Department of Biomedical Engineering and Computational Science

Brain Mechanisms Underlying Perception of Naturalistic Social Events

---

Juha Lahnakoski

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 183/2014

Brain Mechanisms Underlying Perception of Naturalistic Social Events

Juha Lahnakoski

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the Auditorium F239a of the Department of Biomedical Engineering and Computational Science on 16th January 2015 at 12 noon.

Aalto University
School of Science
Department of Biomedical Engineering and Computational Science
Brain and Mind Laboratory

Supervising professor
Professor Mikko Sams

Thesis advisors
Assistant professor Lauri Nummenmaa
Adjunct professor Iiro Jääskeläinen

Preliminary examiners
Professor Christian Keysers, Netherlands Institute for Neuroscience, Netherlands
Professor Aapo Hyvärinen, University of Helsinki, Finland

Opponent
Ph.D. Andreas Bartels, University of Tübingen, Germany

Aalto University publication series
DOCTORAL DISSERTATIONS 183/2014

© Juha Lahnakoski

ISBN 978-952-60-5957-0 (printed)
ISBN 978-952-60-5958-7 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5958-7

Unigrafia Oy
Helsinki 2014

Finland

NORDIC ECOLABEL
441 697
Printed matter

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Juha Lahnakoski

Name of the doctoral dissertation
Brain Mechanisms Underlying Perception of Naturalistic Social Events

Publisher School of Science

Unit Department of Biomedical Engineering and Computational Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 183/2014

Field of research Cognitive neuroscience

Manuscript submitted 13 August 2014 **Date of the defence** 16 January 2015

Permission to publish granted (date) 30 September 2014 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

Understanding the brain basis of the wide variety of skills needed to seamlessly interact with other people in the social world is one of the most important goals of social cognitive neuroscience. However, it has remained unclear how the elementary processes of social interaction that have been studied so far generalize to complex naturalistic settings where multiple social cues have to be dynamically tracked at the same time.

The studies presented here employ movies to depict real-life-like social interactions and map the brain systems that participate in the perception of different aspects of the stimuli, particularly focusing on their social signals. Brain activity during viewing of the movies was recorded with functional magnetic resonance imaging. Additionally, the eye gaze of the subjects was recorded and behavioral measures were acquired in a subset of the experiments. Subjects' viewing patterns and interpretation of the movie were manipulated in one of the studies by asking the subjects to adopt two different perspectives. The brain activity was analyzed by multiple methods: parametric models of stimulus contents, inter-subject correlations of brain activity, independent component analysis, and network analysis.

First study compared methods for analyzing both stimulus-related activity and covariation of networks of brain regions during naturalistic conditions to find networks responding to speech, sound, motion categories and low-level visual information. Second study elucidated the organization of the brain regions participating in processing several types of social contents. The results highlight the role of the posterior superior temporal sulcus as a key structure potentially integrating multiple types of social information. Finally, the third study provides first direct experimental evidence for the hypothesis that shared brain activity across individuals reflects shared understanding of the external world. Consistency of the findings demonstrate the feasibility of studying brain responses to simple stimulus features, social movie content as well as high-level perspective taking tasks during very rich naturalistic audiovisual stimulus conditions.

Keywords Social cognition, perspective taking, brain, fMRI, ICA, GLM, ISC

ISBN (printed) 978-952-60-5957-0 **ISBN (pdf)** 978-952-60-5958-7

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2014

Pages 127 **urn** http://urn.fi/URN:ISBN:978-952-60-5958-7

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Juha Lahnakoski

Väitöskirjan nimi
Luonnollisenkaltaisten sosiaalisten tapahtumien havaitsemisen aivomekanismit

Julkaisija Perustieteiden korkeakoulu

Yksikkö Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 183/2014

Tutkimusala Kognitiivinen neurotiede

Käsikirjoituksen pvm 13.08.2014 **Väitöspäivä** 16.01.2015

Julkaisuluvan myöntämispäivä 30.09.2014 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Eräs sosiaalisen kognitiivisen neurotieteen tärkeimmistä tavoitteista on ymmärtää miten aivot mahdollistavat kaikki ne taidot, joita tarvitaan saumattomaan sosiaaliseen vuorovaikutukseen ihmisten kesken. On kuitenkin epäselvää, kuinka aiemmin yksinkertaistetuissa kokeissa havaitut koetulokset yleistyvät monimutkaisiin luonnollisenkaltaisiin tilanteisiin, joissa useita sosiaalisia vihjeitä pitää pystyä seuraamaan yhtä aikaa.

Tässä väitöskirjassa koehenkilöille esitettiin luonnollisen kaltaisia sosiaalisia vuorovaikutustilanteita elokuvien avulla ja tutkitaan aivojärjestelmiä, jotka osallistuvat ärsykkeiden eri piirteiden havainnointiin keskittyen erityisesti elokuvien sosiaaliseen sisältöön. Aivojen toimintaa mitattiin toiminnallisen magneettikuvauksen avulla. Lisäksi osassa kokeista mitattiin silmänliikkeitä ja kerättiin koehenkilöiltä arvioita heidän toiminnastaan. Yhdessä kokeista koehenkilöiden katselutapaa ja elokuvan tulkintaa manipuloitiin ohjeistamalla heitä katsomaan elokuvaa eri näkökulmista. Aivotoiminnan analysointiin käytettiin useita lähestymistapoja: parametrisia malleja ärsykepiirteistä, koehenkilöiden aivoaktivaatioiden välistä korrelaatiota ja riippumattomien komponenttien analyysia sekä verkostoanalyysia.

Ensimmäisessä osatyössä vertailtiin ärsykepiirteiden havaitsemiseen liittyvää aivoaktivaatiota ja aivojen toiminnallisia verkostoja luonnollisen kaltaisen ärsykkeen aikana ja löydettiin verkostoja, jotka osallistuvat puheen, äänen, liikekategorioiden ja alhaisen tason visuaalisten piirteiden analysointiin. Toinen osatyö valottaa useiden eri sosiaalisten piirteiden analysointiin osallistuvien aivoalueiden organisaatiota. Tulokset viittavat siihen, että ylemmän ohimouurteen takaosa on sosiaaliselle havainnoinnille tärkeä aivoalue, joka mahdollisesti yhdistää hyvin monipuolista sosiaalista informaatiota. Kolmannen osatyön tulokset puolestaan tukevat hypoteesia, että ihmisten tulkitessa ulkomaailmaa samalla tavoin heidän aivotoimintansakin muuttuu samankaltaisemmaksi. Yhtenevät tulokset osoittavat, että sekä yksinkertaisten ärsykepiirteiden ja sosiaalisen sisällön analysointiin, että korkean tason perspektiiviin ottamiseen liittyvää aivotoimintaa voidaan tutkia monimutkaisen audiovisuaalisen ärsykkeen aikana.

Avainsanat Sosiaalinen kognitio, perspektiivin ottaminen, aivot, fMRI, ICA, GLM, ISC

ISBN (painettu) 978-952-60-5957-0 **ISBN (pdf)** 978-952-60-5958-7

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2014

Sivumäärä 127 **urn** http://urn.fi/URN:ISBN:978-952-60-5958-7

Preface

The work presented in this thesis was done in the Department of Biomedical Engineering and Computational Science (BECS) in the Aalto University School of Science. I would like to thank Professor Mikko Sams for making the whole endeavor possible by supervising the thesis work and Lauri Nummenmaa and Iiro Jääskeläinen for valuable instructions throughout the project. It has been a pleasure to work with all of you and your enthusiastic attitudes have given great inspiration for pursuing my own scientific endeavors.

I would like to thank the "Brain and Mind" doctoral program for the support making the writing of this thesis possible, and my follow-up group in the doctoral program, Jussi Tohka and Aapo Hyvärinen for valuable advice for finishing the thesis. I would also like to thank the aivoAALTO research project both for financial support and giving the opportunity to perform this exciting and unconventional research. Additional thanks for financial support go to the Academy of Finland (National Centre of Excellence Program 2006–2011).

I wish to thank the pre-examiners of my thesis, Prof. Aapo Hyvärinen and Prof. Christian Keysers, for their insightful comments that helped immensely to improve the final thesis.

This work would not have been possible without the excellent co-authors participating in various stages of the publications. First, thank you to Academician Riitta Hari from the Brain Research Unit of O.V. Lounasmaa Laboratory for all the late-night "red pen"-notes that always came in large portions and still managed to be always insightful and important; despite the time of day.

Many thanks go to Juha Salmitaival and Enrico Glerean for the discussions about the serious and sometimes not-so-serious topics at the office and to everybody else at the Brain and Mind Laboratory for making it a fun and educational place to work and play.

All the fMRI data used in this study were collected at the Advanced Magnetic Imaging Centre of Aalto University School of science. I am very grateful to Marita Kattelus for practical assistance and company during the long and loud measurement sessions and of course Toni Auranen, Simo Vanni and everybody I forgot or don’t even know about who have made it possible to conduct research at the AMI centre.

Thank you to people working at BECS administration, both past and present, for doing all the background work that I know nothing about. It has been a joy working at a department where everything just seems to work, and even when something goes wrong help is nearby and friendly.

Finally, I would like to thank my family for all the love, support, and opportunities you have given throughout my life, my darling Sonja for always trusting and encouraging me, as well as her mother for all her support. Additional “thanks” go to the furry little creatures that have kept waking me up in the middle of most nights during this thesis project, thus making sure I made it to work on time.

Espoo, October 16th, 2014

Juha Lahnakoski

Contents

List of Publications

This thesis comprises of three journal publications, which are referred to by their roman numerals.

I. Lahnakoski JM, Salmi J, Jääskeläinen IP, Lampinen J, Glerean E, Tikka P and Sams M. (2012) Stimulus-Related Independent Component and Voxel-Wise Analysis of Human Brain Activity during Free Viewing of a Feature Film. PLoS ONE 7(4): e35215. doi:10.1371/journal.pone.0035215

II. Lahnakoski JM, Glerean E, Salmi J, Jääskeläinen IP, Sams M, Hari R and Nummenmaa L (2012) Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. Frontiers in Human Neuroscience. 6(233). doi:10.3389/fnhum.2012.00233

III. Lahnakoski JM, Glerean E, Jääskeläinen IP, Salmi J, Hyönä J, Hari R, Sams M and Nummenmaa L (2014) Synchronous brain activity across individuals underlies shared psychological perspectives. NeuroImage 100: 316–324. doi:10.1016/j.neuroimage.2014.06.02

Author's contribution

Study I The candidate analyzed the data, and wrote the manuscript. All of the co-authors gave valuable input during writing of the manuscript.

Study II The candidate gathered and analyzed the data, and wrote the manuscript. Assistance for the data acquisition was received from Lauri Nummenmaa and Marita Kattelus. All of the co-authors gave valuable input during writing of the manuscript.

Study III The candidate designed the data analysis approach, analyzed and gathered the data, and wrote the manuscript. Assistance for the data acquisition was received from Lauri Nummenmaa, Heini Heikkilä and Marita Kattelus. All co-authors gave valuable input during writing of the manuscript.

Abbreviations

AC	Auditory cortex
ACC	Anterior cingulate cortex
aIns	Anterior insula
ASD	Autism spectrum disorder
aSTS	Anterior superior temporal sulcus
dmPFC	Dorsomedial prefrontal cortex
EBA	Extrastriate body area
EPI	Echo planar imaging
eyeISC	Inter-subject correlation of eye gaze locations
FBA	Fusiform body area
FEF	Frontal eye field
FFA	Fusiform face area
FG	Fusiform gyrus
(f)MRI	(Functional) Magnetic resonance imaging
GLM	General linear model
HG	Heschl's gyrus
HRF	Hemodynamic response function
IC(A)	Independent component (analysis)
IPS	Intraparietal sulcus
ISC	Inter-subject correlation
lOC/LOC	Lateral occipital cortex
M1L	Primary motor cortex, lip representation area
MOC	Medial occipital cortex

mSTS	Middle superior temporal sulcus
mPFC	Medial prefrontal cortex
MTG	Middle temporal gyrus
OFA	Occipital face area
OFC	Orbitofrontal cortex
OP	Occipital pole
PCC	Posterior cingulate cortex
Pcu	Precuneus
PHG	Parahippocampal gyrus
PMC	Premotor cortex
poIFG	Inferior frontal gyrus pars opercularis
POS	Parieto-occipital sulcus
PPC	Posterior parietal cortex
pSTS	Posterior superior temporal sulcus
ptIFG	Inferior frontal gyrus pars triangularis
RF	Radio frequency
SMA	Supplementary motor area
sOC	Superior occipital cortex
SPL	Superior parietal lobule
STG	Superior temporal gyrus
TE	Echo time
TOJ	Temporo-occipital junction
TPJ	Temporoparietal junction
TR	Repetition time
V5/MT	Middle temporal visual area
vACC	Ventral anterior cingulate cortex
VTC	Ventral temporal cortex

1. Introduction

1.1 Social cognitive neuroscience

Social interactions are ubiquitous in human societies and understanding how our brain makes sense of the flood of information in social situations is a major goal of social cognitive neuroscience. The current research aims to develop and apply methods that allow us to study the neural processes that enable us to understand other individuals in complex naturalistic conditions. Social cognition covers a wide range of functions that are necessary for successful social interactions. A broad definition of social cognition attributes to it all *"processes that subserve behavior in response to conspecifics"* (Adolphs 1999), and social interactions may even span across different species (Frith and Frith 2012). The studies presented here concentrate on perception and understanding of social events, often dubbed *social perception*. This is in contrast to other skills needed in active participation in social interactions, such as speech production and turn-taking in conversations. However, there is no perfect division of social skills into active and perceptual categories as perception and action are strongly linked in social contexts (Garrod and Pickering 2004).

1.1.1 Neural correlates for perception of social cues

Earlier in the 20th century typical locations of brain lesions causing prosopagnosia ("face blindness") in humans had been described (Meadows 1974), and studies had shown, for example, that face sensitive neurons exist in the monkey temporal cortex (Perret et al., 1982). Some electrophysiological studies in humans had also addressed face perception (see e.g. Lu et al., 1991). However, social cognitive neuroscience only took off as a separate field of study around the turn of the millennium (Singer 2012). This was catalyzed by the first functional neuroimaging findings of elementary brain functions underlying social processing, such as face (Haxby et al., 2000; Kanwisher et al., 1997) and body perception (Downing et al., 2001), and the first reviews of neural correlates of social cognition in humans (see e.g. Adolphs 1999; Allison et al., 2000).

In typical experiments, researchers have used stimuli such as isolated pictures of faces and houses and compared the responses elicited by the two stimulus categories to each other. Such studies demonstrated the existence of face and body sensitive brain regions (Kanwisher et al., 1997; Haxby et al., 2000; Downing et al., 2001) in the occipital and temporal regions of the human brain. Similar experimental setups have been used to present subjects, for example, with biological motion (Servos et al., 2002) using simplified stimuli such as point light displays showing the trajectories of the joints of a moving person with dots of light — a simple stimulus that causes vivid perception of a person moving. Point light displays also have the added benefit that changing the location of the dots largely retains the complexity of the stimulus but eliminates the perception of biological motion. While simple motion usually activates particularly the motion sensitive visual area V5/MT (middle temporal visual area) (Thompson and Parasuraman 2012), biological motion and action perception cause stronger and more wide-spread activity of an extended network of brain areas, often referred to as the "action-observation network" including parietal, temporal and premotor areas (Nummenmaa et al., 2014).

Action observation activates some of the same brain systems that are responsible for action production. These so called "mirror neurons" were first reported single-cell recordings in monkeys demonstrating the existence of cells that fire when the monkey performs an action or sees the experimenter perform the same action (for a review, see Rizzolatti and Craighero 2004). Recently, nearly twenty years after the first reports in monkeys, Mukamel and co-workers (2010) were the first to present direct evidence for the existence of mirror neurons in humans using intracortical electrodes. Their results indicated that cells with mirroring properties may be more wide-spread in the human cortex — including neurons in medial frontal and medial temporal areas — than previously shown in monkeys. Moreover, vicarious neural activity is observed not only for motor actions, but also for viewing others being touched, and observing emotional expressions (Frith 2007; Keysers et al., 2010). Thus, representing the internal states of others could more generally recruit the same neural populations as experiencing the state firsthand. However, far-reaching theories of mirror neurons as the basis of understanding others have also recently been criticized (see e.g. Hickock 2008; Caramazza et al., 2014).

Figure 1 summarizes major areas participating consistently in various subprocesses of social cognition in the human brain based on prior literature. However, the list is not exhaustive, and other parts of the brain may participate in the functions mentioned here. For example, in addition to the areas traditionally linked to processing of emotions there is evidence that areas such as the dorsolateral prefrontal cortex may participate in emotion processing with non-linear response profiles (Viinikainen et al., 2010).

