350X`.

M. Mikl, R. Mareček, P. Hluštík, M. Pavlicová, A. Drastich, P. Chlebus, M. Brázdil, and P. Krupa. Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, 26(4):490–503, May 2008. ISSN 0730725X. doi: 10.1016/j.mri.2007.08.006. URL `https://linkinghub.elsevier.com/retrieve/pii/S0730725X07003955`.

M. Misaki, Y. Kim, P. Bandettini, and N. Kriegeskorte. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, 53:103–118, 2010. doi: 10.1016/j.neuroimage.2010.05.051.

M. Moerel, F. De Martino, and E. Formisano. Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, 32(41):14205–14216, 2012. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1388-12.2012. URL `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1388-12.2012`.

A. B. Morland, H. A. Baseler, M. B. Hoffmann, L. T. Sharpe, and B. A. Wandell. Abnormal retinotopic representations in human visual cortex revealed by fMRI. *Acta Psychologica*, 107(1):229–247, 2001. ISSN 00016918. doi: 10.1016/S0001-6918(01)00025-7. URL `https://linkinghub.elsevier.com/retrieve/pii/S0001691801000257`.

C. Narain. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12):1362–1368, 2003. ISSN 1460-2199. doi: 10.1093/cercor/bhg083. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhg083`.

A. J. Newman, R. Pancheva, K. Ozawa, H. J. Neville, and M. T. Ullman. An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research*, 30(3):339–364, 2001.

S. D. Newman and D. Twieg. Differences in auditory processing of words and pseudowords: An fMRI study. *Human Brain Mapping*, 14(1):39–47, 2001. ISSN 1065-9471, 1097-0193. doi: 10.1002/hbm.1040. URL `http://doi.wiley.com/10.1002/hbm.1040`.

K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9):424–430, 2006. ISSN 13646613. doi: 10.1016/j.tics.2006.07.005. URL `https://linkinghub.elsevier.com/retrieve/pii/S1364661306001847`.

C. Nosarti, A. Mechelli, D. W. Green, and C. J. Price. The impact of second language learning on semantic and nonsemantic first language reading. *Cerebral Cortex*, 20(2):315–327, 2010. ISSN 1460-2199, 1047-3211. doi: 10.1093/cercor/bhp101. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp101`.

J. Obleser and S. A. Kotz. Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3):633–640, 2010. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhp128. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp128`.

J. Obleser, R. J. S. Wise, M. Alex Dresner, and S. K. Scott. Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, 27 (9):2283–2289, 2007a. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4663-06.2007. URL `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4663-06.2007`.

J. Obleser, J. Zimmermann, J. V. Meter, and J. P. Rauschecker. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17:2251–2257, 2007b.

V. P. Oikonomou and K. Blekas. An adaptive regression mixture model for fMRI cluster analysis. *IEEE Transaction on Medical Imaging*, 32:649–659, 2013. doi: 10.1109/TMI.2012.2221731.

K. Okada, F. Rong, J. Venezia, W. Matchin, I.-H. Hsieh, K. Saberi, J. T. Serences, and G. Hickok. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, 20:2486–2495, 2010. doi: %10.1093/cercor/bhp318.

N. N. Oosterhof, A. J. Wiggett, J. Diedrichsen, S. P. Tipper, and P. E. Downing. Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, 104(2):1077–1089, 2010. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00326.2010. URL `https://www.physiology.org/doi/10.1152/jn.00326.2010`.

B. Osnes, K. Hugdahl, and K. Specht. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage*, 2011. doi: 10.1016/j.neuroimage.2010.09.078.

J. E. Peelle. The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Frontiers in Human Neuroscience*, 6, 2012. ISSN 1662-5161. doi: 10.3389/fnhum.2012.00309. URL `http://journal.frontiersin.org/article/10.3389/fnhum.2012.00309/abstract`.

G. Pendse, D. Borsook, and L. Becerra. Enhanced false discovery rate using Gaussian mixture models for thresholding fMRI statistical maps. *NeuroImage*, 47:231–261, 2009. doi: 10.1016/j.neuroimage.2009.02.035.

W. Penny and K. Friston. Mixtures of general linear models for functional neuroimaging. *IEEE Transaction On Medical Imaging*, 22:504–514, April 2003. doi: 10.1109/TMI.2003.809140.

F. Pereira and M. Botvinick. Information mapping with pattern classifiers: a comparative study. *NeuroImage*, 2011. doi: 10.1016/j.neuroimage.2010.05.026.

F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, 45(1):199–209, 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.11.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811908012263`.

C. R. Pernet, P. McAleer, M. Latinus, K. J. Gorgolewski, I. Charest, P. E. Bestelmeyer, R. H. Watson, D. Fleming, F. Crabbe, M. Valdes-Sosa, and P. Belin. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, 119:164–174, 2015. ISSN 10538119. doi: 10.1016/j.neuroimage.2015.06.050. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811915005558`.

R. A. Poldrack. Precision Neuroscience: Dense Sampling of Individual Brains. *Neuron*, 95(4):727–729, Aug. 2017. ISSN 08966273. doi: 10.1016/j.neuron.2017.08.002. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627317306931`.

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

C. J. Price. The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Science*, 1191(1):62–88, 2010.

T. Raettig, S. Frisch, A. D. Friederici, and S. A. Kotz. Neural correlates of morphosyntactic and verb-argument structure processing: An EfMRI study. *Cortex*, 46(5):613–620, May 2010. ISSN 00109452. doi: 10.1016/j.cortex.2009.06.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S0010945209001841`.

F. M. Richardson and C. J. Price. Structural MRI studies of language function in the undamaged brain. *Brain Structure and Function*, 213(6):511–523, Oct. 2009. ISSN 1863-2653, 1863-2661. doi: 10.1007/s00429-009-0211-y. URL `http://link.springer.com/10.1007/s00429-009-0211-y`.

A. Riecker, D. Wildgruber, G. Dogil, W. Grodd, and H. Ackermann. Hemispheric Lateralization Effects of Rhythm Implementation during Syllable Repetitions: An fMRI Study. *NeuroImage*, 16 (1):169–176, May 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1068. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811902910682`.