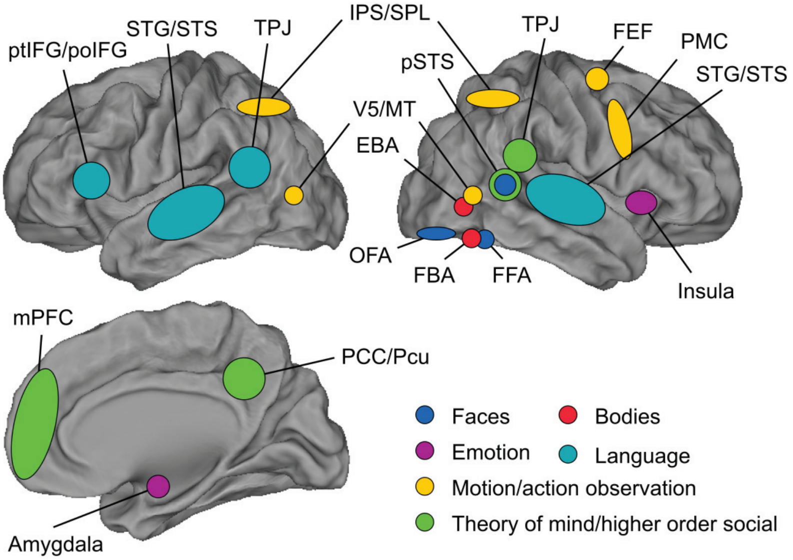

Figure 1. Schematic depiction of the social brain areas. Areas participating in different subprocesses of social cognition are color-coded and approximate locations on the PALS12 atlas template (Van Essen 2005) are shown with the ellipses. Most of the depicted regions are bilateral but some are only shown on one hemisphere for simplicity. Abbreviations: EBA – extrastriate body area, FBA – fusiform body area, FEF – frontal eye field, FFA – fusiform face area, IPS/SPL – intraparietal sulcus/superior parietal lobule, mPFC – medial prefrontal cortex, OFA – occipital face area, PCC/Pcu – posterior cingulate cortex/precuneus, PMC – premotor cortex, pSTS – posterior superior temporal sulcus, ptIFG/poIFG – inferior frontal gyrus pars triangularis/pars opercularis, TPJ – temporoparietal junction, V5/MT – middle temporal visual area.

Partly overlapping regions of the fusiform gyrus, particularly in the right hemisphere, participate in processing faces and bodies (Schwarzlose et al., 2005). In addition to the fusiform gyrus, a more widespread network of brain areas, including inferior occipital, temporal and parietal regions, participate in face processing (Haxby et al., 2000). Additionally, temporo-occipital parts of the extrastriate visual cortex, dubbed extra-striate body area (EBA), take part in the processing of bodies (Peelen and Downing 2007).

Speech processing has traditionally been attributed to left-lateralized areas in the inferior frontal cortex (“Broca’s area”) and temporoparietal junction (“Wernicke’s area”) although the location and functional role of the latter has recently been questioned (DeWitt and Rauschecker 2012). Human voice also activates the superior temporal sulcus bilaterally (Belin et al., 2000). Listening to longer passages of speech tend to activate larger, bilateral brain regions including large parts of the superior temporal gyrus and sulcus, and parts of the precuneus/posterior cingulate cortex, and premotor/prefrontal cortex (Lerner et al., 2011). However, although the activity is bilateral, the activated areas are often more extensive in the left hemisphere.

1.1.2 Affective neuroscience

Although brain basis of emotions have largely been studied separately from brain functions related to social cognition — partly because much of affective neuroscience studies have focused on basic emotions such as fear and happiness (Hamann 2012) — emotions affect multiple aspects of social cognition. Neuroscientific research of emotions has concentrated on the perception of, for example, emotional facial expressions (Morris et al., 1996) and emotionally arousing pictures (Heinzel et al., 2005). While several brain areas have been linked to some aspects of emotional processing, such as activation of the amygdala for fear and the anterior insula for disgust, it is still largely unknown precisely how different emotions are coded in the brain. Moreover, the functional role of the areas that are active during emotional processing may not be limited to emotions. For example, amygdala may play a more general role in associating positive or negative value to stimuli (Frith 2007). Earlier theories have posited that emotions can be divided into discrete categories, which have their distinct neural correlates in the brain. These theories are partly based on the observation that basic emotions, such as fear, anger and sadness, have distinct facial expressions that are shared across cultures (Ekman et al., 1969). Other theories suggest that emotions are based on more general underlying processes or dimensions of emotions, such as valence and arousal (Russell 1980; for a review, see Russel and Barret 1999). More complex constructionist models posit that all emotions are based on several processes, such as physiological state, cognitive appraisal of the situation and bodily state, and memory processes, that together make up the experience of discrete emotions (e.g. Wilson-Mendenhall et al., 2011). Thus, several emotions may employ partly similar brain areas but the pattern of activation over the whole brain could distinguish between different emotional states (for a review, see Hamann 2012).

1.1.3 Theory of mind and perspective taking

Higher-level social cognitive processes depend on several lower-level processes, including perception of bodies, biological motion, facial identity, gaze direction, and speech. We use this information, as well as memories from previous encounters, to predict other persons' goals and future actions. We tend to automatically predict the intentions of others. The tendency to perceive intentionality is so strong that we often spontaneously describe even inanimate objects or moving geometrical shapes as doing things out of their own volition (see e.g. Heider and Simmel 1944). Understanding conspecifics by observation and prior knowledge is at the core of social perception, and has, at least in humans, developed into a "Theory of mind" (Premack and Woodruff 1978).

Theory of mind refers to skills that allow us to infer what other people know or think, and how they are likely to act in a given situation. It is closely related to perspective taking that allows us to take the point of view of another person or a given role. Theory of mind and perspective taking are complex social skills that develop relatively late in life, approximately between 4 and 6 years of age (Blakemore 2008; Frith and Frith 2003), and often still improve in old age when other cognitive skills may already be in decline (Happé et al., 1998; Hari and Kujala 2009). It has also been suggested that understanding others may partly rely on covert mirroring of their actions and emotions as an initial prediction of their intentions that may be built on as more information becomes available (Frith 2007).

Theory of mind has been studied, for example, by asking people to read the mental states of characters in cartoons and stories (Gallagher et al., 2000). It has often been associated with activity in midline brain structures such as medial prefrontal cortex and posterior cingulate cortex, and/or precuneus, which have been proposed to participate in internally-focused social processing (Lieberman 2007) — were it evaluating the internal states of the self or inferring the thoughts of another person. Additionally, the temporoparietal junction (TPJ) and the adjoining posterior superior temporal sulcus (pSTS) have been consistently associated with several higher order social functions (Decety and Lamm 2007; Hein and Knight 2008) including theory of mind. However, while the right TPJ activates during theory of mind tasks it seems not to participate in spatial perspective taking (David et al., 2008) during similar experimental conditions. This suggests that during simple perspective taking tasks people do not need to read the mind of others. There is also evidence that in some conditions people take other people's presence and perspective into account automatically without explicitly engaging their theory of mind (Frith 2012).

Every person has a unique perspective to events unfolding in the outside world. Not only do people have separate physical bodies, leading to different visual perspectives that enable people to see different things from each other, but prior life experiences affect significantly how we interpret incoming information (Pronin et al., 2002). Changing one's perspective to align it with someone else's may even be very difficult. In everyday life people may, for example, perceive events of a football game, such as fairness of the penalties or roughness of the play, quite differently depending on which team they are rooting for (Hastorf and Cantril 1954). However, the current tasks and goals of an individual may also change their perspective toward the outside world in the short term. Here, the perspectives that affect the way information is gathered and processed in task-relevant ways will be referred to as *psychological perspectives* to differentiate them, for example, from the spatial perspectives that affect what is visible to a person from a particular physical location.

Psychological perspective of a person can be manipulated experimentally to influence encoding, comprehension and recall of a stimulus. For example, if the same text is read from two different perspectives interpretation and recall of text content (Anderson and Pichert 1978), visual sampling of the text (Kaakinen et al., 2002) and even the size of the functional visual field (Kaakinen and Hyönä 2014) may change depending on the relevance of the text content for the current perspective. These studies have used generalized perspectives, such as interior designer *vs.* burglar, to guide the attention to different details while the participants read a story describing the interiors of houses. They have shown, for example, that people recall more perspective-relevant *vs.* irrelevant details about the stories they have read, and that the time spent on fixating on parts of the text depends both on the perspective-relevance of the text passage as well as the particular perspective the individuals are currently taking (Kaakinen and Hyönä 2008). However, little is known about the brain mechanisms underlying perspective taking, particularly during perception of naturalistic social events.

1.2 Studies of human brain function using naturalistic stimuli

Traditional neuroscientific research has concentrated on studying brain responses to isolated and often very impoverished stimuli, and tried to piece together how the brain would combine these processes to represent increasingly complex features of the natural world. Through these well-controlled experiments, we have gained important insight into the processing of simple sensory features, and even some higher-level processes. However, recent work has challenged the assumptions that neural responses to, for example, complex speech sounds can be derived from activity elicited by simple synthetic stimuli often used in research (David et al., 2009). Response properties of neurons seem to dynamically adapt to stimuli that are currently behaviorally relevant (Fritz et al., 2007). Moreover, the same brain region may participate in multiple brain networks depending on the current task; anatomically similar areas of the superior temporal sulcus have been implicated in a wide variety of functions including audiovisual integration, theory of mind, and speech perception (Hein and Knight 2008). Such context dependence highlights the need for more complex and naturalistic experimental paradigms to address the question of how the dynamic natural world is represented in the brain.

There are major obstacles in understanding how the brain processes naturalistic scenes because of the complexity of the incoming sensory streams where multiple features overlap and are correlated with each other. However, during the last decade first steps have been taken toward analyzing brain activity during complex stimulation (for a review, see Spiers and Maguire 2007) such as viewing movies (Bartels and Zeki 2004a; Hasson et al., 2004), listening to continuous auditory stimuli such as stories (Hasson et al., 2008; Lerner et al., 2011) or audio dramas (Boldt et al., 2013), playing video games (Kätsyri et al., 2013) or driving virtual car simulations (Calhoun and Pearlson 2012).

First functional brain imaging studies using movie stimuli demonstrated that brain functions related to some features of movies, such as speech and perceived intensity of colors (Bartels and Zeki 2004a) or global and local motion (Bartels et al., 2008) could be mapped during free viewing conditions to brain areas consistent to those previously revealed by simple stimuli. Furthermore, Hasson and co-workers (2004) demonstrated that when subjects are watching the same movie stimulus their brain activity becomes highly similar (temporally correlated across subjects) in large parts of the brain. While the first study of inter-subject correlation (ISC) of brain activity revealed that activity mostly in sensory and associative cortices becomes correlated across participants during movie viewing, subsequent studies revealed that ISCs can also be observed in the prefrontal cortex (Jääskeläinen et al., 2008) when the subjects were first allowed to watch the opening scenes of the movie outside the scanner, presumably getting them more engaged in the movie events. Furthermore, the spatial extent of the correlated areas depends on the type of stimulus: carefully directed movies cause the activity of large portions of the cerebral cortex to become correlated across individuals while undirected videos of real life cause little correlation across viewers beyond the early sensory cortices (Hasson et al., 2010). This raises the intriguing possibility that similar brain activity time courses reflect not only the similarity of the sensory features of the stimuli but also the similarity of the subsequent higher-order processing of the incoming sensory streams.

While ISC is simple to apply and does not require any model of the stimuli or hemodynamic response it alone cannot give direct information about what the brain regions exhibiting ISC are doing during the experiment. However, there are various ways in which such information can be gained, for example, by examining the stimulus content at instances when a given brain region showing high ISC is particularly active (Hasson et al., 2004) or, more recently, comparison of ISC strength across different conditions. For example, researchers have presented with video or audio clips of the same stimulus material shuffled in different time scales (Hasson et al., 2008; Lerner et al., 2011). These experiments have shown that while the ISC in early visual and auditory areas is strong even when the stimulus is played backwards or scrambled in short time scales, areas such as posterior temporal lobe and precuneus require longer continuous passages to show significant ISC. These observations suggest that the latter areas may integrate information over longer temporal windows.

Researchers have also started to address how ISC changes in time. A recent study showed that the strength of ISC within a group of participants followed the experienced emotional valence and arousal (Nummenmaa et al., 2012) during viewing of emotional videos. Using a related method of inter-subject phase synchrony (Glerean et al., 2012) another recent study revealed that activity in the action observation network becomes synchronized across participants when they mentally simulate boxers' actions depicted in video clips (Nummenmaa et al., 2014).

In addition to stimulus-model based analysis methods and ISC, several studies have used independent component analysis (ICA) to study the brain networks activating during stimuli (Bartels and Zeki 2004b; Bartels and Zeki 2005; Malinen et al., 2007) and particularly during resting state in the absence of any particular task (Kiviniemi et al., 2003; Damoiseaux et al., 2006; Biswal et al., 2010). Unlike ISC or model-based methods, ICA reveals brain regions that covary within the brains of single subjects. It does not necessarily require that the brains of the subjects are activating similarly as long as the spatial patterns are similar. This allows the mapping of "resting state" networks that are largely similar to those observed during movie viewing (Lahnakoski 2010) as well as other tasks (Smith et al., 2009). Similar networks and large-scale properties of brain network architecture have also been studied using the tools of network analysis (Sporns 2011).

Generally, methods based on covariation of activity across brain areas, such as ICA and network analysis methods, reveal converging spatial patterns of functional brain networks. Typically, ICA is used to reveal a few tens of networks that are thought to reflect the underlying functional architecture of the brain (Damoiseaux et al., 2006; Biswal et al., 2010) such as the sensorimotor, visual, auditory and the ”default mode” networks. First observation of such covarying functional network was made in the motor system (Biswal et al., 1995) using a seed-voxel based approach where the time course of one voxel was correlated against all other voxels in the brain. Similar approach can also be extended to large-scale brain networks by calculating the temporal correlation between all pairs of regions. Various network analysis methods can be used to reveal more diverse information about the underlying brain networks at different scales (Sporns 2011). Calculating a network representation of the brain also allows direct comparison of various network measures and, for example, personality traits or developmental abnormalities of individuals. In contrast, because many ICA algorithms are stochastic the results may depend on the initialization and sampling of the data (Himberg et al., 2004). This may add noise to the results thus complicating comparisons between datasets.

Traditional experiments employing simplified stimuli have usually used linear modeling methods to analyze brain activity during stimulation or tasks. However, there is no single approach that can address the multitude of possible research questions during naturalistic stimulation experiments. Thus, additional work is still required to find the most appropriate analysis approaches and experimental designs to test specific hypotheses in future experiments employing naturalistic stimulation (Hasson and Honey 2012). The studies presented here assess the applicability of stimulus modeling, functional connectivity, ICA, and ISC in analysis of brain imaging data gathered during complex stimulation and tasks. The studies demonstrate how various brain processes can be studied in naturalistic stimulus conditions. The results reveal brain areas that participate in processing of both simple stimulus features and social stimulus content, and areas that enable higher-order perspective taking during naturalistic stimulus conditions.

2. Methods

2.1 Magnetic resonance imaging

Magnetic resonance imaging (MRI) has quickly become a widely used tool in brain research after its first clinical applications for studying anatomical brain images in the 1980s and the rapid development of imaging sequences in the 1990s allowing different contrasts between brain tissues (Mandeville and Rosen 2002). The first 1-dimensional MR images were produced already in the beginning of 1950s (Carr 2004). However, equipment and mathematical techniques for practical 2- and 3-dimensional imaging were developed in 1970s (see e.g. Lauterbur 1973) allowing the first images of live animals (Lauterbur 1974) and finally leading to MR imaging of humans toward the end of the decade (Damadian et al., 1977).

The basis of the MRI signal is an intrinsic property of elementary particles called spin. Unpaired spins in matter have a magnetic moment, which tends to align to external magnetic fields (**Figure 2**). Because the nuclei of hydrogen atoms, the most abundant atoms in the human body, have such unpaired spins, magnetic fields can be used to image the soft tissue in the body (Mandeville and Rosen 2002). Other nuclei with unpaired spins can also be imaged with MRI (Huettel et al., 2004:pp. 49–73) but this chapter will focus on hydrogen atoms.

In the absence of an external magnetic field the spins of the hydrogen protons can be oriented in any direction (**Figure 2a**). However, when a strong magnetic field (usually 1.5 or 3 T in current commercial devices) is introduced in the bore of the MRI device, the spins of the hydrogen nuclei in the imaged tissue tend to align either parallel or anti-parallel to the magnetic field (**Figure 2b**). This strong magnetic field is referred to as B_0 (Huettel et al., 2004:pp. 49–73) and it is desirable that it has a constant strength in the entire tissue being imaged.

In addition to aligning along the direction of the external field, the magnetic moments precess around the direction of the field with an angular frequency, called Larmor frequency. The angular frequency is dependent on the strength of the field and is denoted by ω (**Figure 3**). This frequency dependence on the magnetic field is exploited in MRI by introducing smaller magnetic fields, called gradient fields, the strength of which has a maximally linear gradient in a given direction. Thus, the gradient fields cause the magnetic moments to precess around the field B_0 with slightly different angular frequency depending on the location within the field (Huettel et al., 2004:pp. 49–73).

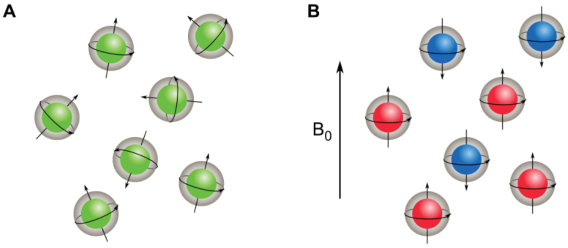

Figure 2: a) Spins of hydrogen nuclei in the absence of a magnetic field may have any orientation. b) In external magnetic field spins arrange so that they are either parallel (red) or anti-parallel (blue) to the external field.

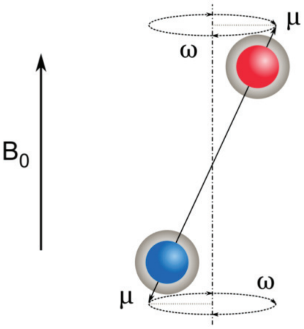

Figure 3: The magnetic moments (μ) of hydrogen protons in a magnetic field (B_0) precess around the direction of the field at Larmor frequency (ω) that is dependent on the magnitude of the field.

Because the characteristic frequency of the precession is dependent on the field strength and gradient fields are used to cause the magnetic field to have a known spatial distribution, properly designed radio-frequency (RF) pulse sequences can be used to deposit energy selectively into a given spatial location. The RF-pulses cause the axis of precession to turn in relation to the external magnetic field. Once the RF pulse ends the spins return to their natural state precessing around the external field emitting the deposited energy as a nuclear magnetic resonance signal in a process called relaxation (Huettel et al., 2004:pp. 49–73).

Relaxation happens at different rates in the longitudinal and transverse directions in relation to the field B_0. Longitudinal relaxation is often referred to as T_1 and transverse as T_2 relaxation according to the time constants describing the relaxation times in the differential equations describing the relaxation process. Critically, the relaxation takes place at different rates in different tissues. This enables the tissues to be separated in the acquired images (Huettel et al., 2004:pp. 99–126) by selecting appropriate imaging parameters. Main imaging parameters defining the contrast between different tissues are the repetition time (TR) of the RF-pulses that defines the time between two successive excitations, and echo time (TE), which defines the time between the excitatory RF-pulse and data acquisition. By proper selection of these parameters the desired contrast between tissues can be maximized. **Figure 4** schematically depicts typical time courses of relaxation in two hypothetical tissue types and shows how the T_1 and T_2 contrasts between these tissues vary as a function of time.

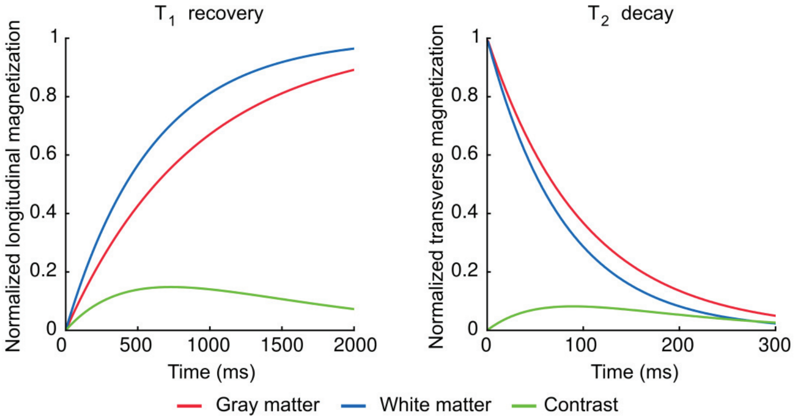

Figure 4: Schematic time courses of T_1 recovery and T_2 decay in two theoretical tissues with typical time constants for gray matter (red) and white matter (blue) after an excitation pulse at time t=0. By properly selecting the parameters TR and TE the T_1 or T_2 contrast between the tissues can be maximized (green curve).

2.1.1 Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is a set of techniques developed to extend MRI sequences to allow the study of hemodynamic brain functions. These techniques include quicker image acquisition schemes, and imaging parameters optimized for detecting small inhomogeneities in magnetization of brain tissue as a function of blood flow and oxygen content. The development of these techniques was initiated by the findings of Ogawa and co-workers (1990) that discovered that measuring the inhomogeneity of the magnetic field inside the brain gives information about local blood oxygenation. The next year, Belliveau and co-workers (1991) demonstrated that visual stimulation increased the local blood volume in the visual cortex, which could be measured using MRI equipment using intravenously administered paramagnetic contrast agent. Subsequently, both Bandettini (1992) and Kwong (1992) independently demonstrated with their co-workers that the blood oxygenation level-dependent (BOLD) signal can be used to measure brain activity completely non-invasively.

The local inhomogeneities in the MRI signal yield information of brain function because oxygenated hemoglobin (i.e. hemoglobin bound to oxygen) has different magnetic properties than deoxygenated hemoglobin. Oxygenated hemoglobin is diamagnetic, meaning that it has no unpaired electrons and zero magnetic moment. In contrast, deoxygenated hemoglobin is paramagnetic meaning that it has unpaired electrons and nonzero magnetic moment causing it to have a positive magnetic susceptibility (i.e. it is magnetized in a magnetic field). This causes local distortions in the magnetic field, and reduces the relative signal strength received from the nearby brain areas. During neural activity, excess of fresh oxygenated blood is rushed to the active sites in the brain causing a net increase in the oxygenated hemoglobin. Thus, the signal from the active brain areas is increased (Huettel et al., 2004:pp. 159–184) causing the observed BOLD effect.

The transverse, T_2 relaxation mentioned in the previous chapter happens through two different processes: pure T_2 relaxation caused by the loss of phase coherence between the precessing protons, and T_2^+ relaxation caused by inhomogeneities in the magnetic field. These two types of relaxation combine to give the total transverse relaxation time constant termed T_2^* (Huettel et al., 2004:pp. 49–73). Because the total transverse relaxation is sensitive to the inhomogeneity of the magnetic field, and the local inhomogeneity of the magnetic field depends on the relative concentrations of oxygenated and deoxygenated hemoglobin, the imaging parameters in functional MRI are typically optimized for T_2^* contrast.

Because fMRI measures the brain at the level of blood flow, rather than neuronal activity, it gives an indirect measure of brain function. The signal changes are relatively slow in comparison with the underlying electrical signaling. Blood oxygenation changes evolve over several seconds after the onset of the electrical activity and reach a peak typically between 4 and 6 seconds after the neural activation. After the peak, the blood oxygen content starts to decrease usually going slightly below the baseline after approximately 10 seconds after the onset, and finally returning to the baseline around 30 seconds after the initial neuronal activation (Huettel et al., 2004:pp. 159–184). In data analysis, the hemodynamic response function (HRF) is often assumed to be a linear response to neural activity, and a simple convolution is used to model the hypothetical hemodynamic response to the experimental conditions (**Figure 5**). However, there are situations when this assumption is violated. For example, there are cases when a hemodynamic response may be triggered without underlying neural activity, for example, when an excitation is anticipated due to repetitive stimulation (Sirotin and Das 2009). Although the linear relationship of hemodynamic and neural activity is a simplification, and the precise mechanism of the neurovascular coupling responsible for the hemodynamic response is still largely unresolved, this model has proven useful in practice. In practice, the HRF is often approximated as a linear combination of two gamma functions.

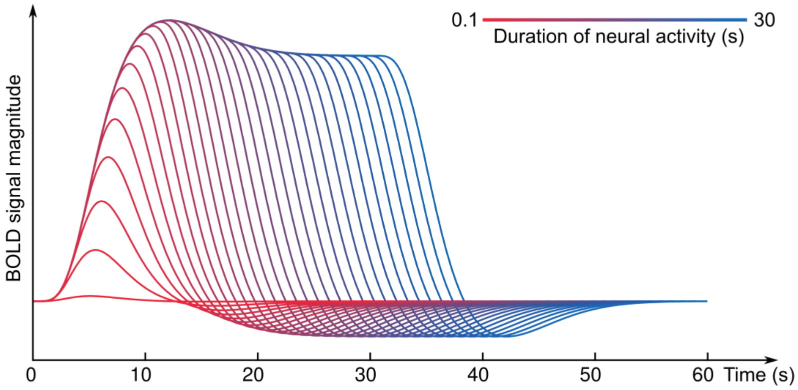

Figure 5. Stereotyped hemodynamic responses caused by underlying neural activity of durations between 0.1 to 30 seconds at 1 s intervals. Typically the HRF is modeled as a linear, time-invariant system and a simple convolution of the stimulus and the hemodynamic impulse response function is used to estimate the shape of the response.

Typical anatomical MRI scans of the whole brain take several minutes. Thus, quicker image acquisition schemes had to be developed to enable higher data acquisition rates that allow the blood oxygenation changes to be sampled in time. Most common acquisition approach for fMRI is the echo planar imaging (EPI) scheme where an entire slice of data is recorded after a single excitation pulse. Gradient fields are then switched systematically to sample the frequency and phase encoding direction of the slice (Mansfield 1977). In contrast, in traditional MRI, a separate excitation pulse precedes the acquisition of every voxel. Additionally, a lower resolution is used in fMRI compared to conventional MRI to further reduce the acquisition time.

These techniques allow imaging of the entire brain in approximately 2 seconds (Mandeville and Rosen 2002) making the sampling of the relatively slow hemodynamic changes possible.

2.2 Eye tracking

Gaze direction is an important cue of the focus of attention and intentions of an individual in everyday life. Tracking of gaze direction has been used as a tool in experimental psychology for decades to study how people gather information from text and images (Henderson 2003), and more recently videos (e.g. Wang et al. 2012)

Eye movements consist of fixations on spatial locations, during which the eyes stay still on a target and visual information is encoded, and saccades, or rapid eye movements from one location to another, during which visual encoding is suppressed. These are typically analyzed by calculating the durations of the fixations on given targets, which is usually strongly related to the current focus of attention of the subject, and amplitudes of the saccades, which may relate to the current attentional mode (e.g. focal vs. ambient processing; Pannasch and Velichkovsky 2009). In addition smooth pursuit eye movements are employed when a person follows a moving object with their eyes (Henderson 2003).

One of the pioneers of using gaze tracking was the Russian physiologist Alfred Yarbus, who employed contact lenses equipped with small mirrors to enable the online tracking of gaze direction while his subjects viewed photos (Yarbus 1967). He demonstrated, for example, that the eyes, which provide the most salient social cues in the face, are fixated disproportionately more than other parts of the face. However, while viewing videos of people speaking, this tendency seems to be replaced by more dynamic patterns of fixations oriented to the locations that are most informative at any given time (Võ et al., 2012).

While the classical studies of eye gaze required somewhat invasive methods to gather accurate data on the direction of eye gaze, the current methods are completely noninvasive. However, most of them still require the viewer to stay relatively still in relation to the equipment used to record the eye movements. Most common methods use algorithms based on corneal reflections (Horimoto and Mimica 2004) recorded using video cameras from a known position to calculate the location of the gaze at each moment of time. The sampling rates can range from tens of hertz to 1000 Hz or more in commercial eye tracking systems.

During complex dynamic stimulation, region-based analysis of fixations is more demanding. Therefore, in addition to traditional analysis of fixations and saccades, recent studies have also analyzed the similarity of the eye movement time courses between individuals. Studies have compared eye movements across healthy human subjects (Wang et al., 2012), between typically developed humans and individuals with autism spectrum disorder (ASD) (Hasson et al., 2009), and even across humans and non-human primates (Shepherd et al., 2010). These studies demonstrated, respectively, that the reliability of gaze across participants increases the longer the stimulus clips are, that the eye movement patterns of individuals with ASD are less reliable than those of control subjects, and that there are striking similarities in the fixation patterns of monkeys and humans, although there are differences as well. Similarly to the inter-subject correlation of brain activity, which will be introduced in chapter 2.3.2, the analysis of the reliability of eye movements has the benefit that there is no need for an explicit model of where subjects should or should not look. This can be beneficial, particularly when comparing the gaze between groups, or when the behavioral effects of a task are previously unknown.

2.3 fMRI signal analysis methods

There are two major classes of analysis methods employed in this thesis: model based and data driven. Traditionally, model based approaches, such as the general linear model (GLM), have been used to compare the brain activity during a task to a baseline condition. In contrast, data driven approaches, as the name implies, do not require any instartiation about the external stimulus. However, there are important differences between data driven methods. Blind source separation techniques, such as ICA, use computational means to find networks of brain areas co-activating in time. Similar co-activations can also be studied with network analysis methods, which can be used to study brain networks at various scales. These methods can often be used without any external source of synchrony, such as auditory or visual stimulation. In contrast, ISC reveals the brain areas that share a temporal fingerprint across subjects. Therefore, ISC analysis requires that the subjects are exposed to an external source of synchronization, such as a movie, but does not need a model of the experiment.

Prior to analysis, several preprocessing steps are typically performed. Temporal filtering is performed to remove slow signal drifts, and possibly other sources of noise. Motion correction is used to transform the brain into the same orientation in all acquired images. Spatial smoothing is done both to average out uncorrelated noise and to increase the overlap of activated areas across participants despite the anatomical differences between their brains. Finally, the subjects' brains are transformed into a standard space to allow comparison of analysis results across individuals.

2.3.1 General linear modeling

General linear modeling (GLM) has long been the standard way of analyzing fMRI data. In GLM analysis the data is explained by a weighted sum of explanatory variables and additive noise as follows:

$$Y_j = X_{j1}\beta_1 + \cdots + X_{jn}\beta_n + \cdots + X_{jN}\beta_N + \varepsilon_j$$

The β_n are the weights of each explanatory variable x_{jn}, Y_j are the observations and ε_j are the noise terms that are assumed to be independent and identically distributed Gaussian random variables with zero mean and variance σ^2 (Kiebel and Holmes 2007).

Typically the experimental conditions have been modeled as box car functions where a condition is either on or off. These conditions are represented in a so-called design matrix that contains the explanatory variables (x_{jn}) in its columns. A linear sum of the columns is then fitted to the observations in the least squares sense to find the optimal column vector containing the weights (β_n) of each explanatory variable, and the model fit is estimated (Kiebel and Holmes 2007).

More recently parametric modulations of the modeled experimental conditions have been introduced where the activity level of a given brain region may depend (linearly or non-linearly) on, for example, the loudness of an auditory stimulus, or subject's performance in a given task (Friston et al., 2007). GLM has been implemented in several software packages for fMRI analysis, such as Statistical Parametric Mapping (SPM, http://www.fil.ion.ucl.ac.uk/spm/), and FMRIB Software Library (FSL, http://fsl.fmrib.ox.ac.uk/fsl/).

2.3.2 Inter-subject correlation

Inter-subject correlation approach for fMRI analysis was first introduced by Hasson and co-workers (2004) in an experiment where five people watched a movie during fMRI recording. They demonstrated that a large part of the brain functioned similarly across subjects during movie viewing using an elegantly simple approach where they calculated the pairwise Pearson's correlation coefficients of voxel time series across subjects at each location in the brain. When the correlation coefficients were averaged across the subject pairs they showed that statistically significant correlations were seen particularly in the posterior parts of the brain. The ISC approach is demonstrated in **Figure 6**.

The power of ISC approach in naturalistic stimulation experiments is that there is no need for creating an *a priori* model of the experiment. In fact it would be impossible to exhaustively model all the content in a movie. It has been demonstrated (Pajula et al., 2012) that despite the lack of knowledge of external stimuli ISC produces highly similar results as a traditional GLM approach in simplified experimental conditions. However, connecting the ISC map to the external stimulus causing it can be difficult if ISC is calculated over a long and complex stimulus such as an entire movie. One option to gain more specific knowledge of what causes the observed ISCs is to calculate the measure in shorter time windows. However, reliable calculation of correlation coefficients requires several time samples (Fisher 1921). Thus, in a signal as slow as fMRI there could be several stimulus features that could be responsible for the observed synchronization of brain activity even if a sliding temporal window was used. To alleviate the problem, differences in observed ISCs caused by experimental manipulations of the stimulus content have been used to elucidate principles of brain functions without the need for a traditional stimulus model (Hasson et al., 2008; Lerner et al., 2011). ISC method has also been extended with non-parametric statistical comparison of ISC strength across different frequency bands (Kauppi et al., 2010). Similar statistical tests can also be applied to across conditions testing to gain knowledge of significance of observed changes in ISCs, as has been done in publication III of the present work. Recently, these comparisons have also been implemented in the ISC toolbox (https://code.google.com/p/isc-toolbox/ (Kauppi et al., 2014)).

2.3.3 Independent component analysis

Independent component analysis (ICA) is a blind source separation technique that has been used to find statistically independent signal sources particularly in the analysis of resting state fMRI data (Kiviniemi et al., 2003; Damoiseaux et al., 2006; Biswal et al., 2010), and denoising of the fMRI signal by removing artifacts (Tohka et al., 2008). The networks revealed by ICA during rest are also remarkably similar during tasks (Smith et al., 2009; Lahnakoski 2010).

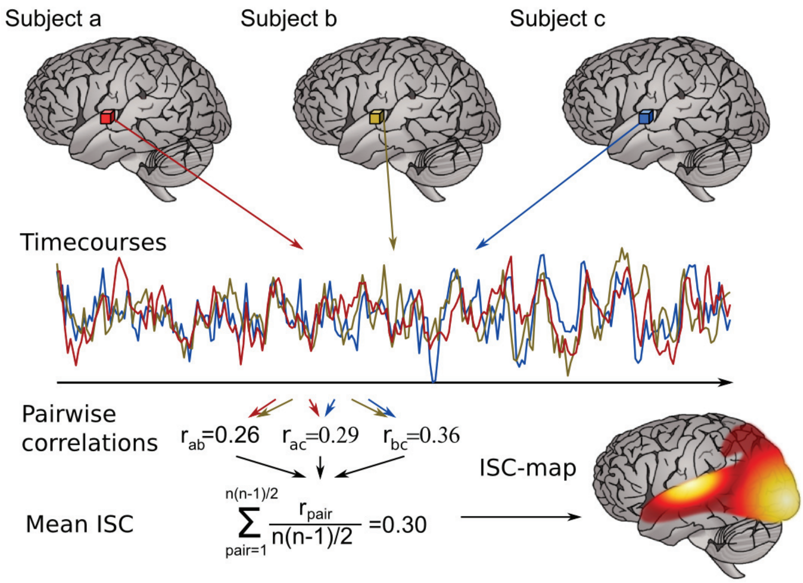

Figure 6: Schematic depiction of inter-subject correlation method. The pairwise correlations of voxel time series are calculated for the same voxel across subjects. The average inter-subject correlation is then calculated over the pairs. The procedure is repeated for all voxels and a statistical threshold is calculated to reveal which areas activate in a similar manner across subjects. The plots and the r-values in the figure are based on actual fMRI data from study I.