J. Rissmann, J. C. Eliassen, and S. E. Blumstein. An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, 15:1160–1175, 2003. doi: 10.1162/089892903322598120.

J. M. Rodd, M. H. Davis, and I. S. Johnsrude. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 2005. doi: doi:10.1093/cercor/bhi009.

J. M. Rodd, I. S. Johnsrude, and M. H. Davis. Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, 22(8):1761–1773, 2012. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhr252. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr252`.

C. Rogalsky and G. Hickok. Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, 19(4):786–796, Apr. 2009. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhn126. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhn126`.

J. C. Rogers and M. H. Davis. Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of Cognitive Neuroscience*, 29(5):919–936, 2017. ISSN 0898-929X, 1530-8898. doi: 10.1162/jocn_a_01096. URL `http://www.mitpressjournals.org/doi/10.1162/jocn_a_01096`.

K. Rothermich and S. A. Kotz. Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *NeuroImage*, 70:89–100, Apr. 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.12.013. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811912011901`.

R. J. Schafer and T. Constable. Modulation of functional connectivity with the syntactic and semantic demands of a noun phrase formation task: A possible role for the default network. *NeuroImage*, 46(3): 882–890, 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.02.017. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909001608`.

G. L. Schmidt and C. A. Seger. Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition*, 71(3):375–386, Dec. 2009. ISSN 02782626. doi: 10.1016/j.bandc.2009.06.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S0278262609000980`.

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0262194759.

S. K. Scott. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123 (12):2400–2406, 2000. ISSN 14602156. doi: 10.1093/brain/123.12.2400. URL `https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/123.12.2400`.

M. Seghier, H. Lee, T. Schofield, C. Ellis, and C. Price. Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words. *NeuroImage*, 42(3):1226–1236, 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.05.029. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811908006526`.

M. L. Seghier and C. J. Price. Visualising inter-subject variability in fMRI using threshold-weighted overlap maps. *Scientific Reports*, 6(1):20170, Apr. 2016. ISSN 2045-2322. doi: 10.1038/srep20170. URL `http://www.nature.com/articles/srep20170`.

M. L. Seghier, F. Lazeyras, A. J. Pegna, J.-M. Annoni, I. Zimine, E. Mayer, C. M. Michel, and A. Khateb. Variability of fMRI activation during a phonological and semantic language task in healthy subjects. *Human Brain Mapping*, 23(3):140–155, 2004. ISSN 1097-0193. doi: 10.1002/hbm.20053. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.20053`.

D. J. Sharp, M. Awad, J. E. Warren, R. J. Wise, G. Vigliocco, and S. K. Scott. The neural response to changing semantic and perceptual complexity during language processing. *Human Brain Mapping*, pages NA–NA, 2009. ISSN 10659471, 10970193. doi: 10.1002/hbm.20871. URL `http://doi.wiley.com/10.1002/hbm.20871`.

D. Smirnov, E. Glerean, J. M. Lahnakoski, J. Salmi, I. P. Jääskeläinen, M. Sams, and L. Nummenmaa. Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia*, 63:293–303, 2014. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2014.09.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0028393214003121`.

T. M. Snijders, T. Vosse, G. Kempen, J. J. Van Berkum, K. M. Petersson, and P. Hagoort. Retrieval and Unification of Syntactic Structure in Sentence Comprehension: an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex*, 19(7):1493–1503, July 2009. ISSN 1460-2199, 1047-3211. doi: 10.1093/cercor/bhn187. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhn187`.

E. Sohoglu, J. E. Peelle, R. P. Carlyon, and M. H. Davis. Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25):8443–8453, 2012. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5069-11.2012. URL `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5069-11.2012`.

K. A. Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. In A. M. Segre, editor, *Proceedings of the Sixth International Workshop on Machine Learning*, pages 160 – 163. Morgan Kaufmann, San Francisco (CA), 1989. ISBN 978-1-55860-036-2.

C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 1904.

J. Stelzer, , Y. Chen, and R. Turner. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, 65:69–82, 2013. doi: 10.1016/j.neuroimage.2012.09.063.

B. Thirion, P. Pinel, S. Mériaux, A. Roche, S. Dehaene, and J.-B. Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage*, 35(1):105–120, Mar. 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.11.054. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811906011682`.

P. Tremblay and A. S. Dick. Broca and wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, 162:60–71, 2016. ISSN 0093934X. doi: 10.1016/j.bandl.2016.08.004. URL `https://linkinghub.elsevier.com/retrieve/pii/S0093934X16300475`.

L. K. Tyler and W. Marslen-Wilson. Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society*, 363, 2007. doi: 10.1098/rstb.2007.2158.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1):273–289, Jan. 2002. ISSN 10538119. doi: 10.1006/nimg.2001.0978. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811901909784`.

L. Q. Uddin. Complex relationships between structural and functional brain connectivity. *Trends in Cognitive Sciences*, 17(12):600–602, 2013. ISSN 13646613. doi: 10.1016/j.tics.2013.09.011. URL `https://linkinghub.elsevier.com/retrieve/pii/S1364661313002155`.

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

M. P. van den Heuvel and H. E. Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010. ISSN 0924977X. doi: 10.1016/j.euroneuro.2010.03.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S0924977X10000684`.

T. Vincent, L. Risser, and P. Ciuciu. Spatially adaptive mixture modeling for analysis of fMRI time series. *IEEE Transaction on Medical Imaging*, 29:1059–1074, 2010. doi: 10.1109/TMI.2010.2042064.

K. von Kriegstein, E. Eger, A. Kleinschmidt, and A. L. Giraud. Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, 17(1):48–55, 2003. ISSN 09266410. doi: 10.1016/S0926-6410(03)00079-X. URL `https://linkinghub.elsevier.com/retrieve/pii/S092664100300079X`.

J. Warren, S. K. Scott, C. J. Price, and T. D. Griffiths. Human brain mechanisms for the early analysis of voices. *NeuroImage*, 31:1389–1397, 2006.