The assumption underlying the ICA model is that the observed signals are unknown mixtures of relatively few underlying latent variables. The goal of ICA is to estimate both the underlying independent sources and their relative mixing weights from the observed data (Hyvärinen and Oja 2000). **Figure 7** illustrates this by an example where five different mixes of two images are subjected to ICA. The estimated independent components in this noise-free example are very similar to the original images, but slight mixing of the original images remains in one of the source estimates.

The ICA mixing model can be expressed formally by the equation

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S},$$

where $\boldsymbol{X}$ is the matrix of observed signals, $\boldsymbol{A}$ is the matrix of weights of the source signals, and $\boldsymbol{S}$ is the matrix of source signals to be estimated (Hyvärinen and Oja 2000). Thus, the ICA model is very similar to the GLM but, critically, neither the source signals nor their weights are predefined and both have to be estimated from the data.

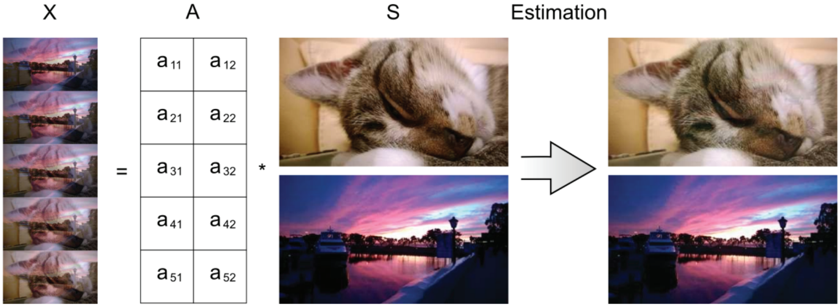

Figure 7: Example of ICA in practice. Five different mixed images (X) of the original source images (S) are observed, where the mixing of the images is defined through the mixing matrix (A). ICA estimates the original images and mixing weights from the mixed images. Estimated images resemble the original images, although some mixing remains (silhouettes of the houses on the cheek of the kitten). The images were reshaped into vectors prior to estimation and FastICA algorithm with gauss nonlinearity and symmetric estimation approach was used for the separation.

The goal of ICA is to reveal the underlying source signals with as few assumptions as possible. In fact, it is often enough to assume that the sources are statistically independent and non-Gaussian (Hyvärinen and Oja 2000). Several different approaches can be used to achieve independence of source estimates. Typically independence is achieved by minimizing (or maximizing) some objective function, such as negentropy of source estimates (Hyvärinen 1999) or joint entropy of the outputs of the nonlinear neural network used for the separation (Bell and Sejnowski 1995), as used by the most prevalent ICA algorithms used in fMRI studies, FastICA, and Infomax, respectively. Both of these methods have been shown to reveal similar estimates of the underlying brain networks, and the results are consistent across repetitions (Correa et al., 2007) despite the stochastic nature of the algorithms.

The original four-dimensional single-subject fMRI data is typically converted into a two dimensional matrix with temporal and spatial dimensions before it is subjected to ICA. ICA has also been extended to group studies either by concatenating the data of the subjects or explicitly extending the data by adding a subject/session dimension to the analysis (Beckmann and Smith 2004). ICA has been applied in both spatial and temporal dimensions (Calhoun et al., 2001). Spatial ICA assumes that the data is composed of independent spatial patterns, or functional brain networks. In contrast, temporal ICA assumes the components are independent in time. Although spatial ICA has been more common in fMRI studies, temporal ICA has also been used, for example, to reveal subdivisions of the auditory cortex (Seifritz et al., 2002) showing transient and sustained responses to stimuli.

Because both underlying signals and their weights are estimated at the same time in the ICA model, the signals can only be estimated up to a scalar multiplier. Thus, the estimates are often standardized and the sign of the components is inferred from results. For the same reason, the order of the signals is arbitrary, because the order of the summation does not change the result of the linear sum (Hyvärinen and Oja 2000). Therefore, the selection of relevant components for further study requires some additional information either from the person performing the analysis, or from the properties of the components themselves.

The number of independent components needed to explain the data in fMRI experiments is an open question. Automatic ways to select the dimensionality of the data have been implemented, for example, in toolboxes such as FSL and Group ICA fMRI Toolbox (GIFT, http://mialab.mrn.org/software/gift/index.html), but they have also received criticism (Li et al., 2007; Cordes and Nandy 2006). In practice, a predefined number of ICs is often chosen based on prior literature (see e.g. Biswal et al., 2010). Some studies have also estimated the stability of the results and selected the appropriate number of ICs based on their stability (see e.g. Pamilo et al., 2012). Number of components estimated varies from study to study, but often a few dozen are calculated. However, the typical networks revealed by ICA of fMRI data seem relatively robust to different numbers of components selected. Some of the networks may be split into smaller subunits when a larger number of ICs are estimated (Lahnakoski 2010; Pamilo et al., 2012), but further work should address what is the most informative parcellation of the brain for particular applications.

2.3.4 Network analysis

Network analysis consists of a wide range of methods that can be applied to practically any data that can be represented as a network (for a review, see Newman 2003). A network consists of a set of nodes that are connected by edges. For example, the nodes could depict people and edges could represent social ties, or nodes could be brain regions and edges could be anatomical or functional connections between them. This type of description can give important information on the information flow between brain regions.

In network science systems can be viewed at several levels, starting from global properties of the network of interest — such as average path length, describing the number of links one must on average traverse to reach any node from any other node — to local properties — such as degree of a node, that is the number of neighbors, or links a node has. In between these extremes one can analyze, for example, the sub-networks, or community structure of a network, that may reveal functional modules of a brain network, or social groups within a larger population (for a primer on brain network analysis, see Sporns 2011).

Community structure of a network can be studied with various methods. One of the most successful methods in network science is the Infomap algorithm that uses random walks as a proxy of information flow within a given network (Rosvall and Bergstrom 2007). The method parcellates the network into sub-graphs and describes the paths of the network with the aid of these communities. The aim of this process is to find the optimal parcellation that retains as much information about the original network while minimizing the description length of the community structure — that is listing the nodes and the links in and between the communities with as few bits as possible, while concurrently minimizing the additional information needed to describe the network exactly. Other approaches to parcellation of networks include, for example, several clustering algorithms. Generally, the sub-networks are spatially similarly distributed as the independent components revealed by ICA (see e.g. van den Heuvel et al., 2008). However, whereas spatial independent components have only one value for each voxel that describes the weight of that voxel in the time course of the independent component, functional networks are desribed by the strengths of connections (e.g. temporal correlation) between all pairs of nodes. Thus, a network representation is a more rich description of the covariance structure. However, as the number of nodes and edges increases, the interpretation and visualization of brain networks becomes challenging. In this work, the functional sub-networks of the human brain are studied with both ICA and the Infomap community detection algorithm.

2.4 Stimulus modeling

Modeling the experimental conditions and stimuli is a major challenge in naturalistic neuroscience experiments. It is obvious that it is impossible to exhaustively model every possible perception a viewer may have during movie viewing, even if the color and brightness of each pixel are exactly defined at each time point by the video file. People rarely experience movies as series of color changing dots but rather focus on the people and objects depicted in the images. Similarly, low-level auditory features do not explain the entire experience while listening to the movie soundtrack containing complex stimuli such as speech and music.

In visual domain, various low-level visual features, such as color, luminosity, orientation and different spatial frequencies (Itti and Koch 2001) describing the visual salience in images, have been used to quantify pictures and videos in modeling of visual attention. Researchers have also used computational models based on knowledge on the organization of the visual cortex to classify natural pictures based on brain activity (Kay et al., 2008). In computer vision, methods have been devised to identify objects from images using, for example, local scale invariant features (Lowe 1999), or distributions of a set of basis functions, such as Haar wavelets, to describe edges in images (Papageorgiou and Poggio 2000).

In auditory domain various auditory features, such as loudness, or spectral characteristics (Alluri et al., 2012) can be extracted from sound files automatically. Loudness is often estimated as the root-mean-square (RMS) energy of the audio waveform. Spectral characteristics employed in the current work include the spectral entropy and spectral spread. Spectral entropy characterizes the randomness of the spectrum estimated as the Shannon entropy of the spectrum. Spectral spread is the standard deviation of the spectrum. Additionally the zero crossing rate of the audio waveform, which yields a simple estimate of the noisiness of the audio signal, is used in study I of the current work. The calculation of these and other auditory features are implemented in the MIRtoolbox (Lartillot and Toiviainen 2007).

Relatively few studies have tried to model dynamic audiovisual naturalistic stimuli. In a landmark study, Bartels and Zeki (2004a) showed that subjects' hemodynamic activity in the sensory and associative cortices followed their self-reported intensity of color, faces, language, and human bodies in an action movie that were rated by the subjects over time. More recently, large collections of hierarchical semantic categories have been used to characterize video clips, and analyzed using regularized linear regression analysis (Huth et al., 2012). However, the reliability of these category representations across individual is still relatively unexplored. Although some more complex features of interest, such as faces, can be recognized from images automatically, higher-level descriptions of the stimulus material often need to be manually evaluated. In current work, manual, semi-automatic, and computational methods are used to characterize the content of short video clips and longer movies, and map the brain regions participating in their processing using the signal analysis methods described above.

3. Goals of the current research

The studies presented here aim to bridge the gap between traditional experiments using simple well-controlled stimuli and the complexity of real-world social situations. Overall goal of these studies is to develop and validate approaches to analyze functional neuroimaging data gathered during perception of naturalistic social events and apply these approaches to further the understanding of the brain basis of social perception. The studies further aim to assess the reliability of brain activity related to social cognition across subjects.

Study I tested the hypothesis that through careful modeling of complex dynamic stimuli we could map the brain regions participating in their processing. Moreover, the model driven analyses were expected to give both converging and complementary information about functional architecture of the brain compared to ICA that was used to reveal functional brain networks. Particularly, we expected that with careful stimulus modeling, the functions of the networks extracted by ICA could be revealed by comparing their activity timecourses to those of the stimulus features. We further expected that the networks could possibly be further divided into more specialized sub-units using voxel-wise analysis.

Study II aimed to reveal the organization of brain regions participating in processing multiple types of social content through stimulus modeling methods validated in Study I. A particular aim was to test whether a collection of short videos could be used as an efficient independent localizer of social brain regions.

Study III tested the hypothesis that when people take a similar perspective toward external events, their brain activity in areas participating in the task of perspective taking becomes synchronized across participants. Thus, similarity of brain activity between individuals could be associated with similarity of interpretation and, ultimately, mutual understanding.

4. Summaries of the studies

4.1 Data acquisition and preprocessing

4.1.1 Magnetic resonance imaging

All fMRI data were gathered at the Advanced Magnetic Imaging (AMI) centre of Aalto University. Data were gathered with a General Electric Signa 3-tesla MRI scanner (GE Healthcare Ltd., Chalfront St Giles, UK) except the data for experiment 2 of study III, which were gathered with Siemens MAGNETOM Skyra 3T MRI scanner (Siemens Healthcare, Erlangen, Germany).

All data were gathered with EPI pulse sequence sensitive to BOLD contrast with a 64x64 matrix with field of view of 230 mm. Slice thickness was 3–4 mm, TR ranged from 2000 to 2112 ms, and TE from 30 to 32 ms.

4.1.2 Preprocessing of fMRI data

Standard preprocessing steps including removal of initial volumes to allow for the stabilization of magnetization, temporal high-pass filtering, motion correction, spatial normalization to MNI standard space (Montreal Neurological Institute) through individual anatomical images and spatial smoothing was performed using FSL software.

4.1.3 Eye gaze recording

Eye gaze was recorded in Study III with an EyeLink 1000 eye tracker (SR Research, Mississauga, Ontario, Canada; sampling rate 1000 Hz) both inside and outside of the MRI scanner. A nine-point calibration and validation was completed prior to the experiment. Saccade detection was performed using a velocity threshold of 30°/s and an acceleration threshold of $4000°/s^2$.

4.2 Study I: Stimulus-Related Independent Component and Voxel-Wise Analysis of Human Brain Activity during Free Viewing of a Feature Film

4.2.1 Aim of the study

The study aimed to reveal how low- and higher-level auditory and visual stimulus features activate brain regions and networks in naturalistic conditions. The annotated stimulus features were fitted to both voxel-wise data and brain network time courses extracted by independent component analysis (ICA). Finally, the results of the two approaches were compared with each other.

4.2.2 Materials and methods

Ten healthy volunteers viewed a re-edited 23-minute version of the feature film "The Match Factory Girl" (Dir. Aki Kaurismäki) during fMRI acquisition. Several stimulus features were modeled to allow the mapping of the brain regions participating in their processing. The features included manual annotations of speech, lead and background singing, and instrumental music. Zero crossing rate, spectral spread, spectral entropy and root mean square energy of the sound track were computed from the audio waveform. Visual motion of the heads, hands, bodies, and inferred motion of visually occluded body parts, and mechanical motion was manually estimated. Prominence of sharp contrast edges in the images was estimated using spatial high-pass filtering. The stimulus models are described in **Figure 8**. In addition to the manual annotation of visual motion, semi-automatic motion capture software was used to validate the ratings of motion of body parts (heads, bodies and hands), and non-biological objects.

The analysis of the data was two-fold: the functionally connected networks of the brain were estimated using ICA and, additionally, the stimulus models were used to predict the activity time courses of both functional networks, and individual voxels within the brain. Inter-subject correlations of IC time series was used to select stimulus-related ICs of interest for further analysis. Finally, the results obtained using these two methods were compared to reveal the similarities and differences between the approaches.

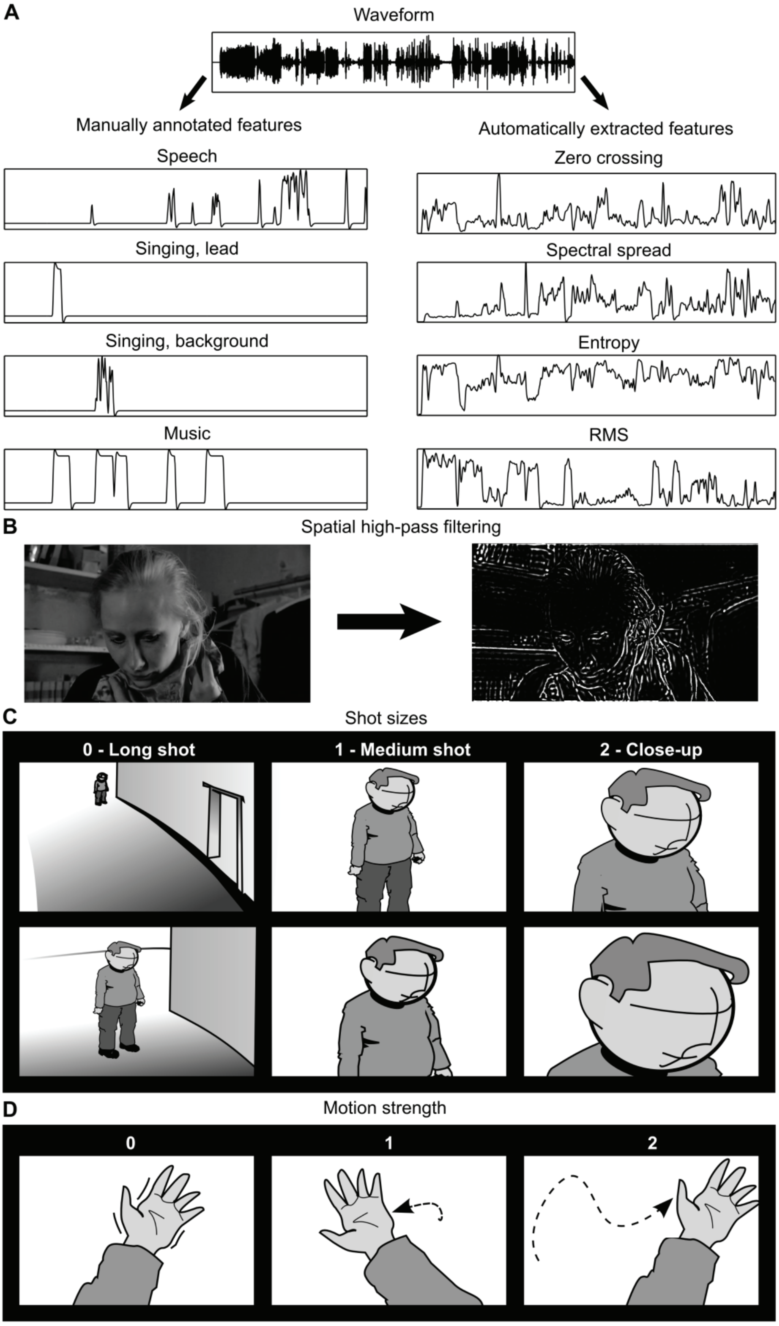

Figure 8. Stimulus models used in Study I. A: Sound was annotated manually for speech, lead and background singing, and instrumental music. Low-level features (zero-crossing rate, spectral spread, entropy and RMS energy) were extracted computationally. **B:** High-frequency spatial content from the movie was extracted using discrete cosine transform. The size (**C**) and motion strength (**D**) of tracked objects in images were graded on a three point scale (0–2), and final score for the motion of specific body part or object was calculated as the sum of the motion and size for those time points where motion was present.

4.2.3 Results

The results of the ICA analysis revealed networks of brain areas that are very similar to the intrinsic functional networks revealed in several resting state studies (Damoiseaux et al., 2006; Lahnakoski 2010; Biswal et al., 2010). The stimulus model based analysis of the brain activity revealed that areas included in a particular IC had distinct sensitivity profiles to stimulus features compared with the brain areas included in other ICs. Two ICs were found to be sensitive to the auditory model and four ICs sensitive to the visual features.

Figure 9 depicts the two ICs that were sensitive to auditory features. Particularly, IC1 was sensitive to the overall loudness of the soundtrack, although more so to speech, while IC2 was more selectively activated only during occurrences of speech in the movie, and did not follow the low-level features of the soundtrack (**Figure 10**). However, a more fine-grained separation of brain areas was revealed with stimulus modeling approach in the auditory cortex (**Figure 11**). Furthermore, the ICs often covered a larger area of the cortex than the results revealed by model based analysis. In particular, the areas in the posterior temporal, and lateral occipital cortices were strongly functionally connected to the other areas of the ICs but not significantly correlated with the stimulus model (**Figure 12**).

Parts of the ICs may activate independently of each other during parts of the experiment (see time-points A–D in **Figure 9**). This may cause them to not be significantly correlated with the stimulus model in a voxel-wise analysis, even if they still are significantly functionally connected with other regions of the IC based on the activity timecourses over the entire experiment. Similarly to the auditory ICs, the ICs sensitive to the visual features showed variable sensitivity profiles to stimulus features, and the ICs were spatially larger than the areas responsive to the full visual model. Together, the visual ICs covered brain regions very similar to those responsive to the different visual features in the single-feature analysis presented in **Figure 11 B**.

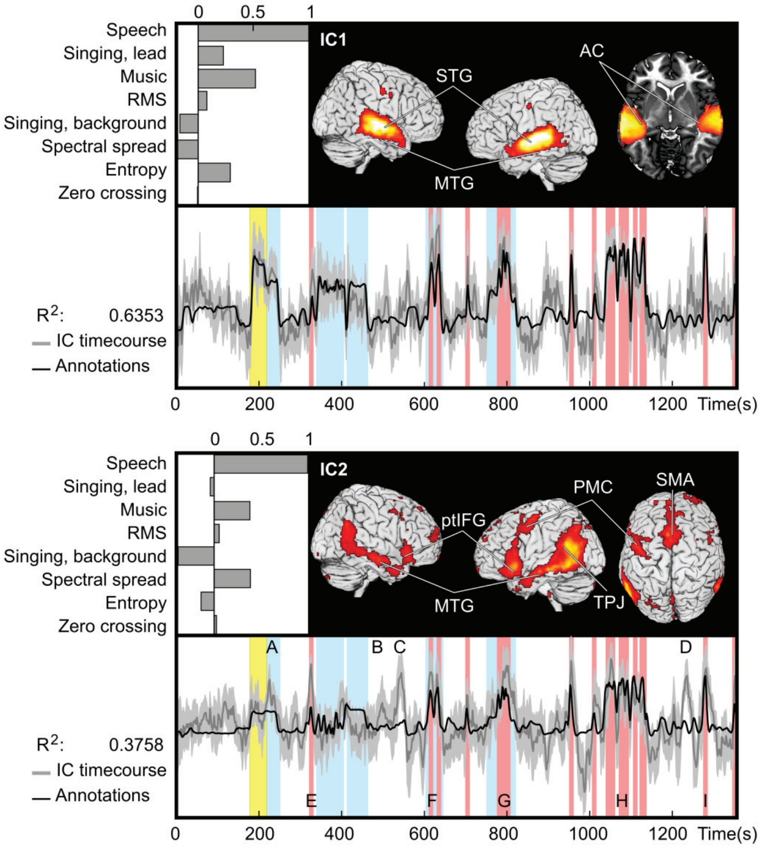

Figure 9: Two ICs that were found to be sensitive to auditory features in Study I. Spatial maps show the locations of the strongest voxels of the ICs. The bar plots show the normalized weights used to fit the auditory model to the mean activity of the IC over subjects. The mean time courses (dark gray; ±95% confidence interval light gray) of the ICs are plotted with the time course of the fitted stimulus model (black) below the spatial maps. R^2 indicates the coefficient of determination of the fitted model and time course of the ICs. Vertical bars show time intervals when there is speech (red), singing (yellow), and music (blue) in the sound track. Additional abbreviations: AC – auditory cortex, MTG – middle temporal gyrus, SMA – supplementary motor area. Activation peaks labeled A–D are not explained by stimulus model and are caused by activity in isolated sub-regions of IC2, whereas speech (peaks E–H) causes robust activity in the entire network represented by IC2.

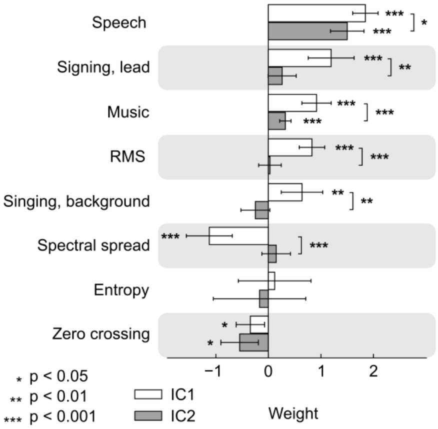

Figure 10: Weights of auditory features in single-feature GLM analysis of IC1 and IC2. Asterisks indicate weights that differ significantly from zero or between the ICs.

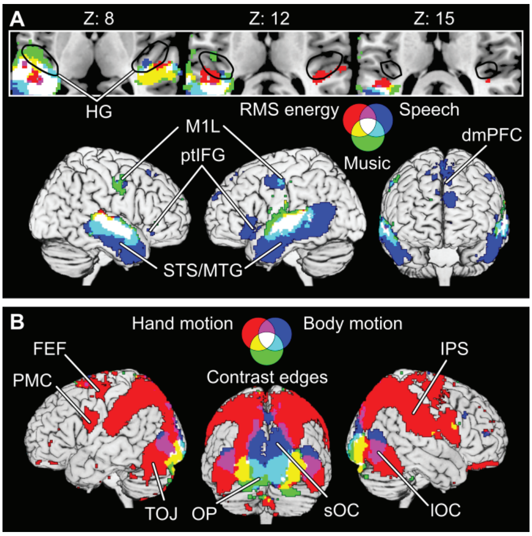

Figure 11: Model based voxel-wise results. A: Spatial maps of brain areas activated by sound loudness (red), speech (blue), and music (green). Overlap of loudness, speech, and music sensitive areas are indicated by mixed colors. **B:** Spatial maps of brain areas activated by hand motion (red), body motion (blue), and contrast edges (green) in the image. Overlapping areas are indicated by mixed colors. Additional abbreviations: M1L — primary motor cortex, lip area, dmPFC – dorsomedial prefrontal cortex, HG – Heschl's gyrus, lOC/sOC – lateral/ superior occipital cortex, OP – occipital pole, TOJ – temporo-occipital junction.

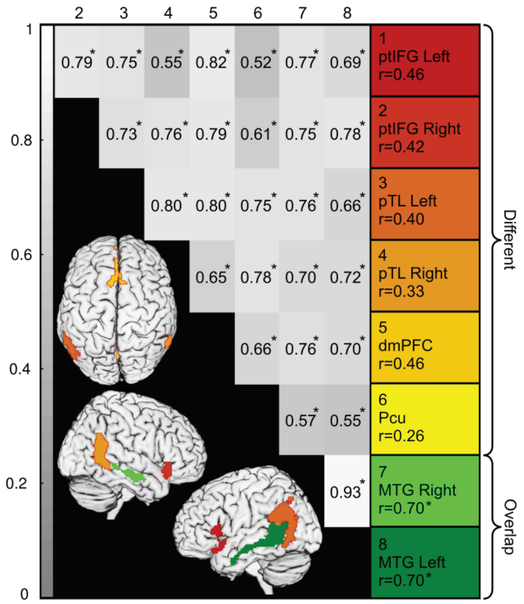

Figure 12: Comparison of speech sensitive IC2 and areas correlated with the auditory model in the voxel-wise analysis. LEFT: Areas where ICA and voxel-wise results differed are color-coded with red–yellow, and overlapping areas are green. The upper-triangle entries of the correlation matrix between the regions are presented on grey background. The grey shades correspond to the magnitudes of correlation. **RIGHT:** The correlation coefficients of the auditory model with each ROIs time courses. Color coding corresponds to the colors on the brain images. Asterisks indicate significant correlations in both panels.

4.2.4 Conclusions

The results show that both converging and complementary information can be revealed using stimulus modeling and model-free methods that reveal the functional covariance structure of the fMRI data. In this case, more fine-grained details could be revealed about some of the brain networks with a carefully constructed stimulus model. However, the covariance, or network structure, between brain regions may contain important additional information. Some regions may generally covary with other regions of the networks represented by the ICs but act independently at given instances. Thus, to understand the brain more fully, both covariation between brain regions as well as their activity time courses should be studied.

4.3 Study II: Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception

4.3.1 Aim of the study

The aim was to reveal brain areas that are particularly important in perception of socially relevant stimulus content compared with contrasting non-social content. A particular aim was to see to what extent the results previously reported in reductionist experiments generalize to more naturalistic stimulus environments. Furthermore, the study aimed to test whether the use of a large set of video clips would allow efficient localization of brain regions participating in processing of multiple types of social content in a relatively short experiment. This would allow the stimulus material and model to be used as an independent localizer to pinpoint social brain regions of interest for further studies.

4.3.2 Materials and methods

Twenty healthy volunteers watched 137 carefully pre-selected short movie clips that were rated (on a continuous scale from 0 to 1 at 5 Hz sampling frequency) by two independent raters for eight social (faces, bodies, biological motion, goal-oriented action, emotion, social interaction, pain, and speech) and six contrasting non-social features (houses, objects, rigid motion, non-goal-oriented action, humans not participating in social interaction, and non-human sounds). **Figure 13** shows illustrations of all the modeled features included in the study. The models of the stimulus content were used as explanatory variables in a GLM analysis to reveal which areas participated in processing of the single variables, and areas that responded preferentially to social *vs.* non-social content. Finally, a functional network was built based on the temporal correlation of the brain areas that responded to multiple social features, and community detection was used to parcellate the network to describe the functional sub-networks participating in social perception.

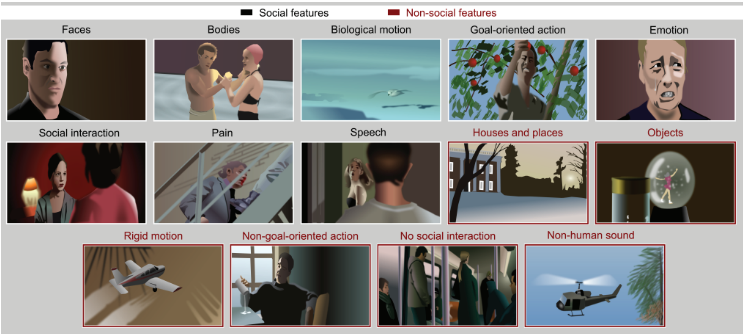

Figure 13: Examples of movie scenes representing the stimulus features modeled in Study II. Frame color of the images indicates the social (white) and non-social (red) features.

4.3.3 Results

Contrasting social *vs.* non-social features in the GLM analysis revealed that temporal areas, amygdala, thalamus, premotor cortex and parts of medial prefrontal cortex were significantly more sensitive to social than non-social features. The opposite was true for many visual areas (**Figure 14a**). Areas similar to those showing a preference to social *vs.* non-social features, particularly in the posterior temporal lobe, were sensitive to several different social categories (**Figure 14b**), and specifically, posterior STS was significantly activated by all social features.

Figure 15 shows the parameter estimates of the GLM analysis reflecting the relative response strengths of the brain areas that were sensitive to multiple social features. The pSTS is the only region that both (i) significantly responds to all social features with nearly equal amplitude, and (ii) does not respond to any of the non-social features.

Figure 16 visualizes the extended brain network specialized for social perception based on the correlation between the response time courses of areas that were responsive to multiple social features. These regions are split into four separate subnetworks using network community detection, and the results largely follow networks also revealed in Study I using ICA as well as other prior studies (Damoiseaux et al., 2006; Biswal et al., 2010). The fronto-temporal network forms a central core of this network, and particularly, the right pSTS is the strongest hub of the network. Compared with other regions of interest, areas that are part of the same sub-network also share more similar distributions of parameter estimates for stimulus features than areas that are in different networks, although the response amplitudes may differ (**Figure 15**).

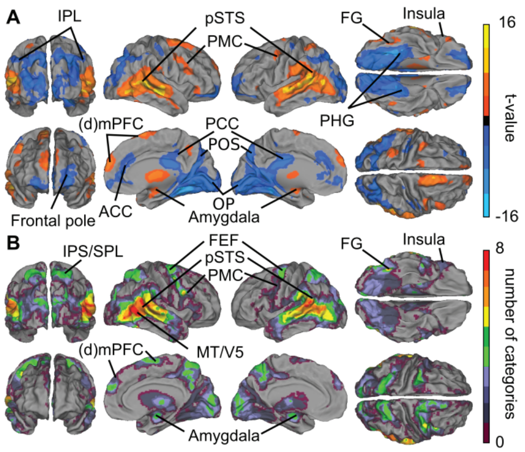

Figure 14: Summary of brain areas showing preference to social stimulus features. A: Contrast showing which brain areas were activated more by social than non-social features (hot colors) or vice versa (cold colors). **B:** Cumulative activation map showing the number of individual social features that significantly correlated with the activity of each brain area. The results indicate that particularly the posterior temporal lobe areas are important for processing all types of social content considered in this study. Additional abbreviations: ACC – anterior cingulate cortex, FG – fusiform gyrus, IPL – inferior parietal lobule, PHG – parahippocampal gyrus, POS – parieto-occipital sulcus.

4.3.4 Conclusions

The results demonstrate the feasibility of mapping brain regions participating in several subtasks of social perception during a single relatively short experiment. The results are largely concordant with the results of prior studies. The stimulus material and corresponding models of stimulus content can be used as an independent localizer of the "social brain" regions participating in the processing of the annotated stimulus content for future studies. Furthermore, the results reveal that the posterior temporal lobe, and particularly the pSTS region, plays an important role in social perception. This area may act as a connector hub of the social brain integrating multiple types of social information together with sub-networks more specialized to particular types of social information.

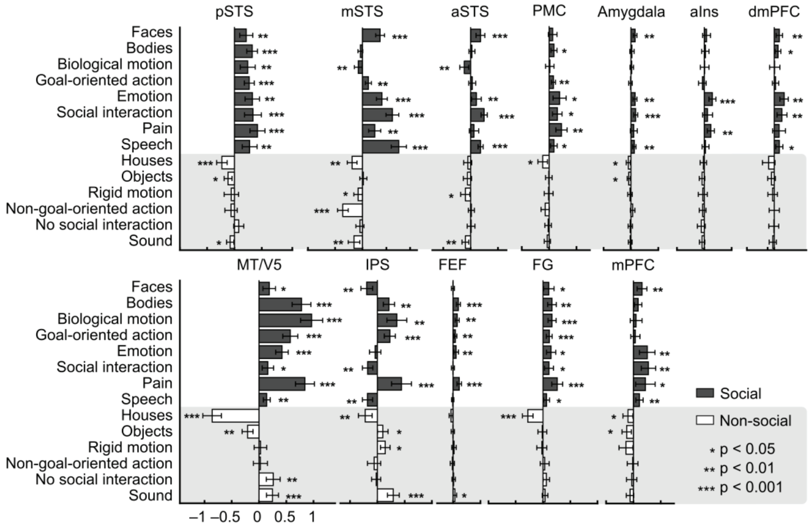

Figure 15: Beta weight parameter estimates in the GLM analysis in regions that were responsive to multiple social categories. Social features are plotted with gray bars and non-social features in white bars on gray background. Error bars correspond to the 95% confidence interval of the mean beta weight across subjects. Asterisks indicate the significance of the contrast in the GLM analysis. The areas responding more to non-social *vs.* social features presented in the article have been omitted. Additional abbreviations: aIns – anterior insula a/mSTS – anterior/middle superior temporal sulcus.

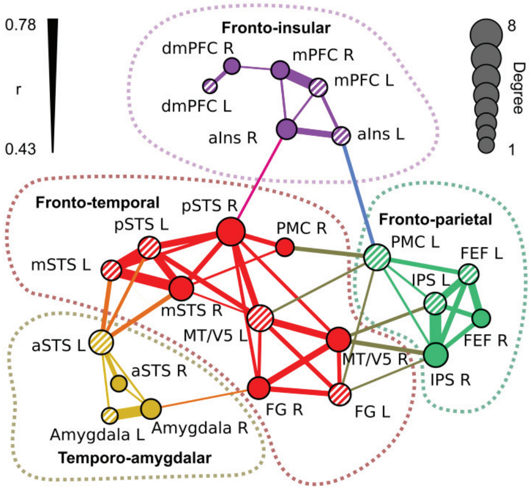

Figure 16: Functional similarity structure of the brain areas activated by multiple social features. The thickness of the edges indicates the correlation coefficients between the time courses of the regions of interest, and the diameter of the nodes indicate the number of edges connected to the node (degree). Left hemisphere nodes are indicated by striped colors.

4.4 Study III: Synchronous brain activity across individuals underlies shared psychological perspectives

4.4.1 Aim of the study

Study III aimed to demonstrate how taking different perspectives toward naturalistic scenes influences the subjects' attention and behavior while viewing a movie, and how the differences in processing manifest in the brain activity. The specific hypothesis was that similar mind states would be reflected in similar brain states when people view the same stimulus having two different perspectives.