X. Wei, S.-S. Yoo, C. C. Dickey, K. H. Zou, C. R. Guttmann, and L. P. Panych. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *NeuroImage*, 21(3):1000–1008, Mar. 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2003.10.039. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811903007092`.

C. Wernicke. *Der aphasische Symptomencomplex - Eine psychologische Studie auf anatomischer Basis.* Breslau, Cohn & Weigert, 1874.

C. M. Wessinger, M. H. Buonocore, C. L. Kussmaul, and G. R. Mangun. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, 5:18–25, 1997.

S. Whitfield-Gabrieli and A. Nieto-Castanon. *Conn* : A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, 2(3):125–141, 2012. ISSN 2158-0014, 2158-0022. doi: 10.1089/brain.2012.0073. URL `http://www.liebertpub.com/doi/10.1089/brain.2012.0073`.

C. Whitney, S. Weis, T. Krings, W. Huber, M. Grossman, and T. Kircher. Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *Journal of Cognitive Neuroscience*, 21(4):697–712, 2009. ISSN 0898-929X, 1530-8898. doi: 10.1162/jocn.2009.21056. URL `http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21056`.

L. Whitten. Functional magnetic resonance imaging (fMRI): An invaluable tool in translational neuroscience. *Functional Magnetic Resonance Imaging (fMRI)*, pages 1–21, 2012. doi: 10.3768/rtipress.2012.op.0010.1212. URL `http://www.rti.org/publication/functional-magnetic-resonance-imaging-fmri-invaluable-tool-translational-neuroscience`. Edition: 1.

C. J. Wild, M. H. Davis, and I. S. Johnsrude. Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, 60:1490–1502, 2012. doi: 10.1016/j.neuroimage.2012.01.035.

M. Wittfoth, C. Schroder, D. M. Schardt, R. Dengler, H.-J. Heinze, and S. A. Kotz. On Emotional Conflict: Interference Resolution of Happy and Angry Prosody Reveals Valence-Specific Effects. *Cerebral Cortex*, 20(2):383–392, Feb. 2010. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhp106. URL `https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhp106`.

Z. Ye and X. Zhou. Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, 48(1):280–290, Oct. 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.06.032. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811909006582`.

A. Zalesky, A. Fornito, and E. T. Bullmore. Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4):1197–1207, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.06.041. URL `https://linkinghub.elsevier.com/retrieve/pii/S1053811910008852`.

A. Supplementary Material Chapter 2

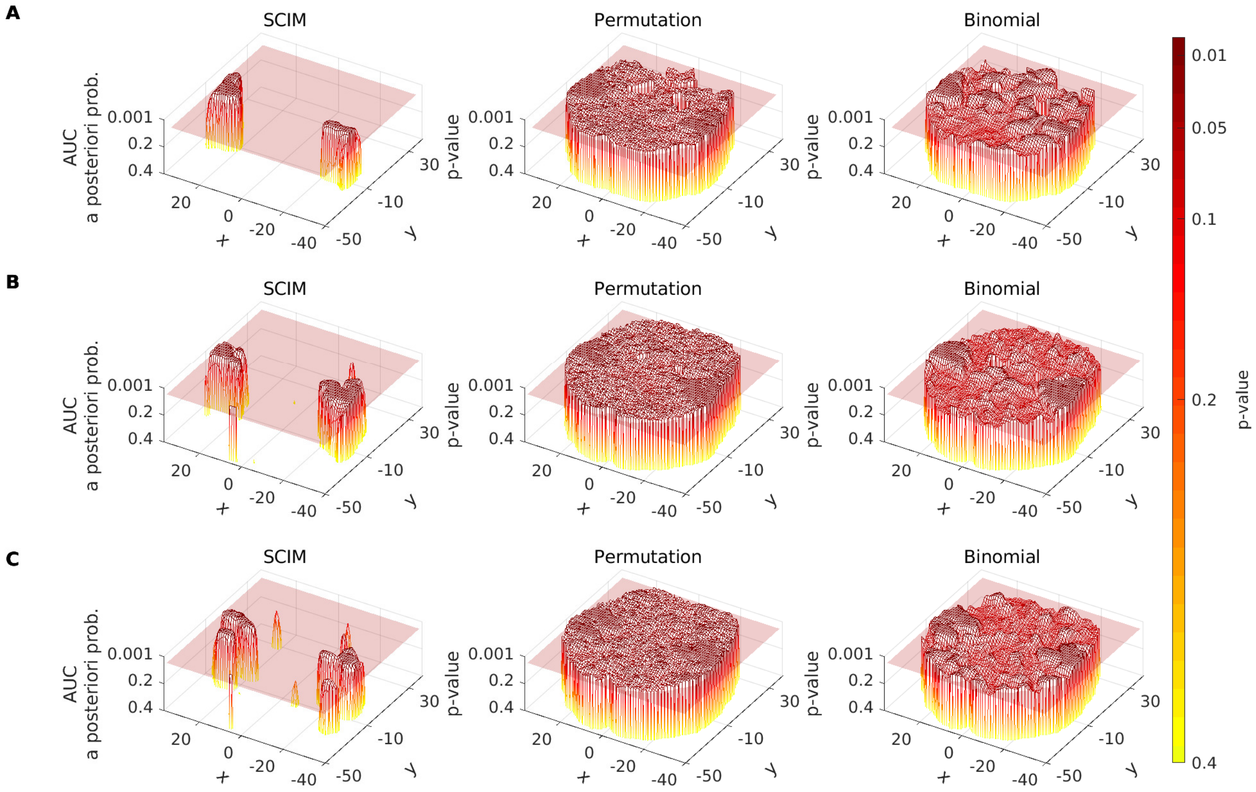

Figure A.1.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutationtest and binomial test) across a single slice (z= -2mm, single subject, evaluation measure AUC, spatialsmoothing for SCIM, random permutation and binomial) of single subject results from three different subjects. Panel A depicts results for subject 1, panel B depicts results for subject 2 and panel C depicts results for subject 3.

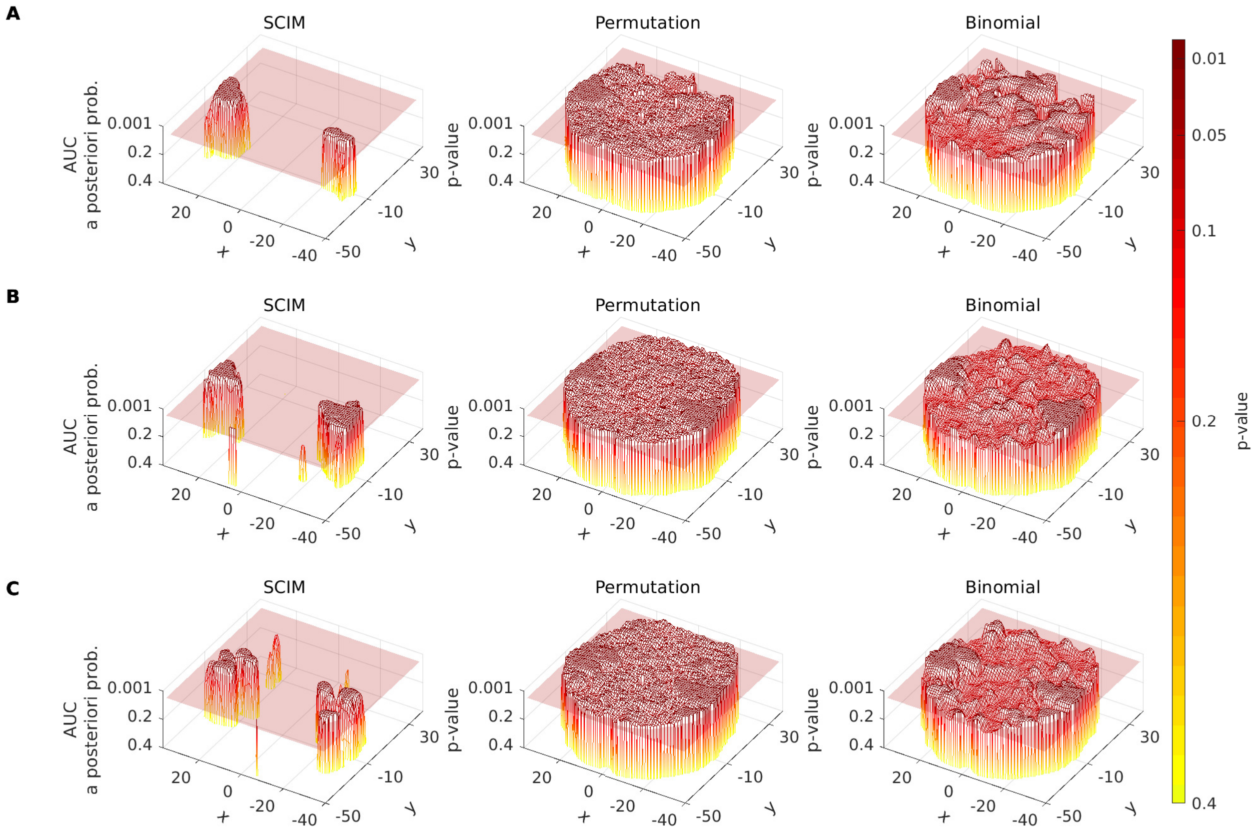

Figure A.2.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutationtest and binomial test) across a single slice (z= 2mm, single subject, evaluation measure AUC, spatialsmoothing for SCIM, random permutation and binomial) of single subject results from three different subjects. Panel A depicts results for subject 1, panel B depicts results for subject 2 and panel C depicts results for subject 3.