4.4.2 Materials and methods

Initially, in Experiment 1 of the study, the viewing patterns of 31 participants were recorded using EyeLink1000 gaze tracker while they watched the first 10 minutes of a television show (Desperate Housewives, Season 1, Episode 15, Cherry Alley Productions, 2005). All participants watched the video twice either taking a perspective hypothesized to bias their attention to social content (being a police detective) or to non-social, inanimate objects (being an internal/external decorator) (see **Figure 17 A**). The perspective taking was motivated by a short background story for both perspectives. Same experimental paradigm was presented to 20 different subjects during fMRI. The order of the perspectives was counterbalanced in both experiments.

In Experiment 2 further 13 subjects were recruited for concurrent recording of fMRI and eye tracking data to verify that the comparison of eye movements and fMRI responses was not biased because separate subjects were analyzed in the two cases. The eye tracking data of two of the subjects in Experiment 2 were lost due to technical difficulties leaving 11 subjects left for the final eye gaze analysis. The subjects were further asked to fill in questionnaires on the difficulty of the perspective taking task and their attention towards perspective-relevant and irrelevant content in the stimulus. Finally, in their freeform accounts the subjects described their behavior during the different tasks (Experiments 1 and 2) and what in particular in the stimulus they found important for the different tasks (Experiment 2 only).

The saccade amplitudes and fixation durations between perspectives were compared and the inter-subject similarity of gaze locations was analyzed based on the spatial correlation of fixation heat maps during each 2-second time window corresponding to the TR of the fMRI data. The fixation heat maps were created by placing a Gaussian kernel (FWHM ~ 2.35° in the visual field) at each fixation position within the viewing area. Additionally, the average fixation patterns between perspectives were compared using similar fixation heat maps created over the whole video presentation.

The fMRI data were analyzed using ISC. The pairwise correlations between subjects were calculated for both perspectives. Non-parametric permutation testing was used to find significant differences in the strength of ISC across the perspectives. Furthermore, Mantel tests (see **Figure 17 B**) and k-nearest-neighbors (kNN) classifiers (see **Figure 17 C**) were used to reveal areas whose ISC was predictive of the particular perspective the subjects were taking. Classification was performed with all possible odd k-values (1–19 in experiment 1; 1–11 in experiment 2) yielding the percentage of correctly classified subjects for each k-value. Only those areas, whose classification accuracy was higher than chance for at least half of the k-values are reported in the results. Analogous classifier was also used to assess whether gaze patterns alone were predictive of the perspectives of the subjects.

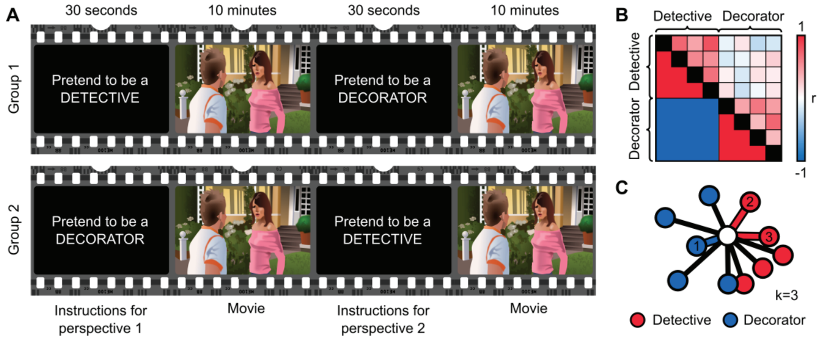

Figure 17: Experimental design and analyses. A: Participants watched the same movie clip twice from social (detective) and non-social (interior decorator) perspectives, with the starting perspective counterbalanced across participants. **B:** Mantel test was used to compare the pairwise ISC values (upper triangle entries) with a correlation matrix template (lower triangle entries) where ISC in same-perspective pairs (red) was higher than different-perspective pairs (blue). **C:** Subjects were classified using a k-nearest-neighbors classifier according to the labels of the training subjects (detective–red, decorator–blue) with whom their ISC was highest. In the visualization the proximity between two dots reflects the strength of the ISC between those subjects. The nearest three neighbors are indexed according to their proximity to the current subject, and the links are highlighted with the color corresponding to their class. For k=3 the current subject (white dot) would be classified as a detective because two of the three nearest neighbors (neighbors 2 and 3) are detectives.

4.4.3 Results

The subjects reported that they paid more attention to task-relevant *vs.* task-irrelevant features in both conditions (**Figure 18 A**). The reported bias of attention toward task relevant features did not differ across conditions. The subject also reported that they found both tasks equally difficult (**Figure 18 B**). In their freeform reports, the subjects further indicated that they were behaving in perspective-relevant manner. For example, subjects in the Detective perspective reported assessing the motives, utterances and facial expressions of the characters. In contrast, in the Decorator perspective, subjects reported thinking of ways how to improve the interiors of the houses, and focusing on the background and furniture visible in the stimulus.

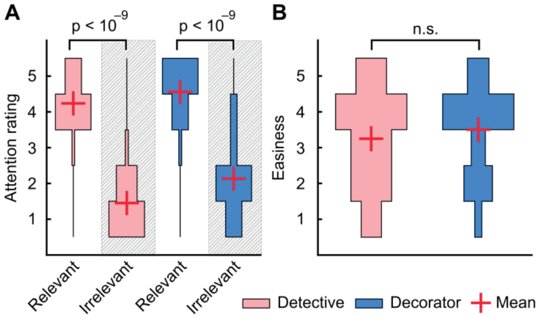

Figure 18: Behavioral results. A: Distributions of ratings across subject of their attention (on scale 1–5) to perspective-relevant and irrelevant items. **B:** Subjective evaluations of the difficulty of the tasks on a scale from 1 (very difficult) to 5 (very easy). Detective perspective is indicated by pink and decorator perspective with light blue color. The red crosses indicate the mean of the ratings across subjects.

Detective ('social') perspective biased fixations towards the center of the screen where the actors were typically shown, whereas decorator ('non-social') perspective biased fixations towards the edges of the screen where the interiors/exteriors were shown (**Figure 19 A**). Furthermore, participants made shorter saccades and longer fixations during detective than decorator perspective (**Figure 19 B**). Inter-subject synchronization of eye movements (eyeISC) was significantly stronger during the detective than in the decorator perspective (**Figure 19 C**). Crucially, the difference in the strength of eyeISC manifested only after the opening credits, which contained no task-relevant information (see the beginning of the timecourses in **Figure 19 A**). The results of the eye-tracking data analysis were similar both inside and outside of the scanner (**Figure 19 D–E**).

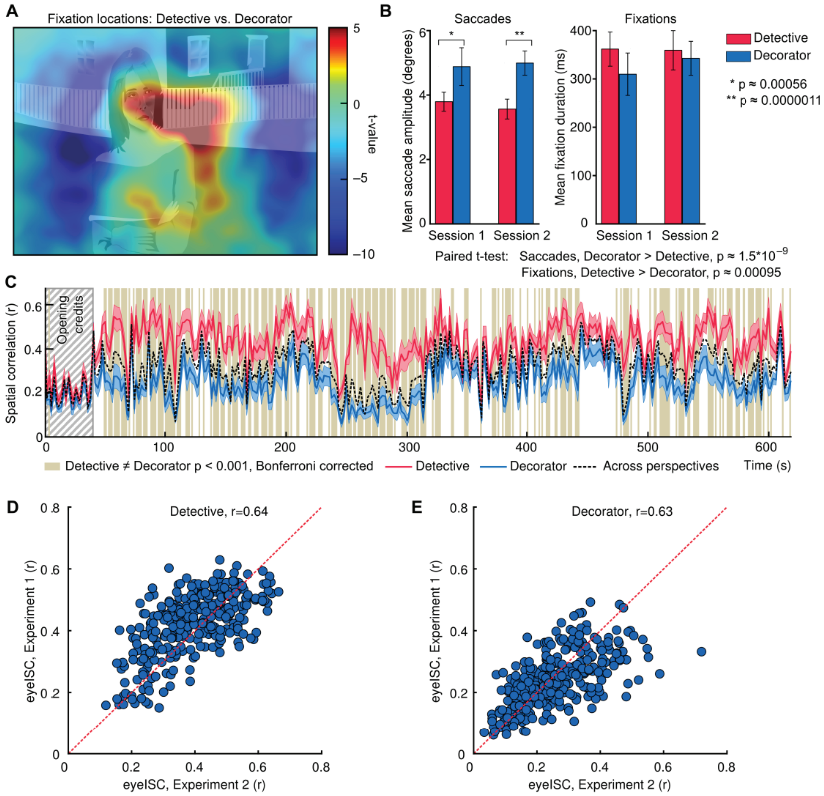

Figure 19: Eye movement patterns across tasks. A: The subtraction heatmaps (T-scores, unthresholded) show regions receiving more fixations in the social (turquoise to blue) and non-social (yellow to red) perspective conditions. Heatmaps were computed over the entire experiment and are here shown overlaid on a sketch of a representative frame of the movie. **B:** Saccades were longer in the decorator condition and fixations were longer in the detective condition. **C:** Time courses of inter-subject synchronization (±95% confidence interval) of gaze position within perspectives (red and blue) and across perspectives (black dashed line). Time points with significantly different eyeISC across conditions are indicated by vertical bars. Opening credits are indicated by gray striped background. **D and E:** Correspondence of eyeISC values in Experiment 1 *vs.* Experiment 2 in detective and decorator conditions, respectively. Dashed red line indicates the region where eyeISC in both experiments would be of the same magnitude.

Figure 20 shows the contrast of ISC strength between the perspective taking conditions. Activity in most of the occipital lobe and parts of the STG, STS and TPJ was more correlated across participants in the Detective *vs.* Decorator condition. Only region showing the opposite effect was located in the junction of ventral anterior cingulate cortex and orbitofrontal cortex, but this effect was not replicable in either experiment alone.

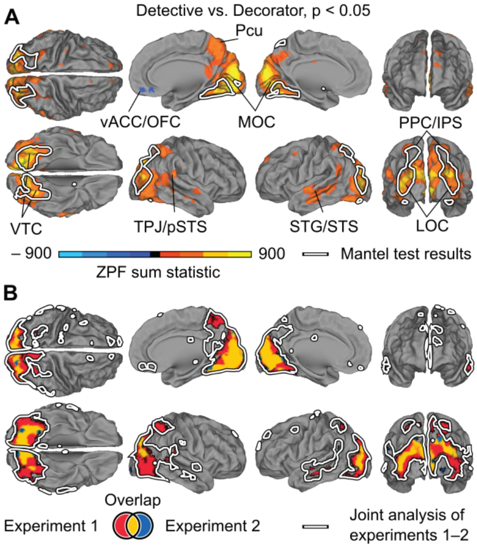

Figure 20: Brain regions exhibiting stronger ISC in the detective *vs.* decorator perspective (orange to yellow) and vice versa (blue to turquoise). A: Results are calculated on the pooled data of both experiments. Results are thresholded at $p < 0.05$ (FWE controlled). White outlines show areas where ISC was higher within *vs.* across (see **Figure 21**). Additional abbreviations: MOC – medial occipital cortex, PPC – posterior parietal cortex, vACC/OFC – ventral anterior cingulate cortex/orbitofrontal cortex, VTC – ventral temporal cortex. **B:** Areas where ISC was stronger in Detective *vs.* Decorator perspective in Experiment 1 (red), Experiment 2 (blue) or in both experiments (yellow). White outlines indicate the results based on the pooled data in panel A.

Lateral occipital, inferior temporal, and posterior parietal cortical regions showed significantly higher ISC in subject pairs assuming the same perspective than in pairs assuming different perspectives (**Figure 21 B**). Therefore, we could predict the correct perspective of the left-out subjects based on the ISC with significantly above-chance accuracy in these higher-order areas (**Figure 21 A**). The results of the classification analysis were replicable in both fMRI datasets as well as in the joint analysis. Highest classification accuracies were approximately 93% in single-experiment analysis (N=13) and approximately 88% in the pooled data (N=33).

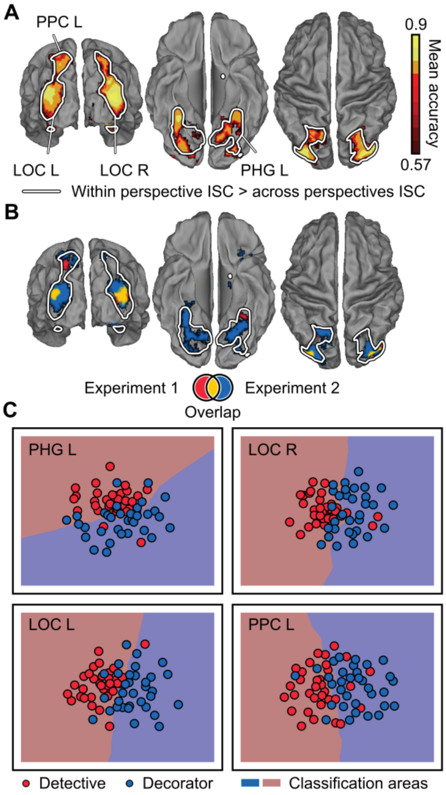

Figure 21: Areas showing higher ISC within *vs.* across conditions. A: Brain regions where accuracy of voxel-by-voxel classification based on pairwise ISC values calculated over the entire stimulus was significantly above chance level ($p<0.001$, uncorrected) with at least half of the k-values. The color coding (red—yellow) indicates the average accuracy over the classification results. White outlines indicate areas exhibiting higher within *vs.* across perspectives ISC in the Mantel test. **B:** Areas where classification accuracy was significantly above chance level (threshold as in panel A) in Experiment 1 (red), Experiment 2 (blue) or both (yellow). **C:** Scatter plots show the subjects plotted on a 2D plane using multidimensional scaling where the proximity between two subjects corresponds to their ISC at the brain regions annotated in panel A. Blue and red background colors indicate areas where new subjects would be classified as decorators and detectives, respectively, using a kNN classifier trained on the entire group ($k = 33$).

4.4.4 Conclusions

The results of this series of experiments demonstrate the feasibility of studying complex social and cognitive tasks, such as perspective taking, during naturalistic experimental conditions. Similarity of brain activity in lateral occipital cortex, parahippocampal gyrus, and posterior parietal cortex, and eye gaze reflect both the similarity of psychological perspectives as well as the particular perspective the subjects are taking. Additionally, taking the “social” *vs.* “non-social” perspective increases the ISC in the pSTS shown to be important for social perception in Study II as well as large parts of the occipital lobe. Moreover, the results were replicated in two independent experiments demonstrating the robustness of the findings. The increased ISC of brain activity while subjects were taking similar *vs.* dissimilar perspectives toward the movie events supports the hypothesis that similar mental states are associated with similar brain activity patterns.

5. Discussion

The studies presented in this thesis reveal how the brain processes complex social information during viewing of naturalistic audiovisual stimulation. Study I demonstrated that both converging and complementary information of brain areas and networks participating in the processing of complex stimuli may be revealed using linear modeling methods and ICA. Study II both confirmed results from Study I, such as the location of speech sensitive regions, and extended these by revealing the pSTS region as the integrating hub of a distributed brain network for social perception. This network consisted of partially specialized subnetworks participating in more specific social tasks. Finally, Study III demonstrated that even complex brain processes such as perspective taking could be studied in naturalistic experimental conditions with carefully planned experimental designs, yielding high replicability across experiments. Together, these studies show the feasibility and power of naturalistic experimental designs in probing multiple brain processes simultaneously that have previously required lengthy series of experiments. Furthermore, the dynamic stimuli and tasks employed in these studies increase the ecological validity of experimental conditions, and enable probing more complex social cognitive brain functions than is possible with simplified static stimuli.

5.1 Brain regions and networks processing naturalistic stimuli

Study I revealed convergent results between stimulus driven analyses and model free independent component analysis that was subsequently correlated with the stimulus models. However, some regions that were included in the brain networks found by ICA were not significantly correlated with the relatively simple stimulus models employed in the study although they were functionally connected with the other regions of the functional networks. These regions included parts of the posterior temporal lobe, temporo-occipital visual areas, posterior parietal cortex, as well as some prefrontal regions. Together with the results of Study II, these differences suggest that these regions participate in processing of more diverse set of stimulus features than were modeled in Study I, while they still show significant functional connectivity with other regions processing more selectively the biological motion categories and speech to which the ICs were sensitive.

Similar observation of specialization of the temporo-frontal network of IC2 to speech, and superior temporal IC1 more generally to sounds in Study I was also found in a recent study using an audio-story as a stimulus (Boldt et al., 2013). Thus, the observation seems to generalize to different stimuli and modalities. However, since some parts of the stimulus in Study I activated sub-regions of the speech-sensitive IC2 independently, the network may consist of distinct nodes that work together during speech comprehension and independently when other type of processing is required. For example, in Study I, sub-regions of the speech-sensitive network seemed to activate during some instances of visual presentation of language. Furthermore, in Study II, the activity of posterior regions of the network followed multiple social signals presented in short videos as discussed below.