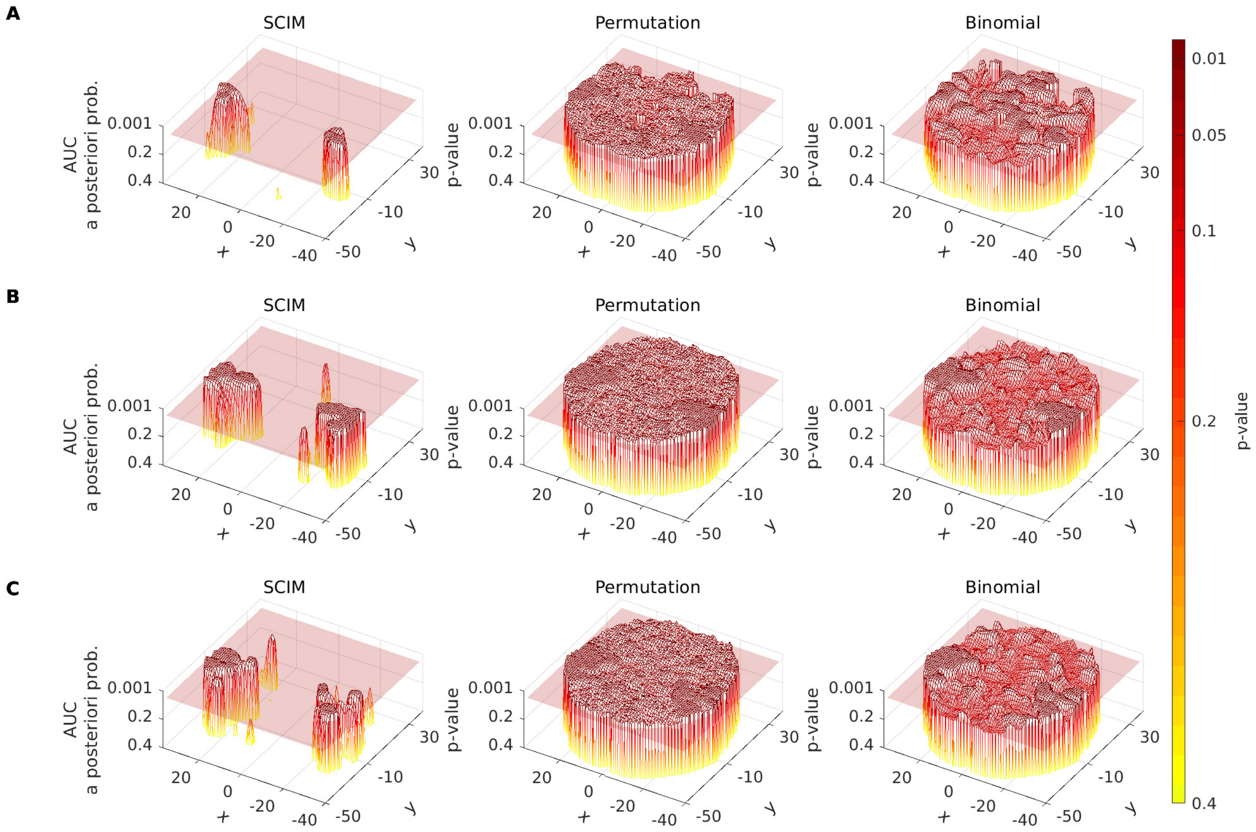

Figure A.3.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutationtest and binomial test) across a single slice (z= 10mm, single subject, evaluation measure AUC, spatial smoothing for SCIM, random permutation and binomial) of single subject results from three different subjects. Panel A depicts results for subject 1, panel B depicts results for subject 2 and panel C depicts results for subject 3.

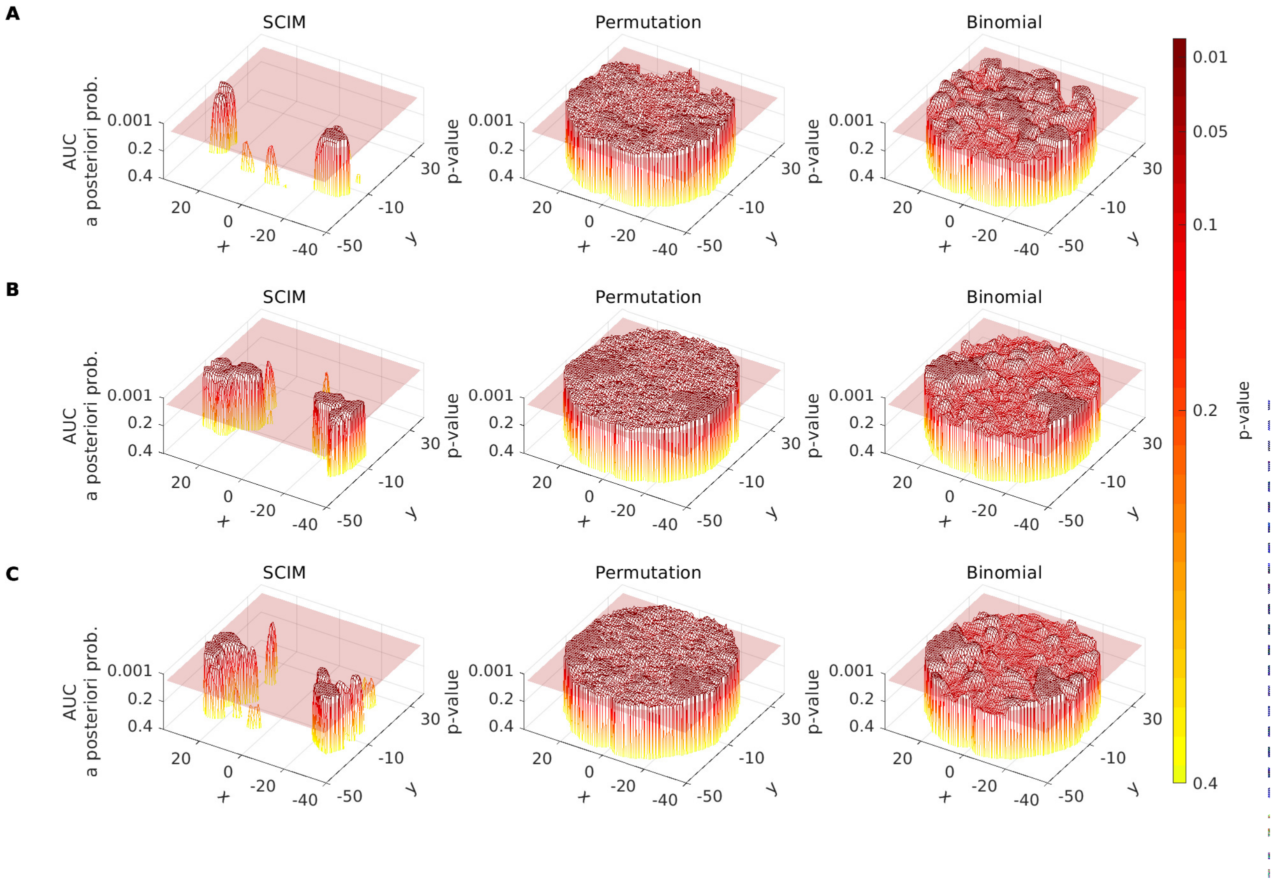

Figure A.4.: Spatial distribution of a-posteriori probabilities p_{SCIM} (SCIM) and p-values (random permutationtest and binomial test) across a single slice (z= 14mm, single subject, evaluation measure AUC, spatial smoothing for SCIM, random permutation and binomial) of single subject results from three different subjects. Panel A depicts results for subject 1, panel B depicts results for subject 2 and panel C depicts results for subject 3.

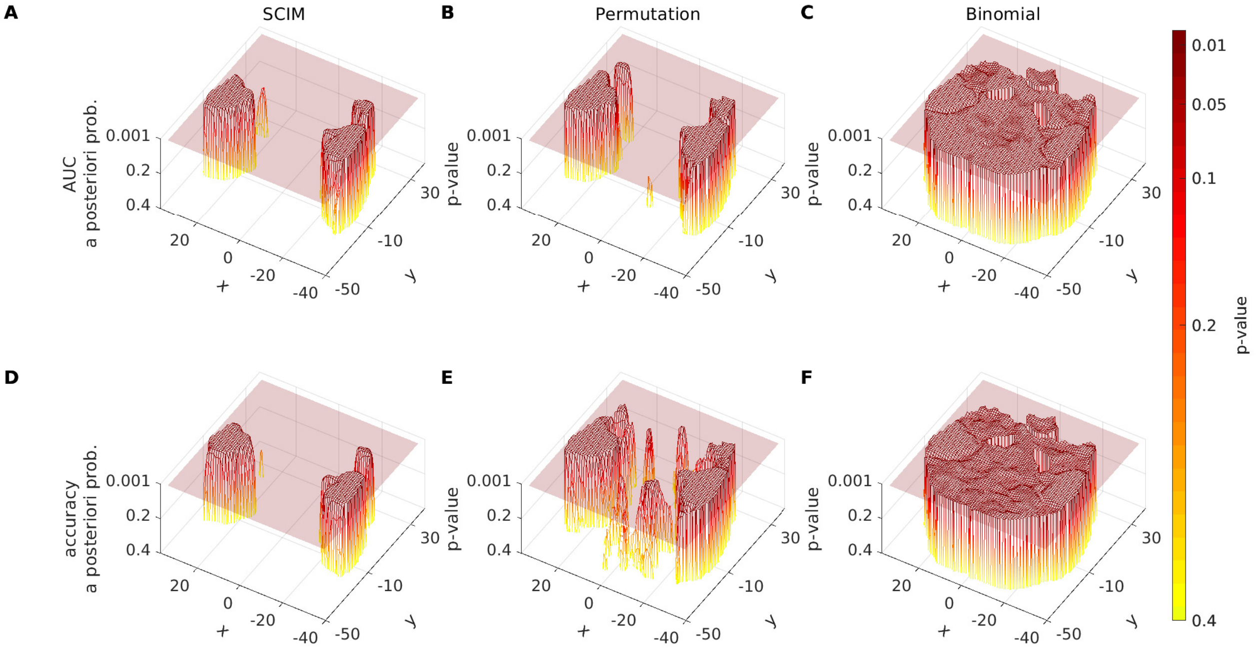

Figure A.5.: Spatial distribution of a-posteriori probabilities p_{SCIM} (in the panels A and D) and p-values (random permutation test in the panels B and E and binomial test in the panels C and F) across a single slice from group result maps (z= -2mm, group results, evaluation measure AUC (A-C) and accuracy (D-F), spatial smoothing.

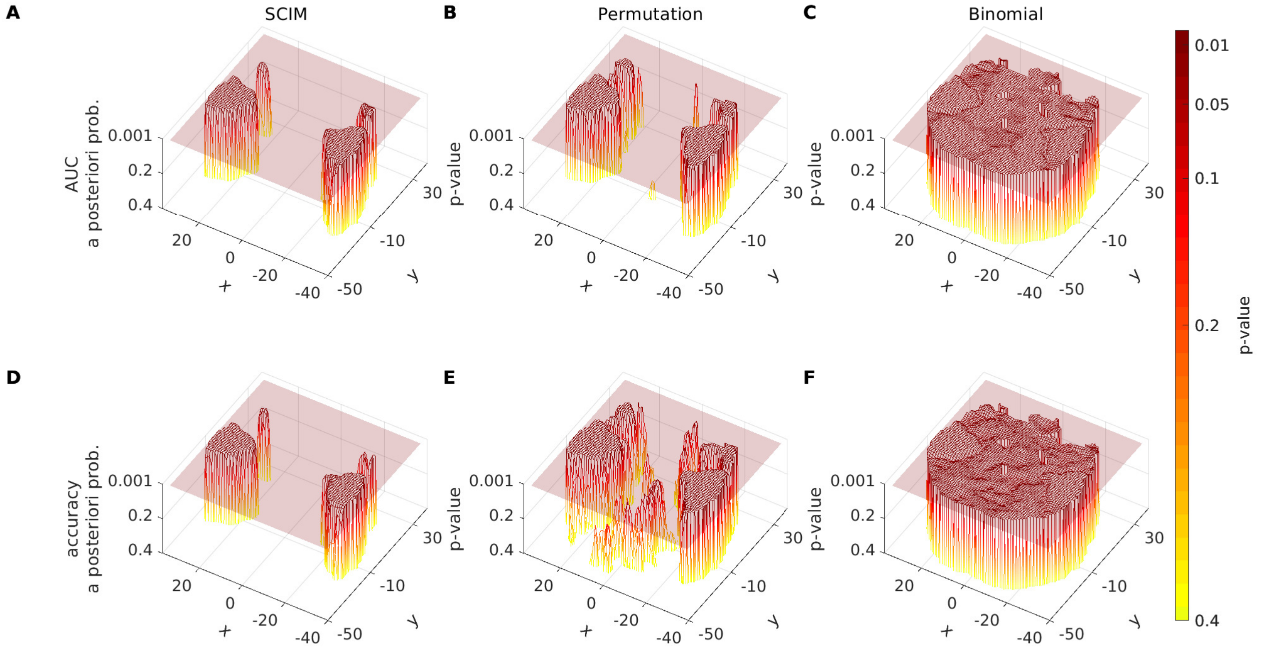

Figure A.6.: Spatial distribution of a-posteriori probabilities p_{SCIM} (in the panels A and D) and p-values (random permutation test in the panels B and E and binomial test in the panels C and F) across a single slice from group result maps (z= 2mm, group results, evaluation measure AUC (A-C) and accuracy (D-F), spatial smoothing.