5.2 Posterior STS as an integrating hub of the social brain

Study II revealed that the pSTS region responds to a very wide range of social signals during presentation of short audiovisual movie clips. Moreover, the responses seem selective to social *vs.* non-social contents of the videos. Recently, Watson and co-workers (2014) found additional evidence of the preference of the pSTS region to social stimuli supporting the results of Study II presented here. They showed that large parts of the STS respond selectively to people in visual, audiovisual and auditory stimuli. Particularly, posterior regions of the STS were further found to be heteromodal areas that integrated auditory and visual information so that the responses were higher to audiovisual stimuli than either modality alone. However, it is obvious from the results of Study II that there are differences in the spatial extent and location of the activated regions of posterior temporal lobe adjacent to the pSTS depending on the social features being analyzed. Thus, there likely are multiple overlapping subregions in the posterior temporal lobe processing different aspects of the incoming sensory streams. While the functions of the motion sensitive middle temporal visual area (MT), and body sensitive extrastriate visual area (extrastriate body area; EBA) in the inferior parts of posterior temporal lobe have been studied extensively using controlled experimental paradigms (Downing et al., 2001; Peuskens et al., 2005; Thompson and Parasuraman 2012) the functions of the other areas between MT/EBA region and the temporoparietal junction are less clear. Several higher-order processes, such as intentionality (Nummenmaa and Calder 2009), empathy, agency and theory of mind, but also attentional reorienting, have been attributed to the temporoparietal junction and adjoining areas of the posterior superior temporal sulcus (for a review, see Decety and Lamm 2007). Recent research has also revealed that the pSTS is a central node in the large scale structural network of the brain (Hagmann et al., 2008) and posterior temporal areas have been suggested to play a role in integrating information over long time scales (Hasson et al., 2008). These observations are in line with the results of Study II that suggests that the pSTS and adjacent regions may play a key role in integrating social information, and they form an important functional hub of the social brain, although the aforementioned studies did not test whether the structural centrality of pSTS and its long time-window of integration are specifically related to its role in social processing.

Temporarily disrupting the function of pSTS with repetitive transcranial magnetic stimulation may impair perception of biological motion (Grossman et al., 2005) and stimulation of the adjacent TPJ can trigger “out-of-body” experiences (Blanke et al., 2005), which was interpreted to suggest that the TPJ is critical for conscious experience of the coherent self. Others have suggested that pSTS may be an entry point of information to large scale brain network participating in action observation and recognition (Peuskens et al., 2005). It is clear that the pSTS region is active in a large variety of conditions that are relevant for social processing. However, it is unclear how the pSTS would categorize something to be socially relevant. If pSTS is a central integrating hub this could be achieved by the structural and functional connections of the more specialized sub-networks participating in social perception. However, further studies are required to test whether the main role of the pSTS and TPJ is in integration of information about the self and social environment or is the region an entry point from where information spreads to other parts of the network, and whether pSTS region has a causal role in classifying something as social.

5.3 Shared brain activity reflects mutual understanding

Recent studies indicate that the lateral occipital and ventral temporal areas revealed in Study III to be modulated by perspective taking show a more complicated response profile to scene statistics than previously thought. The responses in these areas may depend on the co-occurrence of scene contents (Stansbury et al., 2013) rather than a single target. Moreover, the representations of objects in these areas are modified, for example, by visual attention (Nishimoto et al., 2013). The results of Study III are in line with these observations and extend them by showing that while taking a similar psychological perspective with others the attention toward and interpretation of movie events become more similar across individuals, which is also reflected in the similarity of brain responses.

Other recent studies have revealed additional evidence that sharing a similar point of view with others is mediated, in part, by having the brains of those individuals function in similar ways. For example, people who perceived the H1N1 pandemic as a great risk exhibited enhanced ISC in the anterior cingulate cortex compared with those who perceived the risk to be smaller (Schmälzle et al., 2013) while they viewed TV reports on the topic. The synchrony of brain activity across individuals is also increased while they view emotional episodes in movies (Nummenmaa et al., 2012) and thus presumably share a similar emotional state. Similarly, ISC is also increased while people actively mentally simulate the actions of boxers depicted in video clips (Nummenmaa et al., 2014).

Together with these studies, the results of Study III suggest that similar brain activity could underlie mutual understanding between individuals. However, it is not obvious whether enhanced synchrony is a consequence or a mechanism for sharing a similar mind state (Stolk 2014). The extent to which the similarity of brain responses can predict the similarity of views between people is an interesting topic for further research. To this end, behavior of the subjects, as well as their thoughts during the tasks should be carefully probed to quantify the similarity of their states of mind. Importantly, perspective-taking tasks should also be directly compared with explicit attention tasks during the same stimulus to reveal which brain mechanisms are specific to higher-level perspective taking, and which are more generally involved in spatial attention.

5.4 Dynamic changes in functional brain networks and ISC

Traditional neuroscientific studies have mapped brain responses to isolated stimulus categories and features, often focusing on single brain areas of interest. However, as Study II demonstrates, single brain areas can be sensitive to multiple features during perception of naturalistic dynamic scenes, and as seen in Study I, individual parts of brain networks may act independently during some parts of the experiment and coactivate with the network at other time points. Many recent studies are beginning to highlight the importance of considering the brain as a network where multiple areas participate in several tasks in different configurations depending on the task demands (Sporns 2011). The connectivity and co-activation of areas with other brain regions affects their function at any given moment. Thus, the use of simplified and isolated stimuli may reveal only part of the role they play when participating in complex processing of the natural world.

To address the dynamically changing processing demands on brain areas and networks, recent studies have started to investigate time-varying inter-subject synchrony of brain activity across subjects, and temporal dynamics of brain networks. For example, inter-subject correlation strength is modulated by subjectively evaluated emotional valence and arousal of movie clips (Nummenmaa et al., 2012). In a similar vein, in Study III the behavioral similarity (eyeISC) was used to predict the ISC of brain activity calculated in sliding temporal windows, but no consistent link was found when the analysis was performed over the entire experiment. To further address the temporal modulations in functional connectivity and inter-subject synchrony with greater temporal accuracy, new methods, such as phase synchrony (Glerean et al., 2012), and Kalman filtering (Havlicek et al., 2010) can be applied to fMRI data.

Uncovering the principles of the temporal behavior of brain networks is a major challenge and opportunity for future work in neuroscience. The number of possible connections in the brain is vast and we will need sophisticated analysis and visualization approaches to make sense of the complex temporal patterns. One of the first studies to systematically investigate how the time-to-time connectivity changes between brain regions (Smith et al., 2012) revealed temporally independent modes of functional networks that characterize the different network configurations a region participates in at different times. However, additional work on the temporal variability of functional connections is of great importance when trying to understand the human brain. Recent work in different fields of cognitive neuroscience (for examples, see reviews on emotions and memory, respectively: Hamann 2012; Fuster 2009) is highlighting the need to view the brain as a dynamically reorganizing network.

5.5 Advantages of using naturalistic stimuli

Complex cognitive processes such as perspective taking may be difficult to address using simplified experiments. Furthermore, the methods developed in Study III could be further applied to study abnormal brain functions in various clinical conditions. In a recent paper we showed that the inter-subject correlations in high-functioning autistic individuals are decreased compared to matched controls (Salmi et al., 2013) during viewing of the full version of the movie presented in Study I of the present work. The high classification accuracies reached in Study III suggest that with careful design of the stimuli, novel insight could be gained into various abnormalities of social cognition.

There is evidence that subtle functional abnormalities may be poorly detected using simplified psychological tests. Thus, they may be better studied, and possibly even diagnosed, using more naturalistic experiments and new analysis methods such as the ones described in the current work. For example, while some individuals with autism pass standard theory of mind tests, they may struggle in attributing mental states to characters depicted in movies designed to probe their theory of mind capacity more thoroughly (Heavey et al., 2000). It has also been shown that eye gaze patterns of individuals with autism differ from those of typically developed controls during natural viewing of movies (Hasson et al., 2009). Moreover, the fixation times on different parts of the face predict the social competence of individuals with autism (Klin et al., 2002) during viewing of naturalistic social situations. Finally, naturalistic tests have also been shown to yield different assessments of memory performance compared with laboratory experiments, although this may be related to different methodologies probing different aspects of memory (Koriat and Goldsmith 1994) in the two environments. The brain correlates of these behavioral effects should be addressed in more detail in future experiments.

5.6 Challenges of naturalistic neuroimaging experiments

While using naturalistic stimuli and experimental conditions increase the ecological validity of neuroscientific experiments and may significantly increase their detection power, there are still obstacles that hinder their use in research. Despite technological advances, MRI scanners are still noisy and cramped, and the subjects are required to stay stationary for extended periods of time. Thus, it may be difficult to fully immerse oneself into the movies during fMRI scanning. However, similar problems are true for traditional experiments as well. When the stimuli are more interesting the subjects may in fact find it easier to maintain their attention during the experiment. This may be particularly beneficial in experiments that require subjects to attend a given target for long periods of time.

Recently, researchers have further expressed demand on moving beyond passive viewing conditions toward interactive experiments where two communicating people would be imaged at the same time (Hari and Kujala

2009; Hasson and Honey 2012). However, compared with passive viewing of naturalistic stimuli, the problems of subject motion make the study of naturalistic social interactions even more difficult. Although subjects have been imaged, for example, during story telling (Stephens et al., 2010) and gestural communication (Schippers et al., 2010), having natural conversations with others can be problematic while trying to avoid excessive head motion. For example, in the story-telling experiment of Stephens and colleagues (2010) the researchers had the story teller practice the story in the fMRI scanner three times prior to scanning to avoid head motion during scanning. Despite these challenges, to truly understand the brain mechanisms that enable us to participate in everyday social interactions we must eventually be able to measure the brains of people during such episodes.

The modeling of both interactive experiments as well as naturalistic videos is one of the major challenges in the analysis of the data gathered during naturalistic experiments. As mentioned earlier, several features of naturalistic stimuli may coincide during any given moment, and covary in time. In Study II, short video stimuli were selected so that the features of interest were relatively uncorrelated. However, this may not be appropriate if the brain processes to be studied evolve over longer time scales. Furthermore, it is not obvious which features of the stimuli are relevant to be modeled.

While the stimulus-model based results of Studies I and II are consistent with prior literature, it can be difficult to pinpoint the particular stimulus features that elicit the activity in a given brain area when the stimulus is complex. Obviously, it is not possible to exhaustively model a naturalistic stimulus such as a movie, although high-dimensional models have recently been built (Huth et al., 2012; Çukur et al., 2013) using a hierarchical corpus of semantic categories. However, even the selection of semantic categories as the class of explanatory variables may bias the conclusions of the study despite the size of the model. Moreover, as the size of the model increases, the chance of overfitting increases and methods such as regularization become necessary. Moreover, additional work should be undertaken to evaluate whether the assumption of linear superposition of responses to different stimulus features is appropriate, as this may significantly affect the way the models should be fitted to the data. Despite the challenges, however, the conclusions based on simplified experiments should ultimately be validated in more naturalistic conditions. For example, although at least the highest peaks of activity during natural stimulation in the fusiform face area seem to correspond to occurrences of faces (Hasson et al., 2004) the selectivity of the FFA responses should eventually be tested against as large a set of competing hypotheses as possible.

In addition to modeling the stimuli, characterizing the subjects' behavior during social interactions requires careful consideration. Ideally, the behavior of the participants should be probed as comprehensively as possible. However, the fMRI environment makes gathering data on, for example, subjects' facial expressions difficult because most equipment cannot be brought to the high magnetic field inside the scanner, and the view from the outside is severely limited. Thus, it is important that the research questions and experimental set-ups are carefully planned, subjects' behavior is thoroughly probed with appropriate questionnaires, and technology is developed that enables additional behavioral data to be gathered during interactive experiments. Importantly, analyzing the behavior and experiences of the participants could potentially solve some of the bias introduced by the pre-selection of features to be analyzed by the researchers. Moreover, as showed in Study III, the experiences of the subjects are reflected in their brain activity even when the stimulus is exactly the same. Thus, the stimulus model alone, no matter how many features are included, may not explain all aspects of the recorded brain activity.

6. Conclusions

The processing of naturalistic social scenes requires the interplay of multiple brain regions. Brain regions traditionally implicated in perception of specific social features participate in large-scale networks where multiple areas are responsible for creating a unified percept of the incoming information. In particular, the posterior superior temporal sulcus and the surrounding areas of temporoparietal junction, and posterior temporal lobe seem to be key structures potentially integrating multiple aspects of social perception together with other partly segregated subnetworks. It is thus likely that the functional role of a brain region at a given time is dependent on which other brain regions it is active and communicating with. Additionally, the modulation of brain activity by psychological perspectives supports the hypothesis that viewing the world in a similar way with other people may cause the brain activity of the persons to be more similar. Thus, the similarity of brain activity may facilitate mutual understanding. The experimental designs and analysis methods presented in this thesis open new possibilities for studying how humans process multitude of social cues in complex naturalistic contexts, and how these processes are modulated by different tasks. Similar approaches can also be applied to study the neural basis of clinical conditions that are poorly addressed using traditional experimental designs.

Bibliography

Adolphs, R. (1999). Social cognition and the human brain. Trends Cogn Sci 3:469–479.

Allison, T, Puce, A, and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278.

Alluri, V, Toiviainen, P, Jääskeläinen, IP, Glerean, E, Sams, M, and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. NeuroImage 59:3677–3689.

Anderson, R.C. and Pichert, J.W. (1978). Recall of previously unrecallable information following a shift in perspective. J Verb Learn Verb Beh 17:1–12.

Bandettini, PA, Wong, EC, Hinks, RS, Tikofsky, RS, and Hyde, JS. (1992). Time course EPI of human brain function during task activation. Magn Reson Med 25:390–397.

Bartels, A. and Zeki, S. (2004a). Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. 21:75–85.

Bartels, A and Zeki, S. (2004b). The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. NeuroImage 22:419–433.

Bartels, A and Zeki, S. (2005). Brain dynamics during natural viewing conditions—A new guide for mapping connectivity in vivo. NeuroImage 24:339–349.

Bartels, A, Zeki, S, and Logothetis, NK. (2008). Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. Cereb Cortex 18:705–717.

Beckmann, CF and Smith, SM. (2004). Tensorial extensions of independent component analysis for multisubject FMRI analysis. NeuroImage (25):294–311.

Belin, P, Zatorre, RJ, Lafaille, P, Ahad, P, and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belliveau, JW, Kennedy Jr, DN, McKinstry, RC, Buchbinder, BR, Weisskoff, RM, Cohen, MS, Vevea, JM, Brady, TJ, and Rosen, BR. (1991).

Functional mapping of the human visual cortex by magnetic resonance imaging. Science 254: 716–719.

Bell, AJ and Sejnowski, TJ. (1995). An information-maximisation approach to blind separation and blind deconvolution. Neural Comput 7:1129–1159.

Biswal, BB, Mennes, M, Zuo, X-N, Gohel, S, Kelly, C, Smith, SM, Beckmann, CF, Adelstein, JS, Buckner, RL, Colcombe, S, Dogonowski, A-M, Ernst, M, Fair, D, Hampson, M, Hoptman, MJ, Hyde, JS, Kiviniemi, VJ, Kötter, R, Li, S-J, Lin, C-P, Lowe, MJ, Mackay, C, Madden, DJ, Madsen, KH, Margulies, DS, Mayberg, HS, McMahon, K, Monk, CS, Mostofsky, SH, Nagel, BJ, Pekar, JJ, Peltier, SJ, Petersen, SE, Riedl, V, Rombouts, SARB, Rypma, B, Schlaggar, BL, Schmidt, S, Seidler, RD, Siegle, GJ, Sorg, C, Teng, G-J, Veijola, J, Villringer, A, Walter, M, Wang, L, Weng, X-C, Whitfield-Gabrieli, S, Williamson, P, Windischberger, C, Zang, Y-F, Zhang, H-Y, Castellanos, FX, and Milham, MP. (2010). Toward discovery science of human brain function. Proc Natl Acad Sci U S A 107:4734–4739.

Biswal, B, Yetkin, FZ, Haughton, VM, and Hyde, JS. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34:537–541.

Blakemore, S-J. (2008). The social brain in adolescence. Nat Rev Neurosci 9:267–277.

Blanke, O, Mohr, C, Michel, CM, Pascual-Leone, A, Brugger, P, Seeck, M, Landis, T, and Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. J Neurosci 25:550–557.

Boldt, R, Malinen, S, Seppä, M, Tikka, P, Savolainen, PI, Hari, R, and Carlson, S. (2013). Listening to an Audio Drama Activates Two Processing Networks, One for All Sounds, Another Exclusively for Speech. PLoS One 8:e64489. doi:10.1371/journal.pone.0064489.

Calhoun, VD, Adali, T, Pearlson, GD, and Pekar, JJ. (2001). Spatial and Temporal Independent Component Analysis of Functional MRI Data Containing a Pair of Task-Related Waveforms. Hum Brain Mapp (13):43–53.

Calhoun, VD and Pearlson, GD. (2012). A selective review of simulated driving studies: Combining naturalistic and hybrid paradigms, analysis approaches, and future directions. NeuroImage 59:25–35.

Caramazza, A, Anzellotti, S, Strnad, L, and Lingnau, A. (2014). Embodied cognition and mirror neurons: A critical assessment. Annu Rev Neurosci 37:1–15.

Carr, HY. (2004). Field gradients in early MRI. Phys Today 57:83.

Cordes, D and Nandy, RR. (2006). Estimation of the intrinsic dimensionality of fMRI data. NeuroImage 29:145–154.

Correa, N, Adali, T, and Calhoun, VD. (2007). Performance of Blind Source Separation Algorithms for FMRI Analysis using a Group ICA Method. Magn Reson Imaging 25:684–694.

Çukur, T, Nishimoto, S, Huth, AG, and Gallant, JL. (2013). Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–770.

Damadian, R, Goldsmith, M, and Minkoff, L. (1977). NMR in cancer: XVI. FONAR image of the live human body. Physiol Chem Phys 9:97–100.

Damoiseaux, JS, Rombouts, SARB, Barkhof, F, Scheltens, P, Stam, CJ, and Smith, SM, B. C. (2006). Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci U S A 103:13848–13853.

David, N, Aumann, C, Santos, NS, Bewernick, BH, Eickhoff, SB, Newen, A, Shah, NJ, Fink, GR, and Vogeley, K. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. Soc Cogn Affect Neurosci 3:279–289.

David, SV, Mesgarani, N, Fritz, JB, and Shamma, SA. (2009). Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli. J Neurosci 29:3374–3386.