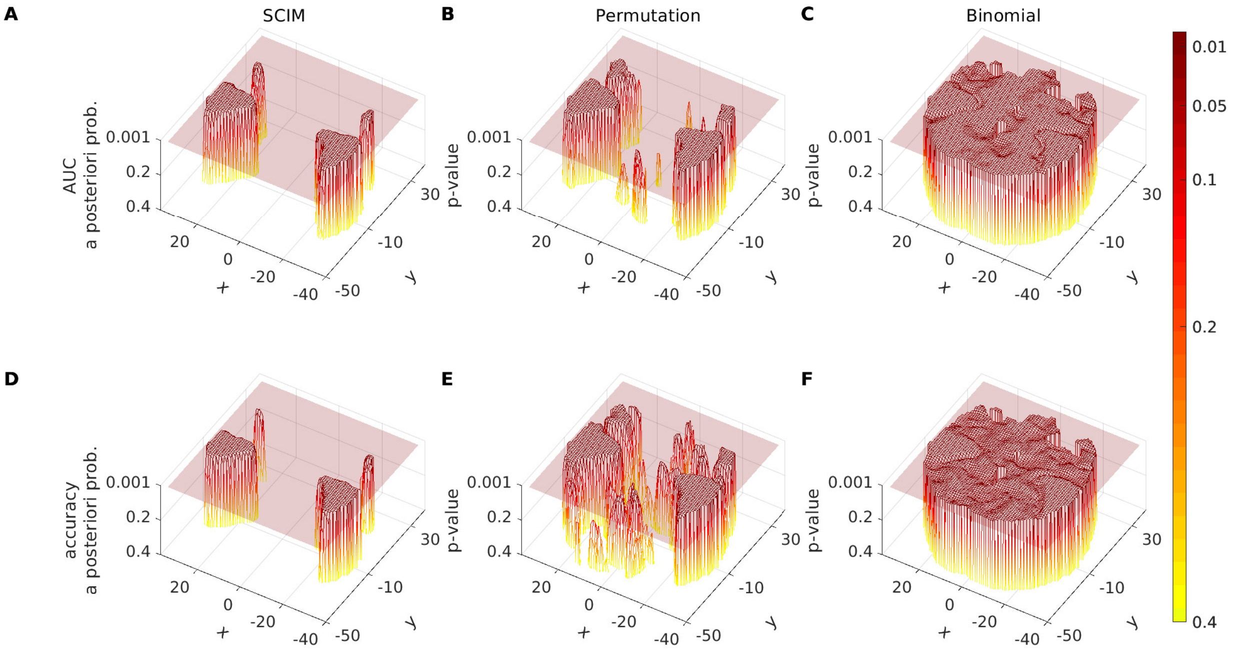

Figure A.7.: Spatial distribution of a-posteriori probabilities p_{SCIM} (in the panels A and D) and p-values (random permutation test in the panels B and E and binomial test in the panels C and F) across a single slice from group result maps (z= 10mm, group results, evaluation measure AUC (A-C) and accuracy (D-F), spatial smoothing.

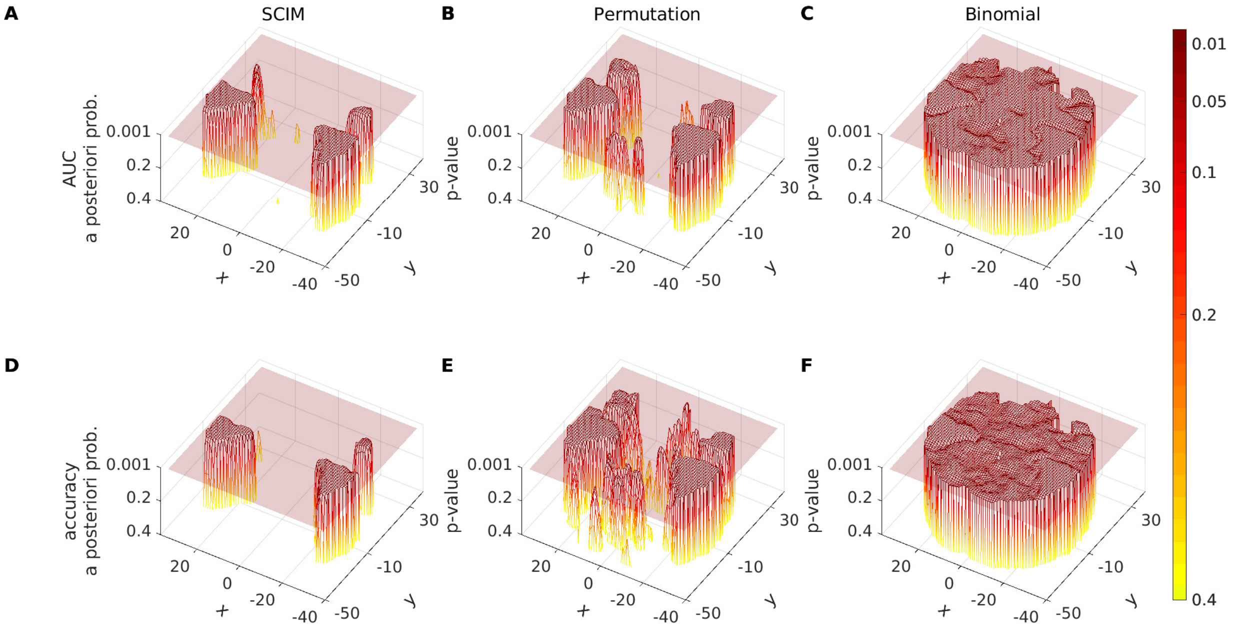

Figure A.8.: Spatial distribution of a-posteriori probabilities p_{SCIM} (in the panels A and D) and p-values (random permutation test in the panels B and E and binomial test in the panels C and F) across a single slice from group result maps (z= 14mm, group results, evaluation measure AUC (A-C) and accuracy (D-F), spatial smoothing.

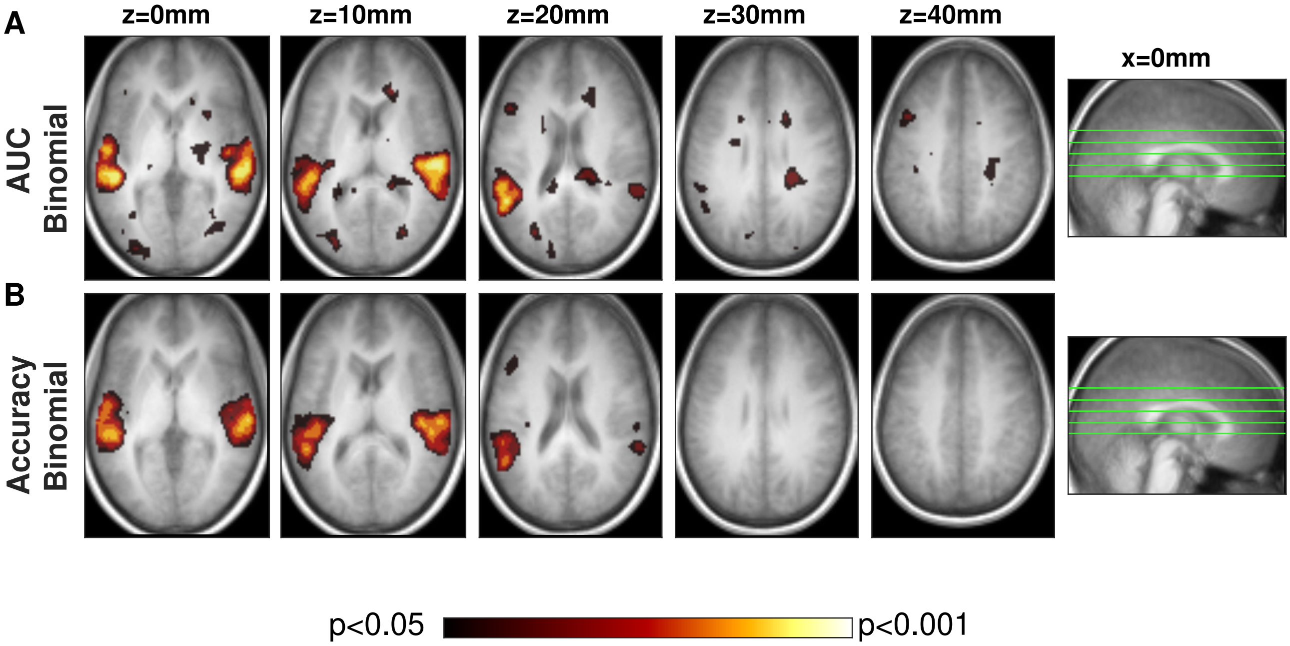

Figure A.9.: Group result maps for the contrast semantic speech vs. non-semantic speech, with binomial test on (A) AUC maps and (B) accuracy maps in five transversal slices and one saggital slice to display location of transversal slices.

Danksagung

Am Ende meiner Zeit als Doktorandin möchte ich mich bei allen bedanken, die mich auf diesem Weg begleitet haben und es mir mit ihrer Unterstützung ermöglicht haben, diese Arbeit zu verfassen.

Ich danke Stefan Uppenkamp für die Betreuung dieser Dissertation und die langjährige Zusammenarbeit und Unterstützung sowie dafür, dass er als Ansprechpartner für inhaltliche Fragen wie auch organisatorische Hürden immer ein offenes Ohr hatte. Dabei wurde mir immer sehr viel Freiraum gelassen, um eigene und neue Ideen zu verfolgen und umzusetzen. Die Arbeitsatmosphäre in der Arbeitsgruppe, sowie Traditionen wie das weihnachtliche Glühweintrinken mit anschließendem Pizzaessen, habe ich immer sehr genossen und werde ich in guter Erinnerung behalten.

Jörn Anemüller danke ich für die Betreuung und Zusammenarbeit bei der Entwicklung des Methodenteils mit vielen langen und spannenden Diskussionen über statistische Verfahren und Ansätze, die oft in vollgeschriebenen Whiteboards und diversen offenliegenden Statistikbuch-Klassikern endeten.

Katharina von Kriegstein möchte ich für die Begutachtung meiner Dissertation danken.

Birger Kollmeier möchte ich für die finanzielle Ermöglichung der Endphase meiner Dissertation und die organisatorische Unterstützung danken.

Der Deutschen Forschungsgemeinschaft DFG danke ich für die Finanzierung der Arbeit über die Projekte SFB 1330/B3 no. 352015383, FOR 1732 und SFB-TRR 31.

Meinen Kollegen und insbesondere Mareike Buhl, David Reineke-Hülsmeier und Bernd T. Meyer danke ich für die moralische, organisatorische und lektorische Unterstützung, sowie aufbauende Worte in besonders herausfordernden Phasen der Dissertation. Außerdem danke ich Mareike Buhl und Sven Herrmann für die gute Zusammenarbeit im Amt der Jüngstenratsprecher und das super Miteinander im gesamten Jüngstenrat mit vielen schönen Bier-, Glüh- und Grillgefühlen. Auch bei Jana Müller und Sabine Hochmuth möchte ich mich für die schöne, gemeinsame Zeit als Bürokollegen bedanken, in der im Mittagstief die eine oder andere Tasse koffeinhaltiger Kakao geflossen ist.

Ein besonderer Dank gilt auch den TAs der Medi, deren Türen immer offen standen und die bereitwillig bei den kleinen Problemen des Alltags Hilfe angeboten haben, sei es bei der Beschaffung von Material oder beim Sortieren und Ausfüllen komplexerer Formulare.

Ich danke der NeuC0p-Gruppe für spannende und inspirierede Diskussionen. Dabei gilt mein Dank insbesondere Manfred Mauermann und Oliver Behler für kritische Nachfragen, die oftmals sehr detaillierte und ausufernde, aber durchaus hilfreiche Gespräche nach sich zogen.

Abschließend möchte ich meinen Eltern Margret und Harry sowie meinem Partner Moritz danken, die mich durch Höhen und Tiefen der Dissertationszeit begleitet haben und auf deren Unterstützung und Rückhalt ich mich immer verlassen konnte.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Die Dissertation hat weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen. Teile der Dissertation wurden bereits veröffentlicht, wie an den entsprechenden Stellen angegeben. Die Beiträge der Autoren entsprechen den Angaben in dem entsprechenden Kapitel. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt und keine kommerziellen Vermittlungs- oder Beratungsdienste in Anspruch genommen habe.

Annika Urbschat
Matrikelnummer 6387227
Oldenburg, October 18, 2021