Decety, J and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. Neuroscientist 13:580–593.

DeWitt, I and Rauschecker, JP. (2012). Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109:E505–E514.

Downing, PE, Jiang, Y, Shuman, M, and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science 293:2470–2473.

Ekman, P, Sorenson, ER, and Friesen, WV. (1969). Pan-cultural elements in facial displays of emotion. Science 164:86–88.

Fisher, RA. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1:3–32.

Friston, KJ, JT Ashburner, S Kiebel, T Nichols, and WD Penny, eds. (2007). Statistical Parametric Mapping: The Analysis of Functional Brain Images. 1st ed. London: Academic Press.

Frith, CD. (2007). The social brain? Phil Trans R Soc B 362:671–678.

Frith, CD. (2012). The role of metacognition in human social interactions. Phil Trans R Soc B 367:2213–2223.

Frith, U and Frith, CD. (2003). Development and neurophysiology of mentalizing. Phil Trans R Soc London B 358:459–473.

Frith, CD and Frith, U. (2012). Mechanisms of social cognition. Annu Rev Psychol 63:287–313.

Fritz, J, Elhilali, M, David, S, and Shamma, S. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1. Hear Res 229:186–203.

Fuster, JM. (2009). Cortex and memory: emergence of a new paradigm. J Cognitive Neurosci 21:2047–2072.

Gallagher, HL, Happé, F, Brunswicka, N, Fletcher, PC, Frith, U, and Frith, CD. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 38:11–21.

Garrod, S and Pickering, MJ. (2004). Why is conversation so easy? Trends Cogn Sci 8:8–11.

Glerean, E, Salmi, J, Lahnakoski, JM, Jääskeläinen, IP, and Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. Brain Connect. 2:91–101.

Grossman, ED, Battelli, L, and Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. Vision Res 45:2847–2853.

Hagmann, P, Cammoun, L, Gigandet, X, Meuli, R, Honey, CJ, Wedeen, VJ, and Sporns, O. (2008). Mapping the Structural Core of Human Cerebral Cortex. PLoS Biol 6:e159. doi:10.1371/journal.pbio.0060159.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. Trends Cogn Sci 16:458–466.

Happé, FG, Winner, E, and Brownell, H. (1998). The getting of wisdom: theory of mind in old age. Dev Psychol 34:358–362.

Hari, R and Kujala, M. (2009). Brain basis of human social interaction: from concepts to brain imaging. Physiol Rev 89:453–479.

Hasson, U, Avidan, G, Gelbard, H, Vallines, I, Harel, M, Minshew, N, and Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. Autism Res 2:220–231.

Hasson, U and Honey, C. (2012). Future trends in neuroimaging: Neural processes as expressed within real-life contexts. NeuroImage 62:1272–1278.

Hasson, U, Malach, R, and Heeger, DJ. (2010). Reliability of cortical activity during natural stimulation. Trends Cogn Sci 14:40–48.

Hasson, U, Nir, Y, Levy, I, Fuhrmann, G, and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640.

Hasson, U, Yang, E, Vallines, I, Heeger, DJ, and Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. J Neurosci 28:2539–2550.

Hastorf, AH and Cantril, H. (1954). They saw a game: A case study. J Abnorm Psychol 49:129–134.

Havlicek, M, Jan, J, Brazdil, M, and Calhoun, VD. (2010). Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. NeuroImage 53:65–77.

Haxby, JV, Hoffman, EA, and Gobbini, MI. (2000). The distributed human neural system for face perception. Trends Cogn Sci 4:223–233.

Heavey, L, Phillips, W, Baron-Cohen, S, and Rutter, M. (2000). The Awkward Moments Test: A Naturalistic Measure of Social Understanding in Autism. J Autism Dev Disord 30:225–236.

Heider, F and Simmel, M. (1944). An experimental study of apparent behavior. Am J Psychol 57:243–259.

Hein, G and Knight, RT. (2008). Superior temporal sulcus–it's my area: or is it? J Cognitive Neurosci 20:2125–2136.

Heinzel, A, Bermpohl, F, Niese, R, Pfennig, A, Pascual-Leone, A, Schlaug, G, and Northoff, G. (2005). How do we modulate our emotions? Parametric fMRI reveals cortical midline structures as regions specifically involved in the processing of emotional valences. Cognitive Brain Res 25:348–358.

Henderson, JM. (2003). Human gaze control during real-world scene perception. Trends Cogn Sci 7:498–504.

Hickock, G. (2008). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J Cognitive Neurosci 21:1229–1243.

Himberg, J, Hyvärinen, A, and Esposito, F. (2004). Validating the independent components of neuroimaging time-series via clustering and visualization. NeuroImage 22:1214–1222.

Horimoto, CH and Mimica, MRM. (2004). Eye gaze tracking techniques. Comput Vis Image Und 98:4–24.

Huettel, SA, Song, AW, and McCarthy, G. (2004). Functional Magnetic Resonance Imaging. Sunderland, Massachusetts, USA: Sinauer Associates, Inc.

Huth, AH, Nishimoto, S, Vu, AT, and Gallant, JL. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210–1224.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans on Neural Networks 10:626–634.

Hyvärinen, A and Oja, E. (2000). “Independent Component Analysis: Algorithms and Applications.”. Neural Networks.

Itti, L and Koch, C. (2001). Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.

Jääskeläinen, IP, Koskentalo, K, Balk, MH, Autti, T, Kauramäki, J, Pomren, C, and Sams, M. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. Open Neuroimag J 2:14–19.

Kaakinen, JK and Hyönä, J. (2008). Perspective-driven text comprehension. Appl Cogn Psych 22:319–334.

Kaakinen, JK and Hyönä, J. (2014). Task Relevance Induces Momentary Changes in the Functional Visual Field During Reading. Psychol Sci doi: 10.1177/0956797613512332.

Kaakinen, J.K., Hyönä, J., and Keenan, J.M. (2002). Perspective effects on online text processing. Discl Process 33:159–173.

Kanwisher, N, McDermott, J, and Chun, MM. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kauppi, J-P, Jääskeläinen, IP., Sams, M, and Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. Front Neuroinform 4:doi: 10.3389/finf.2010.00005.

Kauppi, J-P, Pajula, J, and Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. Front Neuroinform 8:doi: 10.3389/fninf.2014.00002.

Kay, KN, Naselaris, T, Prenger, RJ, and Gallant, JL. (2008). Identifying natural images from human brain activity. Nature 452:352–355.

Keysers, C, Kaas, JH, and Gazzola, V. (2010). Somatosensation in social perception. Nat Rev Neurosci 11:417–428.

Kiebel, SJ and Holmes, AP (2007). The general linear model. P. 101–125 in Statistical Parametric Mapping: The Analysis of Functional Brain Images. London, U.K.: Elsevier Academic Press.

Kiviniemi, V, Kantola, JH, Jauhiainen, J, Hyvärinen, A, and Tervonen, O. (2003). Independent component analysis of nondeterministic fMRI signal sources. NeuroImage 19:253–260.

Klin, A, Jones, W, Schultz, R, Volkmar, F, and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry 59:809–816.

Koriat, A and Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. J Exp Psychol-Gen 123:297–315.

Kwong, KK, Belliveau, JW, Chesler, DA, Goldberg, IE, Weisskoff, RM, Poncelet, BP, Kennedy, DN, Hoppel, BE, Cohen, MS, and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci 89:5675–5679.

Kätsyri, J, Hari, R, Ravaja, N, and Nummenmaa, L. (2013). The opponent matters: elevated fMRI reward responses to winning against a human versus a computer opponent during interactive video game playing. Cereb Cortex 23:2829–2839.

Lahnakoski, J. (2010). “Functional Magnetic Resonance Imaging of Human Brain during Rest and Viewing Movies.” Master's thesis, Faculty of Electronics, Communications and Automation, Aalto University, Espoo.

Lartillot, O and Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10–15, 2007.

Lauterbur, PC. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. Nature 242:190–191.

Lauterbur, PC. (1974). Magnetic resonance zeugmatography. Pure Appl Chem 40:149–157.

Lerner, Y, Honey, CJ, Silbert, LJ, and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci 31:2906–2915.

Li, Y-O, Adali, T, and Calhoun, VD. (2007). Estimating the Number of Independent Components for Functional Magnetic Resonance Imaging Data. Hum Brain Mapp 28:1251–1266.

Lieberman, MD. (2007). Social cognitive neuroscience: A review of core processes. Annu Rev Psychol 58:259–289.

Lowe, DG. (1999). Object recognition from local scale-invariant features. in Proceedings of the International Conference on Computer Vision 2:1150–1157.

Lu, ST, Hämäläinen, MS, Hari, R, Ilmoniemi, RJ, Lounasmaa, OV, Sams, M, and Vilkman, V. (1991). Seeing faces activates three separate areas outside the occipital visual cortex in man. Neuroscience 43:287–290.

Malinen, S., Hlushchuk, Y., and Hari, R. (2007). Towards natural stimulation in MRI - Issues of data analysis. NeuroImage 35:131–139.

Mandeville, JB and Rosen, BR (2002). Functional MRI. P. 315–349 in Brain Mapping: The Methods, 2nd edition. San Diego, California, U.S.A.: Elsevier Science.

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. J Phys C 10:L55–L58.

Meadows, JC. (1974). The anatomical basis of prosopagnosia. J Neurol Neurosurg Psychiatry 37:489-501.

Morris, JS, Frith, CD, Perrett, DI, Rowland, D, Young, AW, Calder, AJ, and Dolan, RJ. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. Nature 383:812–815.

Mukamel, R, Ekstrom, AD, Kaplan, J, Iacoboni, M, and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Curr Biol 20:750–756.

Newman, MEJ. (2003). The structure and function of complex networks. SIAM Rev 45:167–256.

Nishimoto, S, Huth, A, Bilenko, N, and Gallant, J. (2013). Human visual areas invariant to eye movements during natural vision. J Vis 13:article 1061.

Nummenmaa, L and Calder, AJ. (2009). Neural mechanisms of social attention. Trends Cogn Sci 13:135–143.

Nummenmaa, L, Glerean, E, Viinikainen, M, Jääskeläinen, IP, Hari, R, and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. Proc Natl Acad Sci USA 109:9599–9604.

Nummenmaa, L, Smirnov, D, Lahnakoski, JM, Glerean, E, Jääskeläinen, IP, Sams, M, and Hari, R. (2014). Mental action simulation synchronizes action-observation circuits across individuals. J Neurosci 34:748–757.

Ogawa, S, Lee, TM, Kay, AR, and Tank, DW. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868–9872.

Pajula, J, Kauppi, J-P, and Tohka, J. (2012). Inter-subject correlation in fMRI: Method validation against stimulus-model based analysis. PLoS One 7:e41196. doi: 10.1371/journal.pone.0041196.

Pamilo, S, Malinen, S, Y, Hlushchuk, Seppä, M, Tikka, P, and Hari, R. (2012). Functional Subdivision of Group-ICA Results of fMRI Data Collected during Cinema Viewing. PLoS One 7:e42000. doi:10.1371/journal.pone.0042000.

Pannasch, S and Velichkovsky, BM. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. Vis Cogn 17:1109-1131.

Papageorgiou, C and Poggio, T. (2000). A trainable system for object detection. Int J Comput Vision 38:15–33.

Peelen, MN and Downing, PE. (2007). The neural basis of visual body perception. Nat Rev Neurosci 8:636–648.

Perret, DI, Rolls, ET, and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342.

Peuskens, H, Vanrie, J, Verfaillie, K, and Orban, GA. (2005). Specificity of regions processing biological motion. Eur J Neurosci 21:2864–2875.

Premack, D and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behav Brain Sci 1:515-526.

Pronin, E, Puccio, C, and Ross, L (2002). Understanding misunderstanding: Social psychological perspectives. P. 636–665 in Heuristics and biases: The psychology of intuitive judgment, edited by T Gilovich, D Griffin, and D Kahneman. Cambridge, UK: Cambridge University Press.

Rizzolatti, G and Craighero, L. (2004). The mirror-neuron system. Annu Rev Neurosci 27:169-192.

Rosvall, M and Bergstrom, CT. (2007). An information-theoretic framework for resolving community structure in complex networks. Proc Natl Acad Sci U S A 104:7327–7331.

Russel, JA and Barret, LF. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. J Pers Soc Psychol 76:805-819.

Russell, JA. (1980). A circumplex model of affect. J Pers Soc Psychol 39:1161–1178.

Salmi, J, Roine, U, Glerean, E, Lahnakoski, JM, Nieminen-von Wendt, T, Tani, P, Leppämäki, S, Nummenmaa, L, Jääskeläinen, IP, Carlson, S, Rintahaka, P, and Sams, M. (2013). The brains of high functioning autistic individuals do not synchronize with those of others. NeuroImage: Clinical 3:489–497.

Schippers, MB, Roebroeck, A, Renken, R, Nanetti, L, and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. Proc Natl Acad Sci U S A 107:9388–9393.

Schmälzle, R, Häcker, F, Renner, B, Honey, CJ, and Schupp, HT. (2013). Neural Correlates of Risk Perception during Real-Life Risk Communication. J Neurosci 33:10340–10347.

Schwarzlose, RF, Baker, CI, and Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. J Neurosci 11055–11059:25.

Seifritz, E, Esposito, F, Hennel, F, Mustovic, H, Neuhoff, JG, Bilecen, D, Tedeschi, G, Scheffler, K, and Di Salle, F. (2002). Spatiotemporal Pattern of Neural Processing in the Human Auditory Cortex. Science 297:1706–1708.

Servos, P, Osu, R, Santi, A, and Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. Cereb Cortex 12:772–782.

Shepherd, SV, Steckenfinger, SA, Hasson, U, and Ghazanfar, AA. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. Curr Biol 20:649–656.

Singer, T. (2012). The past, present and future of social neuroscience: a European perspective. NeuroImage 61:437–449.

Sirotin, YB and Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. Nature 457:475–479.

Smith, SM, Fox, PT, KL, Miller, Glahn, DC, Fox, PM, MacKay, CE, Filippini, N, Watkins, KE, Toro, R, Laird, AR, and Beckman, CF. (2009). Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci USA 106:13040–13045.

Smith, SM, Miller, KL, Moeller, S, Xu, J, Auerbach, EJ, Woolrich, MW, Beckmann, CF, Jenkinson, M, Andersson, J, Glasser, MF, Van Essen, DC, Feinberg, DA, Yacoub, ES, and Ugurbil, K. (2012). Temporally-independent functional modes of spontaneous brain activity. Proc Natl Acad Sci U S A 109:3131–3136.

Spiers, H and Maguire, EA. (2007). Decoding human brain activity during real-world experiences. Trends Cogn Sci 11:356–365.

Sporns, O. (2011). Networks of the Brain. Cambridge, Massachusetts, USA: The MIT Press.

Stansbury, DE, Naselaris, T, and Gallant, JL. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. Neuron 79:1–10.

Stephens, GJ, Silbert, LJ, and Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. Proc Natl Acad Sci USA 107:14425–14430.

Stolk, A. (2014). In sync: metaphor, mechanism or marker of mutual understanding? J Neurosci 34:5397–5398.

Thompson, J and Parasuraman, R. (2012). Attention, biological motion, and action recognition. NeuroImage 59:4–13.

Tohka, J, Foerde, K, Aron, AR, Tom, SM, Toga, AW, and Poldrack, RA. (2008). Automatic independent component labeling for artifact removal in fMRI. NeuroImage 39:1227–1245.

van den Heuvel, M, Mandl, R, and Hulshoff Pol, H. (2008). Normalized Cut Group Clustering of Resting-State fMRI Data. PLoS One 3:e2001. doi:10.1371/journal.pone.0002001.

Van Essen, DC. (2005). A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. NeuroImage 28:635–662.

Wang, HX, Freeman, J, Merriam, EP, Hasson, U, and DJ, Heegerm. (2012). Temporal eye movement strategies during naturalistic viewing. J Vis 12:doi: 10.1167/12.1.16.

Watson, R, Latinus, M, Charest, I, Crabbe, F, and Belin, P. (2014). People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. Cortex 50:125–136.

Viinikainen, M, Jääskeläinen, IP, Alexandrov, Y, Balk, MH, Autti, T, and Sams, M. (2010). Nonlinear relationship between emotional valence and brain activity: Evidence of separate negative and positive valence dimensions. Hum Brain Mapp 31:1030–1040.

Wilson-Mendenhall, CD, Barrett, LF, Simmons, WK, and Barsalou, LW. (2011). Grounding emotion in situated conceptualization. Neuropsychologia 49:1105–1127.

Võ, ML-H, Smith, TJ, Mital, PK, and Henderson, JM. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. J Vis 12(13):3, doi: 10.1167/12.13.3.

Yarbus, AL. (1967). Eye movements and vision. New York: Plenum Press.

Publications

Aalto-DD 183/2014

Understanding the brain basis of the wide variety of skills needed for seamless social interactions with other people is one of the most important goals of social cognitive neuroscience. However, it has remained unclear how the elementary processes of social interaction that have been studied so far generalize to complex naturalistic settings where multiple social cues have to be dynamically tracked at the same time. The studies presented in this dissertation employ movies to depict real-life-like social interactions and map the brain systems that participate in the processing of different aspects of the stimuli as well as higher-level brain processes that guide our interpretation of the movie events. The consistency of the findings presented here demonstrates the feasibility of studying brain responses to simple stimulus features, social movie content as well as high-level perspective-taking tasks during very rich naturalistic audiovisual stimulus conditions.

ISBN 978-952-60-5957-0 (printed)
ISBN 978-952-60-5958-7 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Biomedical Engineering and Computational Science

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS