From the Institute of Neuroscience and Medicine,
Brain & Behaviour (INM-7),
at Research Centre Jülich, Jülich, Germany

Exploring naturalistic stimuli as an alternative to a traditional functional localizer

Dissertation

to obtain the academic title of Doctor of Philosophy (Ph.D.) in Medical Sciences
from the Faculty of Medicine at Heinrich Heine University Düsseldorf

submitted by

Christian Olaf Häusler

(2023)

As an inaugural dissertation printed by permission of the
Faculty of Medicine at Heinrich Heine University Düsseldorf

signed:
Dean: Prof. Dr. med. Nikolaj Klöcker
Examiners: Prof. Dr. rer. nat. Michael Hanke, Prof. Dr. rer. nat. Gerhard Jocham

“Gladius ultor noster! Pectus amico, cuspis hosti!”

— old studential motto

Parts of this work have been published:

Halchenko, Y. O., Meyer, K., Poldrack, B., Solanky, D. S., Wagner, A. S., Gors, J., MacFarlane, D., Pustina, D., Sochat, V., Ghosh, S. S., Mönch, C., Markiewicz, C. J., Waite, L., Shlyakhter, I., de la Vega, A., Hayashi, S., Häusler, C. O., Poline, J.-P., Kadelka, T., Skytén, K., Jarecka, D., Kennedy, D., Strauss, T., Cieslak, M., Vavra, P., Ioanas, H.-I., Schneider, R., Pflüger, M., Haxby, J. V., Eickhoff, S. B., & Hanke, M. (2021). DataLad: distributed system for joint management of code, data, and their relationship. Journal of Open Source Software, 6(63), 3262. doi: 10.21105/joss.03262.

Häusler, C. O., & Hanke, M. (2021). A studyforrest extension, an annotation of spoken language in the German dubbed movie “Forrest Gump” and its audio-description. F1000Research, 10(54). doi: 10.12688/f1000research.27621.1.

Häusler, C. O., Eickhoff, S. B., & Hanke, M. (2022). Processing of visual and non-visual naturalistic spatial information in the “parahippocampal place area”. Scientific Data, 9(1). doi: 10.1038/s41597-022-01250-4.

Zusammenfassung

Functional localizers sind fMRT-Experimente, die die funktionelle Neuroanatomie von Individuen charakterisieren sollen. Allerdings verwenden diese Paradigmen selektiv ausgewählte, experimentell streng kontrollierte Reize, stützen sich auf die kooperative Mitarbeit des jeweiligen Individuums, und können üblicherweise nur eine Domäne von Gehirnfunktionen abbilden. Im Gegensatz dazu bieten naturalistische Stimuli wie Filme oder auditorische Erzählungen ein fesselndes, aufgabenfreies Paradigma, das der Komplexität und Vielfalt des alltäglichen Erlebens näher kommt, und dadurch eine Vielzahl von Gehirnfunktionen abbilden könnte. Der Schwerpunkt dieser Dissertation richtet sich auf das "Parahippocampal Place Area" (PPA), ein funktionelles Areal höherer visueller Wahrnehmung, das erhöhte hämodynamische Aktivität aufweist, wenn Studienteilnehmer Bilder von Landschaften oder Orientierungspunkten, im Gegensatz Bildern von Gesichtern oder Werkzeugen, betrachten. Unter Berücksichtigung der Prinzipien offener, transparenter und reproduzierbarer Wissenschaft untersucht die Arbeit mit zwei methodischen Ansätzen, ob ein Film und eine auditorische Erzählung einen visuellen Localizer ersetzen könnten.

Als ersten Ansatz führten wir eine modellgetriebene Analyse hämodynamischer Aktivität während des Films "Forrest Gump" und seiner Audiodeskription durch, die die Hauptstimuli des öffentlich zugänglichen StudyForrest-Datensatzes sind (`studyforrest.org`). Zunächst wurde eine umfassende Annotation der im Film und in der Audiodeskription vorkommenden gesprochenen Sprache erstellt, um die Grundlage für die Modellierung hämodynamischer Antworten zu schaffen und das StudyForrest-Projekt als öffentlich zugängliche Wissenschaftsressource zu erweitern. Anschließend führten wir eine Analyse mit dem allgemeinen linearen Modell (GLM) durch, um die PPA in Personen zu lokalisieren, die zuvor bereits an einem Localizer-Experiment teilgenommen hatten. Die Ergebnisse legen nahe, dass eine modellgetriebene Analyse auf der Grundlage von Annotationen eines Films oder eines ausschließlich auditorischen naturalistischen Stimulus verwendet werden kann, um ein visuelles Areal in Individuen lokalisieren.

Als zweiten Ansatz untersuchte die Arbeit ein neuartiges, datengetriebenes Verfahren für eine funktionelle Lokalisierung, das es ermöglicht die Lage der PPA in einem Individuum zu schätzen, indem es sich mittels *functional alignments* Daten anderer Individuen zu Nutze macht. Unter Verwendung des *shared response models* (SRM) erstellten wir einen *common functional space* (CFS) und individuelle Transformationen, um funktionale Daten von Individuen in einer Referenzgruppe durch den CFS in den Gehirnraum des zu untersuchenden Individuums zu projizieren. Darüber hinaus untersuchten wir die Beziehung zwischen der Menge funktioneller Daten, die für das Alignment eines Individuums mit dem CFS genutzt wurden und der sich anschließenden Schätzleistung. Die Ergebnisse legen nahe, dass eine auditorische Erzählung grundsätzlich dazu verwendet werden kann, um die neuroanatomische Position eines visuelles Areal wie der PPA zu schätzen. Darüber hinaus zeigen die Ergebnisse, dass Daten eines 15-minütigen Scans, während dem ein Individuum einen Film schaut, hinreichend sind, um Gehirnmuster genauer zu schätzen als ein Verfahren, das auf einem anatomischen Alignment beruht.

Die Arbeit zeigt jedoch auch Hindernisse in der Entwicklung eines multifunktionalen naturalistischen Localizers auf. Daten naturalistischer Stimuli stellen eine Herausforderung für modellgetriebene Analysen insofern da, weil sie physiologische und statistische Modellannahmen strapazieren. Außerdem zielen traditionelle Localizer darauf ab, die interindividuelle Variabilität zu minimieren und funktionelle Areale in allen gesunden Personen reliabel zu lokalisieren, wohingegen naturalistische Stimuli höhere interindividuelle Variabilität zulassen. Daher hängt das Potenzial eines naturalistischen Stimulus, einen oder mehrere traditionelle Localizer zu ersetzen, von weiteren Entwicklungen ab, die die statistischen und methodologischen Herausforderungen angehen. Jedoch könnte ein datengetriebener Ansatz, der auf functional alignment beruht, möglicherweise einen naturalistischen Stimulus von ähnlicher Dauer eines traditionellen Localizers verwenden, um die Ergebnisse vieler funktioneller Localizer schätzen.

Summary

Functional localizers are fMRI experiments that aim to characterize the functional neuroanatomy on the level of individuals. However, these paradigms employ selectively sampled, tightly controlled stimuli, rely on an individual's compliance, and can typically map only one domain of brain functions. In contrast, naturalistic stimuli, such as movies or auditory narratives, provide an engaging, task-free paradigm that more closely resembles the complexity and richness of real-life experiences, and sample a wide range of brain functions. This dissertation focuses on the "parahippocampal place" (PPA), a high-level visual area, that exhibits increased hemodynamic activity when participants view images of landscapes or landmarks, as opposed to other stimuli, such as faces or tools. Following the principles of open, transparent, and reproducible science, the thesis explores whether a movie and an auditory narrative could replace a visual localizer in two methodological ways.

As the first approach, we performed a model-driven analysis of hemodynamic activity during the movie "Forrest Gump" and its audio-description, which are the core stimuli of the publicly accessible studyforrest dataset (`studyforrest.org`). An exhaustive annotation of speech occurring in the movie and audio-description was created to establish the foundation for modeling hemodynamic responses and to extend the studyforrest project as an open science resource. Subsequently, we performed a general linear model (GLM) analysis to localize the PPA, which had previously been identified in the same group of participants using a visual localizer. Results suggest that a model-driven analysis based on annotations of a movie or an exclusively auditory naturalistic stimulus can be used to localize a visual area on an individual level.

As the second approach, we explored a novel functional alignment procedure that allows estimating the location of the PPA in an individual by leveraging data collected from of a reference group. Using a shared response model (SRM), we created a common functional space (CFS) and subject-specific transformations to project functional data from the reference through the CFS into an individual's brain space. Additionally, we investigated the relationship between the quantity of data used for functional alignment and subsequent estimation performance. Results suggest that an auditory narrative can, in principle, be used to estimate a visual area such as the PPA. Moreover, 15 minutes of functional scanning during movie watching can generate a sufficient amount of data to estimate brain patterns more accurately than a procedure based on anatomical alignment.

The thesis also highlights obstacles in the pursuit of developing a multi-functional naturalistic localizer. Applying a model-driven analysis to naturalistic stimuli is challenging, as these stimuli stress physiological and statistical assumptions. Moreover, traditional localizers aim to minimize interindividual variability and reliably localize functional areas in all healthy individuals, whereas naturalistic stimuli allow for more variability. Therefore, the potential of a naturalistic stimulus to replace one or multiple traditional localizers relies on further developments that address the statistical and methodological challenges. Nevertheless, a data-driven approach based on functional alignment using a naturalistic stimulus of similar duration to that of one traditional localizer could potentially estimate the results of many functional localizers.

List of Abbreviations

BOLD	blood oxygen level-dependent
CAS	common anatomical space
CFS	common functional space
EBA	extrastriate body area
EEG	electroencephalography
FFA	fusiform face area
FoV	field of view
GLM	general linear model
fMRI	functional magnetic resonance imaging
MEG	magnetoencephalography
OPA	occipital place area
PCA	principal component analysis
PET	positron emission tomography
PPA	parahippocampal place area
ROI	region of interest
RSC	retrosplenial complex
TR	time of repetition
SRM	shared response model

Contents

1 Introduction

"A remarkable feature of the vertebrate brain is the anatomical specialization of cortical regions for the processing of different types of information. Since the late 19th century, it has been recognized that restricted lesions of the human brain result in location-specific sensory, motor or cognitive deficits" (Cohen & Bookheimer, 1994, p. 268). Contemporary *human brain mapping* (e.g., Raichle, 2009) investigates in vivo the brain's topographic organization (e.g., Eickhoff, Constable, & Yeo, 2018) by specifying "in as much detail as possible the localisation of function in the human brain" (Savoy, 2001, p. 10). Since the early 1990s, brain imaging studies have used functional magnetic resonance imaging (fMRI) to measure blood oxygen level-dependent (BOLD) activity as a proxy for neural activation. Replicated findings in the domain of high-visual perception, for example, suggest that category-selective brain regions, such as the parahippocampal place area (R. Epstein, Harris, Stanley, & Kanwisher, 1999; R. A. Epstein & Kanwisher, 1998) or the fusiform face area (Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006), exhibit significantly increased BOLD activity correlated with a "preferred" (Op de Beeck, Haushofer, & Kanwisher, 2008, p. 123) stimulus class. The parahippocampal place area (PPA) responds more strongly when participants are viewing images of landscapes or landmarks compared to images of, e.g., faces; vice versa, the fusiform face area (FFA) responds more strongly when participants are viewing images of faces compared to images of scenes.

However, most studies in functional imaging research have averaged data across study participants for practical (e.g., limited scan time per subject), statistical reasons (e.g., improved signal-to-noise ratio), or to generalize from study participants to a broader (neurologically healthy) population. While studies that average data across study participants may draw population-level inferences, they also risk losing individual-level information and may capture only the "common denominator" (Pinel et al., 2007, p. 2) of functional patterns. Yet, in order advance human brain mapping towards clinical applications that assess health and disease, fMRI data need to be interpreted on the level of individuals (Dubois & Adolphs, 2016; Wegrzyn et al., 2018).

1.1 Functional localization

Functional localizers (s. Friston, Rotshtein, Geng, Sterzer, & Henson, 2006; Saxe, Brett, & Kanwisher, 2006, for reviews) are fMRI experiments that aim to characterize the *topography* (i.e. the location, size, and shape) of functional areas whose hemodynamic activity correlates with perceptual processes, such as the perception of object categories (Kanwisher et al., 1997), or cognitive processes, such as theory of mind (Spunt & Adolphs, 2014). Functional localizers are frequently used as a separate experiment to identify a subject-specific functional region of interest (ROI) and reduce the number of voxels investigated in the main experiment (Poldrack, 2007; Saxe et al., 2006). Additionally, they may be employed as a diagnostic tool before neurosurgery (cf. Silva, See, Essayed, Golby,

& Tie, 2018; Szaflarski et al., 2017). However localizers are designed to maximize detection power, use carefully selected, simplified and tightly controlled stimuli presented in a block-wise manner, and are often accompanied by a task to keep study participants attentive. Consequently, one localizer paradigm can typically map only one domain of brain functions. In order to map many different processes despite limited resources, researchers have developed more time-efficient, multi-functional *localizer batteries* (e.g., Barch et al., 2013; Drobyshevsky, Baumann, & Schneider, 2006; Pinel et al., 2007; Pinho et al., 2020, 2018). For example, Pinel et al. (2007) uses a range of dedicated stimuli and specific tasks in a 5-minute routine to map processes of "auditory and visual perception, motor actions, reading, language comprehension, and mental calculation" (Pinel et al., 2007, p. 15). Nonetheless, current paradigms face two challenges. From a theoretical standpoint, localizers rely on selectively sampled, tightly controlled stimuli presented in blocks, and do not resemble how we perceive the real world outside of the laboratory in daily life. From a practical standpoint, localizers rely on participants' comprehension of task instructions and compliance during the scanning session, which can be difficult to achieve in clinical or pediatric populations (Eickhoff, Milham, & Vanderwal, 2020; Vanderwal, Eilbott, & Castellanos, 2019; Vanderwal, Kelly, Eilbott, Mayes, & Castellanos, 2015).

1.2 Naturalistic stimuli

Because a major goal of neuroscience is not to reveal how the brain responds to blocks of stimuli presented in a laboratory setting, but how the brain processes information during everyday perception, *naturalistic stimuli* are gaining popularity in neuroimaging. Naturalistic stimuli are "a class of stimuli that aim to evoke more naturalistic patterns of neural responses than traditional controlled artificial stimuli. Naturalistic paradigms are typically complex and dynamic, and longer in duration than many conventional stimuli." (Vanderwal et al., 2019, p. 2). Currently, the most popular naturalistic stimuli in neuroscience are movies and auditory narratives (cf. Sonkusare, Breakspear, & Guo, 2019, for an introduction) that provide a time-locked event structure, sample a broad range of brain states, and engage multiple perceptual and cognitive systems in parallel (Haxby, Gobbini, & Nastase, 2020).

Naturalistic stimuli have several advantages over traditional paradigms. From a theoretical standpoint, naturalistic stimuli promise an increased extent of both subtypes of external validity, namely population validity and ecological validity (Bracht & Glass, 1968). *Population validity* refers to the extent to which inferences drawn from an experiment's results may generalize from the experiment's sample of subjects and stimuli to the total population of potential subjects and stimuli (Bracht & Glass, 1968; Westfall, Nichols, & Yarkoni, 2016). *Ecological validity* refers to the extent to which inferences drawn from an experiment's results may generalize from the experiment's setting, stimuli, and task to nonexperimental situations (Bracht & Glass, 1968; Orne, 1962; Schmuckler, 2001). Naturalistic stimuli promise an increased population validity of stimuli because the *stimulus features* (i.e. the variables or stimulus classes) that are embedded in a naturalistic stimulus represent a more random sample from the total population of stimuli that might have been used (Westfall et al., 2016). Naturalistic stimuli also promise a higher ecological validity because they more closely resemble how we experience our environment outside of the scanner bore (Hasson & Honey, 2012).

Audio-visual movies and spoken narratives have been used during fMRI (s. Hamilton & Huth, 2020; Hasson et al., 2008; Jääskeläinen, Sams, Glerean, & Ahveninen, 2021; Saarimäki, 2021, for reviews), and electroencephalography (EEG) or magnetoencephalography (MEG) data acquisition (s. Alday, 2019; Kandylaki & Bornkessel-Schlesewsky, 2019, for reviews). Previous studies have shown that watching a movie (Hasson, Malach, & Heeger, 2010; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) or listening to a narrative (Lerner, Honey, Silbert, & Hasson, 2011; S. M. Wilson, Molnar-Szakacs, & Iacoboni, 2008) reliably synchronizes spatiotemporal responses across multiple subjects in a large part of the brain compared to, for example, an unedited video of a concert taken from a fixed viewpoint (Hasson et al., 2008). Importantly, a pioneering study by Bartels and Zeki (2004) demonstrated results at a group level, indicating that functional specialization of cortical areas is preserved during naturalistic stimulation. Hence, on an individual level, a naturalistic stimulus could, in theory, also be used as a more life-like paradigm to replace traditional localizer paradigms.

From a practical standpoint, naturalistic stimuli require minimal instructions given by the experimenter, and place minimal task demands on study participants, who can simply enjoy the movie or audio story. In addition, movies and narratives provide an interesting and engaging stimulation that can put participants at ease in the uncomfortable and noisy MRI scanner. Consequently, naturalistic stimuli promise improved data quality due to reduced fatigue and head movement, particularly in children (Vanderwal et al., 2015), and possibly psychiatric (Eickhoff et al., 2020) or elderly individuals.

Nevertheless, naturalistic stimuli also have disadvantages over traditional paradigms. First, the majority of naturalistic stimuli used in neuroimaging have been created by professional production companies for commercial purposes, rather than for research purposes. The temporal structure of stimulus features embedded in professionally produced naturalistic stimuli is fixed, and thus reproducible, but initially not explicitly known. Consequently, modeling brain activity correlating with stimulus features embedded in a stimulus' time course is challenging (Saarimäki, 2021; Simony & Chang, 2020) because such models, like a traditional *general linear model (GLM)*, rely on the stimulus features being annotated. The lack of extensive annotations has resulted in a "usage bottleneck" (Aliko, Huang, Gheorghiu, Meliss, & Skipper, 2020, p. 16) and may be the main reason why explicit models of embedded stimulus features are "notoriously" (Richard, Martin, Pinho, Pillow, & Thirion, 2019, p. 1), if not "prohibitively" (Nastase, Gazzola, Hasson, & Keysers, 2019, p. 676) difficult to construct. Second, considering practical and monetary constraints in a clinical context, presenting a full feature film typically lasting 90 to 120 minutes is inappropriate as an individual diagnostic procedure.

1.3 Ethics protocol

All data acquisitions were jointly approved by the Ethics Committee of Otto-von-Guericke-University of Magdeburg, Germany (approval reference 37/13).

1.4 Aims of thesis

This dissertation aimed to explore—while adhering to the principles of open, transparent, and reproducible science—whether a movie and the movie's audio-description that was produced for a visually impaired audience could, in principle, replace a traditional localizer paradigm. As a proof of concept, the dissertation focuses on the PPA that exhibits increased hemodynamic activity when participants view images of landscapes, buildings or landmarks, as opposed to, for example, images of faces or tools (s. Aminoff, Kveraga, & Bar, 2013; R. A. Epstein & Vass, 2014, for reviews). To the author's knowledge, only one study (Aziz-Zadeh et al., 2008) compared hemodynamic activity levels in the PPA that were correlated with different categories presented in spoken sentences. Despite mixed results, the findings by Aziz-Zadeh et al. (2008) suggest that the PPA does not solely respond to visually presented scene-related stimuli.

The thesis assessed the feasibility of using the movie and audio-description as alternatives to a visual localizer in two ways. As the first approach, we modeled hemodynamic activity based on annotated stimulus features that are embedded in the audio-visual movie and the movie's audio-description (each lasting ≈120 m). We then created GLM t-contrasts using the modeled hemodynamic time courses (i.e. regressors) to locate the PPA, which was previously identified in the same group of participants by a conventional block-design functional localizer using static images (Sengupta et al., 2016). However, conducting a two-hour long fMRI scan session may not be desirable or feasible due to potential compliance issues or constraints on time and resources. Therefore, we also explored a second approach to localize the PPA in an individual that leverages data collected from an independent sample of other individuals (i.e. data from a *reference group*). To address the issue of anatomical variability across persons, previous studies (Frost & Goebel, 2012; Rosenke, van Hoof, van den Hurk, Grill-Spector, & Goebel, 2021; Weiner et al., 2018; Zhen et al., 2017) estimated the location of a participant's functional area by performing an *anatomical alignment* that relies on anatomical scans. However, the anatomical location of functional regions varies between individuals (Benson, Butt, Brainard, & Aguirre, 2014; Coalson, Van Essen, & Glasser, 2018; Frost & Goebel, 2012; Langers, 2014; Natu et al., 2021; Rosenke et al., 2021; Wang, Mruczek, Arcaro, & Kastner, 2015; Weiner et al., 2014). To address the issue of functional-anatomical variability across persons, we therefore performed a *functional alignment* (cf. Bazeille, Dupre, Richard, Poline, & Thirion, 2021; Haxby, Guntupalli, Nastase, & Feilong, 2020, for reviews), and investigated whether we can estimate the results of t-contrasts (i.e. statistical Z-maps) that we created in the previous studies to localize the PPA. Considering that functional alignment relies on functional scans, we also evaluated the relationship between the quantity of functional data used for functional alignment and subsequent estimation performance.

1.4.1 Open, transparent, and reproducible science

In the last decade, there has been a growing awareness that the results of scientific publications are not reproducible or general scientific findings are not replicable. This prompted some authors to refer to the sciences as being in the midst of a "reproducibility crisis" or "replication crisis" (Baker, 2016; Nosek et al., 2022; Plesser, 2018; Stupple, Singerman, & Celi, 2019). "A study is reproducible if all of the code and data used to generate the numbers and figures in the paper are available and exactly produce the published results" (Leek & Jager, 2017, p. 111). A study is replicable if the same analysis of data from an equivalent experiment yields consistent results (Dubois & Adolphs, 2016; Leek & Jager, 2017). The first objective in the context of open science was to meet the requirements of open, shared, accessible, and transparent science (cf. Fecher & Friesike, 2014; Watson, 2015) as well as a reproducible and replicable research project: to achieve this, the dissertation follows the guidelines and best practices for (a) coding and scientific computing (G. Wilson et al., 2014), (b) procedures and data analyses (Nichols et al., 2017; Poldrack et al., 2017; Poldrack, Huckins, & Varoquaux, 2019), and (c) sharing code, created data, and results (Eglen et al., 2017; Nichols et al., 2017; Pernet & Poline, 2015). The second objective in the context of open science was to reuse a publicly and freely available dataset for a new research question that was not anticipated at the time the data were released, to extend the dataset, and to generate novel findings to be published in open-access journals.

The present work is built upon fMRI data that are part of the *studyforrest* project (`studyforrest.org`). The core of this project are two-hour long BOLD fMRI scans of participants watching the movie "Forrest Gump" (Zemeckis, 1994) and listening to the movie's audio-description that was created for a visually impaired audience by adding a narrator to the movie's audio track. Since its first publication in 2014 (Hanke et al., 2014a), the studyforrest project has served as a resource of raw (and preprocessed) data for international working groups to conduct and publish independent, peer-reviewed research (cf. `studyforrest.org/publications.html`). All data, materials, custom code, analysis steps, and results used or created over the course of this dissertation are version-controlled—i.e. each change of data is logged and documented—and accessible in standardized *DataLad* (`datalad.org`; Halchenko et al., 2021) datasets. Data analysis pipelines are designed for automated processing and implemented in freely available and, where possible, open-source software. Among potential software packages, the tools were chosen that offered the most solid documentation, and a large community of developers and maintainers to ensure long-term support. All custom code is written in open-source programming languages (Python and Bash), version-controlled, documented, and publicly released. Therefore, all executed steps from downloading input data to visualizing the results can be rerun to improve reproducibility of current results and to facilitate replication of findings on other datasets. Finally, because "nature abhors a paywall" (DuPre, Hanke, & Poline, 2020), publications describing generated data, methodological choices, analysis steps, and results are published in open-access journals.

1.4.2 Specific objectives and hypotheses

During the pre-alpha stage of DataLad, I contributed to its development by testing its features on a real-world project, evaluating its user interface and documentation, and providing feedback to the software engineering team. This work laid the technical foundation for conducting a transparent and reproducible research project, and resulted in co-authorship of the software's accompanying publication (s. Halchenko et al., 2021).

In Häusler and Hanke (2021), we created and validated an annotation of speech spoken in the movie, as well as its audio-description, with two objectives in mind. The first objective was to establish the groundwork for creating models of hemodynamic activity in response to the movie and the audio-description in Häusler, Eickhoff, and Hanke (2022). The second objective was to create an exhaustive annotation of speech in order to extend the studyforrest project and supplement the previously published annotations of portrayed emotions (Labs et al., 2015), perceived emotions (Lettieri et al., 2019), and cuts and locations depicted in the movie (Häusler & Hanke, 2016). In order to validate the annotation's quality, we conducted a canonical GLM analysis of modeled hemodynamic activity based on information drawn from the annotation. Regressors correlating with speech-related events were contrasted with a regressor correlating with events without speech. We hypothesized that results would reveal clusters of increased activity in brain regions known to be involved in processing spoken language. The analysis revealed statistically significant increased hemodynamic activity in a bilateral cortical network replicating results of previous studies that used tightly controlled stimuli (s. Friederici, 2011; Hickok & Poeppel, 2007; Price, 2012, for reviews), and studies that used data-driven methods to analyze fMRI data from auditory naturalistic stimulation (Honey, Thompson, Lerner, & Hasson, 2012; Lerner et al., 2011; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). The results of the validation encouraged us to publish the annotation as an extension of the studyforrest dataset, and use it as the foundation to be adapted to our specific needs in Häusler et al. (2022).

In Häusler et al. (2022), our objective was to investigate the possibility of localizing the PPA using GLM t-contrasts based annotations of the naturalistic stimuli. For the analysis of the movie, we relied on a previously published annotation of movie cuts and the depicted location after each cut (Häusler & Hanke, 2016). For the analysis of the audio-description, we extended the annotation of speech by further annotating nouns that the audio-description's narrator uses to describe the missing visual content. We hypothesized that a mass-univariate GLM analysis would reveal increased hemodynamic activity in medial temporal regions that were functionally identified as the PPA in the same set of participants employing a traditional block-design functional localizer (Sengupta et al., 2016). On a group-average level, the results demonstrate that increased activation in the PPA during the perception of static images generalizes to the perception of spatial information embedded in an audio-visual movie and exclusively auditory naturalistic stimulus. Results add evidence (cf. Bartels & Zeki, 2004) that functional specialization of cortical areas is preserved during naturalistic stimulation. On an individual level, the analysis of the movie yielded bilateral clusters of increased hemodynamic activity in the PPA of five participants and a unilateral cluster in seven participants. The analysis of the audio-description revealed bilateral clusters in nine participants and a unilateral cluster in one participant. These results imply that a model-driven GLM analysis based on a naturalistic stimulus' annotation can be used to localize functional areas in individuals.

Conducting a two-hour long fMRI scan session may not be desirable or feasible due to potential compliance issues or constraints on time and resources. In Chapter 5, we therefore aimed to explore an alternative approach to identify the PPA in an individual by leveraging data from a reference group. To address the challenge of functional-anatomical variability across individuals, we employed a functional alignment approach using the shared response model (SRM) (Chen et al., 2015). This approach allowed us to predict an individual's results of t-contrasts, created in previous studies using time series from the visual localizer, movie, and audio-description, by projecting the results of persons in the reference group from their respective voxel space through a common functional space (CFS) into the individual's voxel space. Following a leave-one-subject-out cross-validation, we split the dataset repeatedly in a set of one *test subject* and $N-1$ *training subjects* (the reference group). We then calculated a CFS and subject-specific transformations based on the training subjects' concatenated response time series of the visual localizer, the movie, and the audio-description. We assessed the prediction performance of each paradigm by aligning the test subject's response time series from each paradigm separately with the corresponding time points within the CFS. In an ideal scenario, only a part of a naturalistic stimulus would serve as a "diagnostic" run to align a new individual with a CFS and then estimate the results of other paradigms. Therefore, we also aimed to explore the relationship between the quantity of data of each paradigm used for aligning a test subject with the CFS and subsequent estimation performance. Results suggest that an auditory narrative can in principle be used to estimate a visual area such as the PPA. Moreover, 15 minutes of functional scanning during movie watching can generate a sufficient amount of data to estimate brain patterns with higher fidelity than a procedure based on anatomical alignment.

2 DataLad: distributed system for joint management of code, data, and their relationship

This part of the dissertation has been published:

Halchenko, Y. O., Meyer, K., Poldrack, B., Solanky, D. S., Wagner, A. S., Gors, J., MacFarlane, D., Pustina, D., Sochat, V., Ghosh, S. S., Mönch, C., Markiewicz, C. J., Waite, L., Shlyakhter, I., de la Vega, A., Hayashi, S., Häusler, C. O., Poline, J.-P., Kadelka, T., Skytén, K., Jarecka, D., Kennedy, D., Strauss, T., Cieslak, M., Vavra, P., Ioanas, H.-I., Schneider, R., Pflüger, M., Haxby, J. V., Eickhoff, S. B., & Hanke, M. (2021). DataLad: distributed system for joint management of code, data, and their relationship. Journal of Open Source Software, 6(63), 3262. doi: `10.21105/joss.03262`.

Abstract

DataLad is a Python-based tool for the joint management of code, data, and their relationship, built on top of a versatile system for data logistics (`git-annex`) and the most popular distributed version control system (`Git`). It adapts principles of open-source software development and distribution to address the technical challenges of data management, data sharing, and digital provenance collection across the life cycle of digital objects. DataLad aims to make data management as easy as managing code. It streamlines procedures to consume, publish, and update data, for data of any size or type, and to link them as precisely versioned, lightweight dependencies. DataLad helps to make science more reproducible and FAIR (Wilkinson et al., 2016). It can capture complete and actionable process provenance of data transformations to enable automatic re-computation. The DataLad project (`datalad.org`) delivers a completely open, pioneering platform for flexible decentralized research data management (RDM) (Hanke et al., 2021). It features a Python and a command-line interface, an extensible architecture, and does not depend on any centralized services but facilitates interoperability with a plurality of existing tools and services. In order to maximize its utility and target audience, DataLad is available for all major operating systems, and can be integrated into established workflows and environments with minimal friction.

3 A studyforrest extension, an annotation of spoken language in the German dubbed movie “Forrest Gump” and its audio-description

This part of the dissertation has been published:

Häusler, C. O., & Hanke, M. (2021). A studyforrest extension, an annotation of spoken language in the German dubbed movie “Forrest Gump” and its audio-description. F1000Research, 10(54). doi: `10.12688/f1000research.27621.1`.

Abstract

Here we present an annotation of speech in the audio-visual movie “Forrest Gump” and its audio-description for a visually impaired audience, as an addition to a large public functional brain imaging dataset (`studyforrest.org`). The annotation provides information about the exact timing of each of the more than 2500 spoken sentences, 16,000 words (including 202 non-speech vocalizations), 66,000 phonemes, and their corresponding speaker. Additionally, for every word, we provide lemmatization, a simple part-of-speech-tagging (15 grammatical categories), a detailed part-of-speech tagging (43 grammatical categories), syntactic dependencies, and a semantic analysis based on word embedding which represents each word in a 300-dimensional semantic space. To validate the dataset’s quality, we build a model of hemodynamic brain activity based on information drawn from the annotation. Results suggest that the annotation’s content and quality enable independent researchers to create models of brain activity correlating with a variety of linguistic aspects under conditions of near-real-life complexity.

4 Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area"

This part of the dissertation has been published:

Häusler, C. O., Eickhoff, S. B., & Hanke, M. (2022). Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area". Scientific Data, 9(1). doi: `10.1038/s41597-022-01250-4`.

Abstract

The "parahippocampal place area" (PPA) in the human ventral visual stream exhibits increased hemodynamic activity correlated with the perception of landscape photos compared to faces or objects. Here, we investigate the perception of scene-related, spatial information embedded in two naturalistic stimuli. The same 14 participants were watching a Hollywood movie and listening to its audio-description as part of the open-data resource `studyforrest.org`. We model hemodynamic activity based on annotations of selected stimulus features, and compare results to a block-design visual localizer. On a group level, increased activation correlating with visual spatial information occurring in the movie is overlapping with a traditionally localized PPA. Activation correlating with semantic spatial information occurring in the audio-description is more restricted to the anterior PPA. On an individual level, we find significant bilateral activity in the PPA of nine individuals and unilateral activity in one individual. Results suggest that activation in the PPA generalizes to spatial information embedded in a movie and an auditory narrative, and may call for considering a functional subdivision of the PPA.

5 Assessing the quantity of data for functional alignment to estimate responses in the “parahippocampal place area”

Abstract

Block-design functional localizers are traditionally used to identify functional brain areas in individuals. However, these paradigms rely on selectively sampled, tightly controlled stimuli and participant compliance, and are typically limited to mapping a single domain of brain functions. We explored an alternative procedure to identify functional areas, and investigated whether the results from our previous studies that used a localizer, a movie or the movie’s audio-description to localize the “parahippocampal place area” (PPA), a high-visual area, could be estimated in an individual by leveraging data from others individuals. To address the challenge of functional-anatomical variability across persons, our procedure utilizes the shared response model (SRM) to perform a functional alignment of study participants with a common functional space (CFS) that we derived from the concatenated responses to the localizer, the movie and its audio-description. In particular, we assess the relationship between the estimation performance and the amount of data from each of the three paradigms used to align an individual with the CFS. Our results reveal that data from the audio-description allow to estimate the results of the visual paradigms, however at the expense of a long scanning session. Moreover, results indicate that 15 minutes of movie watching provide a sufficient amount for our functional alignment procedure to more accurately estimate the results of the visual localizer than a procedure based on anatomical alignment. This opens up the possibility of estimating results from many functional localizers using a movie of similar duration to that of one localizer.

5.1 Introduction

In the domain of high-visual perception, functionally defined category-selective brain regions, such as the parahippocampal place area (PPA) (R. A. Epstein & Kanwisher, 1998), the fusiform face area (FFA) (Kanwisher et al., 1997), or the extrastriate body area (EBA) (Downing, Jiang, Shuman, & Kanwisher, 2001) exhibit significantly increased blood oxygen level-dependent (BOLD) activity correlated with a "preferred" (Op de Beeck et al., 2008, p. 123) stimulus category. While the topographies, i.e. the location, size and shape, of these category-selective areas are similarly distributed across individuals, their exact topographies vary interindividually (Frost & Goebel, 2012; Rosenke et al., 2021; Zhen et al., 2017, 2015). To identify the topography of functional areas in individuals, block-design *functional localizer* paradigms are traditionally used that contrast modeled hemodynamic responses correlating with the corresponding stimulus category, such as landscapes, faces, or bodies. Functional localizers are designed to maximize detection power and thus limited to mapping only one domain of brain functions, such as category-selective regions (Stigliani, Weiner, & Grill-Spector, 2015), retinotopic visual areas (Wang et al., 2015), theory of mind (Spunt & Adolphs, 2014), or semantic processes (Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Fernández et al., 2001). However, when mapping multiple functional domains in a limited amount of time is desired, the approach "one paradigm for one domain of functions" becomes impractical. To address this issue, researchers have attempted to create time-efficient, multi-functional *localizer batteries* (e.g., Barch et al., 2013; Drobyshevsky et al., 2006; Pinel et al., 2007). Nevertheless, the diagnostic quality of localizer paradigms heavily depends on a participant's comprehension of the task instructions and general compliance, a criteria that can be difficult to meet in clinical or pediatric populations (Eickhoff et al., 2020; Vanderwal et al., 2019).

In a previous study (Häusler et al., 2022), we demonstrated that a functionally defined region such as the PPA can be localized using a general linear model (GLM) that is based on the annotated temporal structure of a two-hour long naturalistic stimulus. However, conducting a two-hour long fMRI scan session may not be desirable or feasible due to potential compliance issues or constraints on time and resources. An alternative approach that addresses the challenges of a lengthy scanning procedure is to localize a functional area in an individual by leveraging data collected from an independent sample of other individuals (i.e. from a *reference group*). Previous studies have estimated the most probable location of a functional area in an individual from a reference group by performing either a volume-based (Zhen et al., 2017, 2015) or a surface-based (Frost & Goebel, 2012; Rosenke et al., 2021; Wang et al., 2015; Weiner et al., 2018) *anatomical alignment*. First, in order to address the issue of anatomical variability across persons, functional data of persons in the reference group are anatomically aligned with (i.e. projected into) a common anatomical space, such as the Montreal Neurological Institute brain atlas (MNI152 atlas; Fonov et al., 2011). Then, data are projected from the common anatomical space into the individual person's voxel space to provide an estimate of a functional region's location. Volume-based anatomical alignment (s. Klein et al., 2009, for a review) aligns voxels to a three-dimensional common anatomical space (e.g., MNI152 atlas; Fonov et al., 2011). Surface-based anatomical alignment (Fischl, Sereno, & Dale, 1999; Yeo et al., 2009) aligns vertices to a two-dimensional common anatomical space (e.g., FreeSurfer's fsaverage template; Fischl, Sereno, Tootell, & Dale, 1999). Whereas volume-based alignment does not account for individual sulcal and gyral folding patterns, surface-based alignment respects interindividual variability of the cortical surface. Consequently, previous studies that compared volume-based and surface-based alignment to estimate the location of functional regions have shown that surface-based alignment reduces interindividual variability and improves estimation performance (Frost & Goebel, 2012; Rosenke et al., 2021; Wang et al., 2015; Weiner et al., 2018). However, even after surface-based alignment, the anatomical location of functional regions varies between individuals (Benson et al., 2014; Coalson et al., 2018; Frost & Goebel, 2012; Langers, 2014; Natu et al., 2021; Rosenke et al., 2021; Wang et al., 2015; Weiner et al., 2014). Frost and Goebel (2012), for example, localized 13 functional areas of the high-level visual cortex and "found a large variability in the degree to which functional areas respect macro-anatomical boundaries" (Frost & Goebel, 2012, p. 1369). The remaining variability indicates that functional areas a not necessarily bound to anatomical landmarks, and reflects the degree of *functional–anatomical correspondence* between a brain function and its underlying anatomical location.

In order to address the issue of functional-anatomical variability across subjects, *functional alignment* algorithms, such as *hyperalignment* (Guntupalli et al., 2016; Haxby et al., 2011) or the *shared response model* (SRM) (Chen et al., 2015; Zhang et al., 2016), have been developed. Whereas anatomical alignment aligns voxels (or vertices) that share the same anatomical location to a common anatomical space, functional alignment aligns voxels (or vertices) that share similar functional properties to a *common functional space* (CFS). Functional alignment algorithms are typically used to compute both a high-dimensional, functional brain template (i.e. the CFS) and subject-specific transformations based on functional data of a study's participants. A subject-specific transformation allows projecting functional data from a subject's three-dimensional voxel space into the CFS. Conversely, the inverse transformation allows projecting data from the CFS into the subject's voxel space (Haxby, Guntupalli, et al., 2020; Kumar et al., 2021). The CFS and transformations are computed (i.e. *trained*) by either maximizing the interindividual similarity of BOLD response time series correlating with a time-locked external stimulation (Chen et al., 2015; Haxby et al., 2011; Sabuncu et al., 2010), or by maximizing the interindividual similarity of connectivity profiles (Feilong, Nastase, Guntupalli, & Haxby, 2018; Guntupalli, Feilong, & Haxby, 2018; Nastase, Liu, Hillman, Norman, & Hasson, 2019). While connectivity-based functional alignment has been shown to be more effective in aligning connectivity profiles, response-based functional alignment is more effective in aligning response time-series (Guntupalli et al., 2018). Although functional alignment algorithms can be applied to fMRI time series data from paradigms employing simplified stimuli, data from naturalistic stimuli provide improved generalizability of the CFS and transformations to novel stimuli or tasks. This is presumably because naturalistic stimuli sample a broader range of brain states (Guntupalli et al., 2016; Haxby et al., 2011).

Consequently, a more recent procedure (e.g., Guntupalli et al., 2016; Haxby et al., 2011; Jiahui et al., 2020) to estimate the most probable location of a functional area in an individual from a reference performs a functional alignment. First, the functional data of individuals in the reference group are anatomically aligned with a common anatomical space (CAS). Second, to address the issue of functional-anatomical variability across individuals, the data are functionally aligned with (i.e. projected into) a CFS. Finally, data are projected from the CFS into the individual's voxel space, serving as an estimate of a functional region's location. For instance, Jiahui et al. (2020) used surface-based hyperalignment to calculate CFSs and transformations based on data from the movie "Grand Budapest Hotel" (≈50 min; time of repetition (TR)=1 s) and the movie "Forrest Gump" (≈120 min; TR=2 s). Jiahui et al. (2020) then estimated t-contrast maps of a visual localizer that aimed to identify the FFA by projecting the t-contrast maps of a reference group through each CFS into an individual's brain anatomy. Results showed that t-contrast maps of the visual localizer correlated more highly with contrast maps that were estimated via hyperalignment than contrast maps that were estimated via surface-based anatomical alignment.

Here again, our focus is on the PPA (e.g., Aminoff et al., 2013; R. A. Epstein & Vass, 2014, for reviews). We investigated whether we can estimate the results of t-contrasts (i.e. statistical Z-maps) created in previous studies that aimed to identify the PPA using response time series from three different paradigms: (a) a classic visual localizer (Sengupta et al., 2016) as the assumed "gold standard" to localize the PPA, (b) a movie (Häusler et al., 2022), and (c) an auditory narrative (Häusler et al., 2022). To obtain *predicted Z-maps* of these *empirical Z-maps*, we employed a volume-based functional alignment approach that utilizes the SRM (Chen et al., 2015; Richard et al., 2019). The SRM is an unsupervised probabilistic latent-factor model that decomposes response time series of participants who have experienced the same stimulus into a CFS of *shared features* (also known as *shared feature space*; Chen et al., 2015) and subject-specific linear transformations. Specifically, the SRM algorithm uses each n^{th} subject's response time series represented as matrix X_n (v voxels by t time points) to compute the CFS C (k shared responses by t time points) and subject-specific transformation matrices W_n (v voxels by k shared responses) with orthonormal columns ($W_n^T W_n = I_k$). The algorithm randomly initializes and fits the transformation matrices over iterations to minimize the error in explaining the participants' data, while also learning the time course of the shared responses (cf. `brainiak.org/tutorials/11-SRM`). Unlike hyperalignment, the number of dimensions of the CFS is not set by the number of voxels, but rather it is determined by the researcher to a number lower than the number of voxels, a procedure that also filters out noise and reduces overfitting (Chen et al., 2015). Each shared feature can be thought of as a weighted sum of many voxels across subjects (Kumar et al., 2021). A subject-specific transformation matrix represents the weight of each voxel in a subject's voxel space on each shared feature, and allows an functional alignment of subjects by projecting hemodynamic responses within the voxels into the k-dimensional CFS.

In contrast to previous studies (Guntupalli et al., 2016; Haxby et al., 2011; Jiahui et al., 2020) that calculated a CFS based on data from a single paradigm, we calculated a *multi-paradigm CFS* based on data from three different paradigms. Following an exhaustive leave-one-subject-out cross-validation, each *training subject*'s response time series from the movie "Forrest Gump" (≈120 min, split into eight runs; TR=2 s), the movie's audio-description that was produced for a visually impaired audience (≈120 min, split into eight runs runs; TR=2 s), and the visual localizer (≈20 min, split into four runs; TR=2 s) were concatenated and fed into the SRM algorithm in order to calculate the CFS and the training subjects' transformations (s. Fig. 5.1). We then aligned the *test subject* with the CFS to obtain the test subject's transformation. In order to investigate the prediction performance of each paradigm, a test subject's response time series from each of the three paradigms was separately aligned with the paradigm's corresponding TRs within the CFS letting us obtain transformation matrices based on each paradigm. In other words, the time series of each paradigm served as a separate predictor to estimate three different empirical Z-maps (i.e. one *cross-subject-within-paradigm prediction* and two *cross-subject-cross-paradigm predictions* per paradigm). Further, considering that acquiring functional data from a two-hour long naturalistic stimulus to align an individual to a CFS may not be desirable or feasible, we also investigated the relationship between the quantity of data of each predictor used to obtain a test subject's transformation and subsequent performance of estimating each Z-map. Our results indicate that an auditory narrative can be employed to estimate the results of a visual localizer, although it requires a longer functional scanning session. Additionally, we find that ≈15 min of movie data sampled at 0.5Hz used for volume-based functional alignment can estimate the results of the visual

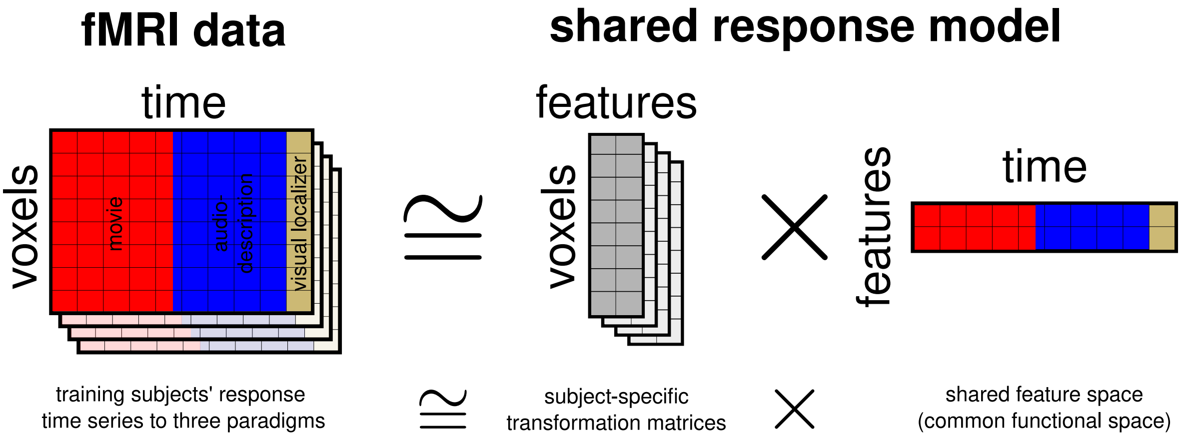


Fig. 5.1: **Overview of the shared response model (SRM).** For each fold of the leave-one-subject-out cross-validation, each training subject's response time series from the movie (≈120 min; time of repetition (TR)=2 s), the movie's audio-description (≈120 min; TR=2 s), and the visual localizer (≈20 min; TR=2 s) were concatenated to serve as the input for the SRM algorithm. From these response time series represented as matrix X_n (v voxels by t time points), the algorithm calculates the common functional space (CFS) C (k shared features by t time points) and subject-specific transformation matrices W_n (v voxels by k shared features) with orthonormal columns ($W_n^T W_n = I_k$).

localizer more accurately than an estimation procedure based on non-linear volume-based anatomical alignment. This opens up the possibility of estimating results from many localizer paradigms using a naturalistic stimulus of similar duration to one localizer to gain insights into individual functional brain anatomy.

5.2 Methods

For the current study, we used the same subset of the studyforrest dataset as in Häusler et al. (2022). The same fourteen participants participated in a six-category block-design visual localizer (Sengupta et al., 2016), watched the audio-visual movie "Forrest Gump" (Hanke, Adelhöfer, et al., 2016), and listened to the movie's audio-description (Hanke et al., 2014a). An exhaustive description of the participants, stimulus creation, procedure, stimulation setup, and fMRI acquisition can be found in the corresponding publications, while a summary is provided in Häusler et al. (2022).

5.2.1 Preprocessing

The analyses in this study were conducted on the same preprocessed fMRI data (s. `github.com/psychoinformatics-de/studyforrest-data-aligned`) that were used for (a) the technical validation of the dataset (Hanke, Adelhöfer, et al., 2016), (b) the localization of high-visual areas (Sengupta et al., 2016), and (c) the investigation of responses of the PPA correlating with naturalistic spatial information (Häusler et al., 2022). We reran the preprocessing and analyses steps performed in Sengupta et al. (2016) and Häusler et al. (2022) using FEAT v6.00 (FMRI Expert Analysis Tool; Woolrich, Ripley, Brady, & Smith, 2001) as shipped with FSL v5.0.9 (FMRIB's Software Library; S. M. Smith et al., 2004) to reproduce the time series that served as input for the previous statistical analyses and their results (i.e. the empirical Z-maps). The preprocessing steps included high-pass temporal filtering (using a Gaussian-weighted least-squares straight line) for every run

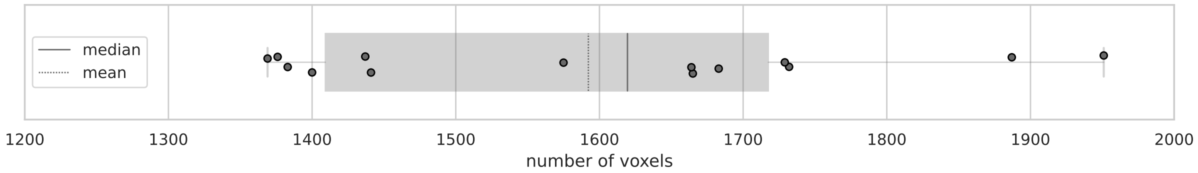


Fig. 5.2: **Size of the bilateral region of interest (ROI) of each participant.** In order to reduce the number of voxels, we warped the union of individual PPAs (cf. Fig. 1 in Häusler et al., 2022) from MNI152 space into each participant's native voxel space. The remaining voxels of each participant were further constrained to those voxels that are included in the respective participant's FoV of the audio-description (cf. Hanke et al., 2014a).

of the visual localizer (cutoff period of 100 s) and naturalistic stimuli (cutoff period of 150 s). Brain extraction was performed using BET (S. M. Smith, 2002), and data from all three paradigms were spatially smoothed using a Gaussian kernel with a full width at half maximum of 4.0 mm. A grand-mean intensity normalization was applied to each run of the functional localizer (four runs, each lasting ≈5 min; TR=2 s) and naturalistic stimuli (eight runs, each lasting ≈15 min; TR=2 s). Further analyses on these reproduced times series were performed using Python (v3.7) scripts that relied on NiBabel v4.0.2 (`nipy.org`), NumPy v1.21.6 (`numpy.org`), Pandas v1.3.5 (`pandas.pydata.org`), Scipy v1.7.3 (`scipy.org`), scikit-learn v1.0.2 (`scikit-learn.org`), BrainIAK v0.11 (`brainiak.org` Kumar et al., 2021, 2020), Matplotlib v3.5.3 (`matplotlib.org`), seaborn v0.11.2 (`seaborn.pydata.org`), and calling command line functions of FSL.

The SRM requires that the number of samples (i.e. TRs) exceed the number of features (i.e. voxels). In order to restrict the number of voxels, we created a bilateral ROI for each participant. Specifically, we warped the union of individual PPAs (s. Fig. 1 in Häusler et al., 2022) from MNI space into each participant's voxel space using subject-specific, non-linear transformation matrices that were previously computed (Hanke et al., 2014a, `github.com/psychoinformatics-de/studyforrest-data-templatetransforms`). The time series of each participant were then masked in their native voxel space by the union of individual PPAs and the subject-specific field of view (FoV) of the audio-description. The number of remaining voxels per participant (range 1369–1951, $\overline{X} = 1592$, $SD = 188$) can be seen in Fig. 5.2. Data of each run were normalized (z-scored) to a mean of zero ($\overline{X} = 0$) and a standard deviation of one ($SD = 1$). Due to an image reconstruction problem (cf. Hanke et al., 2014a), the last 75 TRs of the audio-description were missing in subject 04. The SRM allows for different numbers of voxels across subjects, but the number of TRs must be the same. Consequently, we removed the last 75 TRs of the audio-description from the time series of all other participants. As a result, the data used to fit the SRM in the next step included 3599 TRs of the movie, 3524 TRs of the audio-description, and 624 TRs of the visual localizer (7747 TRs in total). The time series of all three paradigms were concatenated and z-scored.

5.2.2 Estimation via functional alignment

To estimate the empirical Z-maps (i.e. the results of the three t-contrast), we followed a three-step procedure. First, for every fold of the leave-one-subject-out cross-validation (N=14 participants), we fit a SRM to $N-1$ training subjects' response response time series from the movie, the audio-description, and the visual localizer. This step generated a CFS for each fold of the cross-validation and a transformation matrix with orthonormal columns for each training subject. Second, we aligned the test subject's response time series from the movie, audio-description, and visual localizer separately to the corresponding TRs within the CFS. This procedure yielded different transformation matrices for the test subject based on data from different paradigms. In order to examine the relationship between the estimation performance and the amount of data used to generate a transformation matrix, we also varied the number of runs of the paradigms. This additional procedure during step two produced transformation matrices based on an increasing number of runs per paradigm. In the third step, we estimated a test subject's empirical Z-map by first projecting the training subjects' empirical Z-maps from their voxel space into the CFS using their transformation matrices. Then, we projected the training subjects' Z-maps from the CFS into the test subject's voxel space using the transpose of the test subject's transformation matrix (that is equivalent to the inverse of the transformation matrix due to its orthonormal columns). Finally, we obtained the test subject's predicted Z-maps by calculating the arithmetic mean of the respective paradigm's projected empirical Z-maps.

Fitting the SRM

In order to obtain the CFS and the training subjects' transformation matrices, we used the probabilistic SRM algorithm that is implemented in BrainIAK v.11 (Brain Imaging Analysis Kit; Kumar et al., 2021, 2020), and approximates the SRM based on the Expectation Maximization (EM) algorithm as proposed by Chen et al. (2015) and optimized by Anderson et al. (2016). We chose a value of $k = 10$ for the number of shared features (i.e. the number of dimensions in the CFS) based on the temporal and spatial resolution of our data (TR $= 2\,\mathrm{s}$; $2.5 \times 2.5 \times 2.5\,\mathrm{mm}$), the average number of voxels per ROI, and findings from Haxby et al. (2011). Haxby et al. (2011) used hyperalignment to create a CFS of 1,000 dimensions based on functional data (TR $= 3\,\mathrm{s}$) of voxels ($3 \times 3 \times 3\,\mathrm{mm}$) located in the ventral temporal cortex. They then reduced the dimensionality of the CFS by applying a principal component analysis (PCA) in order to determine the subspace that is sufficient to capture the full range of response-pattern distinctions. Results revealed that approximately 35 principal components (i.e. dimensions) were sufficient to represent the information content of a movie from which the CFS was derived. Results also showed that the cortical topographies of category-selective brain regions was preserved in the 35-dimensional CFS. In the present study, we also computed CFSs of $k = 5, 20, 30, 40, 50$ but the prediction performance based on these CFSs barely varied from a 10-dimensional CFS. The algorithm was set to iterate 30 times to minimize the error.

In order to visualize characteristics of the CFS, we calculated the Pearson correlation coefficients between the shared responses and the regressors that we previously created to model hemodynamic responses during the three paradigms (cf. Häusler et al., 2022; Sengupta et al., 2016). As an example, we chose the CFS that was created in the first fold of the cross-validation from $N - 1$ subjects to estimate the Z-maps of subject 01. The time series of the shared features were trimmed to match the corresponding TRs of the respective paradigms. Fig. 5.3 shows the correlations between regressors created to model hemodynamic responses during the visual localizer and shared responses (trimmed to the TRs that match the visual localizer). Fig. 5.4 shows the correlations between regressors created to model hemodynamic responses during the movie (cf. Table 3 in Häusler et al., 2022) and shared responses, while Fig. 5.5 shows the correlations between regressors created to model hemodynamic responses during the audio-description (cf. Table 3 in Häusler et al., 2022) and shared responses.

As a negative control, we created 1000 models based on randomly shuffled time series. We expected that the SRM algorithm would yield “shared” responses that are not correlated with the regressors. For each of the 1000 models, the order of runs of the visual localizer and naturalistic stimuli were shuffled separately for each training subject. Then, we concatenated the time series, fit the SRM, calculated the Person correlation coefficients per model, and finally their means across the 1000 models. As hypothesized, the shared features within CFSs based on shuffled time series show no or just minor mean correlations with the regressors, as shown in Fig. A.1, Fig. A.2, and Fig. A.3.

Alignment of a test subject

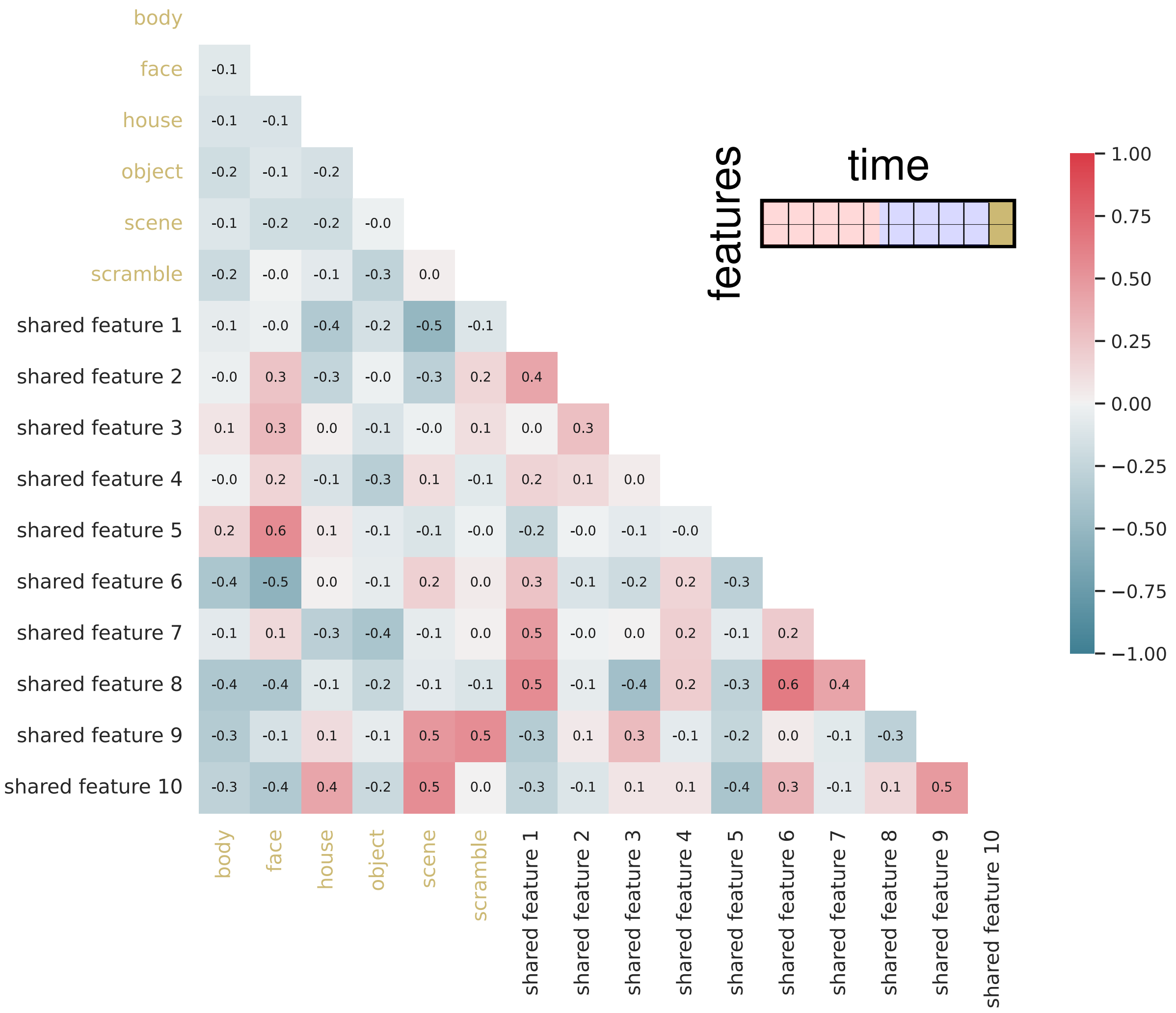
We aligned the test subject’s response time series from the visual localizer, the movie, or the audio-description to the corresponding TRs within the CFS by factorizing the response time series data via singular value decomposition. This step produced transformation matrices with orthonormal columns that allow a linear transformation of data from a test subject’s voxel space into the CFS. To investigate how the amount of data used to acquire a transformation matrix affects the estimation performance, we also varied the number of runs per paradigm. Specifically, we used one up to four runs (each lasting ≈5 min) of the visual localizer, and one up to eight runs (each lasting ≈15 min) of the naturalistic stimuli to align the test subject’s time series with the corresponding TRs within the CFS. Therefore, for each test subject, we obtained four matrices based on data from the visual localizer and eight different matrices per naturalistic stimulus, each transformation matrix having a size of v voxels by k shared responses but being based on an increasing amount of data used to calculate the linear transformation.

Estimation of a test subject’s Z-maps

We estimated the empirical Z-maps of the test subject by projecting the empirical Z-maps of all training subjects trough the CFS into the test subject’s voxel space. First, we masked the empirical Z-maps of the training subjects with the same subject-specific masks that we used to mask the time series data. Then, we used the transformation matrices derived during the training of the CFS to map the masked empirical Z-maps from each training subject’s voxel space into the CFS. Next, we used the transpose of a transformation matrix obtained from the alignment of the test subject to project the Z-maps from the CFS into the test subject’s voxel space. For each of the three t-contrasts, we obtained the test subject’s predicted Z-map by calculating the arithmetic mean of the respective paradigm’s projected empirical Z-maps.

5.2.3 Estimation via anatomical alignment

As a baseline, we used a non-linear anatomical alignment procedure to estimate a test subject’s empirical Z-maps. First, we projected the masked empirical Z-maps of each paradigm and each training subject from their native voxel space into the MNI space via previously computed subject-specific transformation matrices (Hanke et al., 2014a, `github.com/psychoinformatics-de/studyforrest-data-templatetransforms`). We then used the test subject’s pseudoinverse transformation matrix to project the data from MNI space into the test subject’s voxel space. Similar to our functional alignment procedure, we obtained an estimation of the test subject’s empirical Z-maps by calculating the arithmetic mean of the respective paradigm’s projected Z-maps.



	body	face	house	object	scene	scramble	shared feature 1	shared feature 2	shared feature 3	shared feature 4	shared feature 5	shared feature 6	shared feature 7	shared feature 8	shared feature 9	shared feature 10
body																
face	-0.1															
house	-0.1	-0.1														
object	-0.2	-0.1	-0.2													
scene	-0.1	-0.2	-0.2	-0.0												
scramble	-0.2	-0.0	-0.1	-0.3	0.0											
shared feature 1	-0.1	-0.0	-0.4	-0.2	-0.5	-0.1										
shared feature 2	-0.0	0.3	-0.3	-0.0	-0.3	0.2	0.4									
shared feature 3	0.1	0.3	0.0	-0.1	-0.0	0.1	0.0	0.3								
shared feature 4	-0.0	0.2	-0.1	-0.3	0.1	-0.1	0.2	0.1	0.0							
shared feature 5	0.2	0.6	0.1	-0.1	-0.1	-0.0	-0.2	-0.0	-0.1	-0.0						
shared feature 6	-0.4	-0.5	0.0	-0.1	0.2	0.0	0.3	-0.1	-0.2	0.2	-0.3					
shared feature 7	-0.1	0.1	-0.3	-0.4	-0.1	0.0	0.5	-0.0	0.0	0.2	-0.1	0.2				
shared feature 8	-0.4	-0.4	-0.1	-0.2	-0.1	-0.1	0.5	-0.1	-0.4	0.2	-0.3	0.6	0.4			
shared feature 9	-0.3	-0.1	0.1	-0.1	0.5	0.5	-0.3	0.1	0.3	-0.1	-0.2	0.0	-0.1	-0.3		
shared feature 10	-0.3	-0.4	0.4	-0.2	0.5	0.0	-0.3	-0.1	0.1	0.1	-0.4	0.3	-0.1	0.1	0.5	

Fig. 5.3: **Similarity of hemodynamic responses modeled for the analysis of the visual localizer in Sengupta et al. (2016) and shared features calculated by the shared response model (SRM) in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients plotted in the figure, the time series of the shared features within the multi-paradigm CFS were trimmed to match the corresponding TRs of the visual localizer paradigm (Sengupta et al., 2016). The modeled hemodynamic responses (i.e. regressors) represent predicted responses to the six categories of pictures that were presented in blocks.

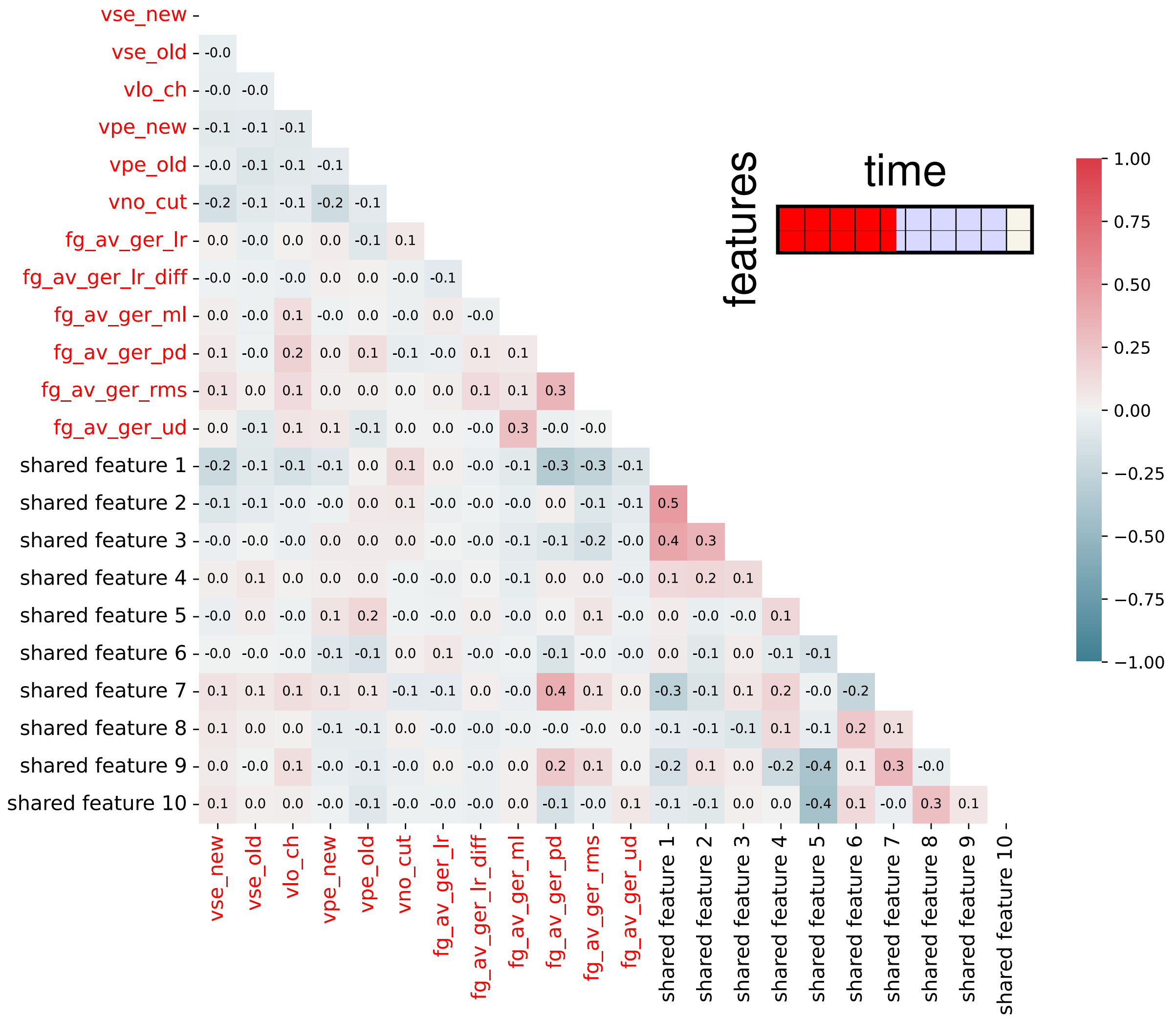
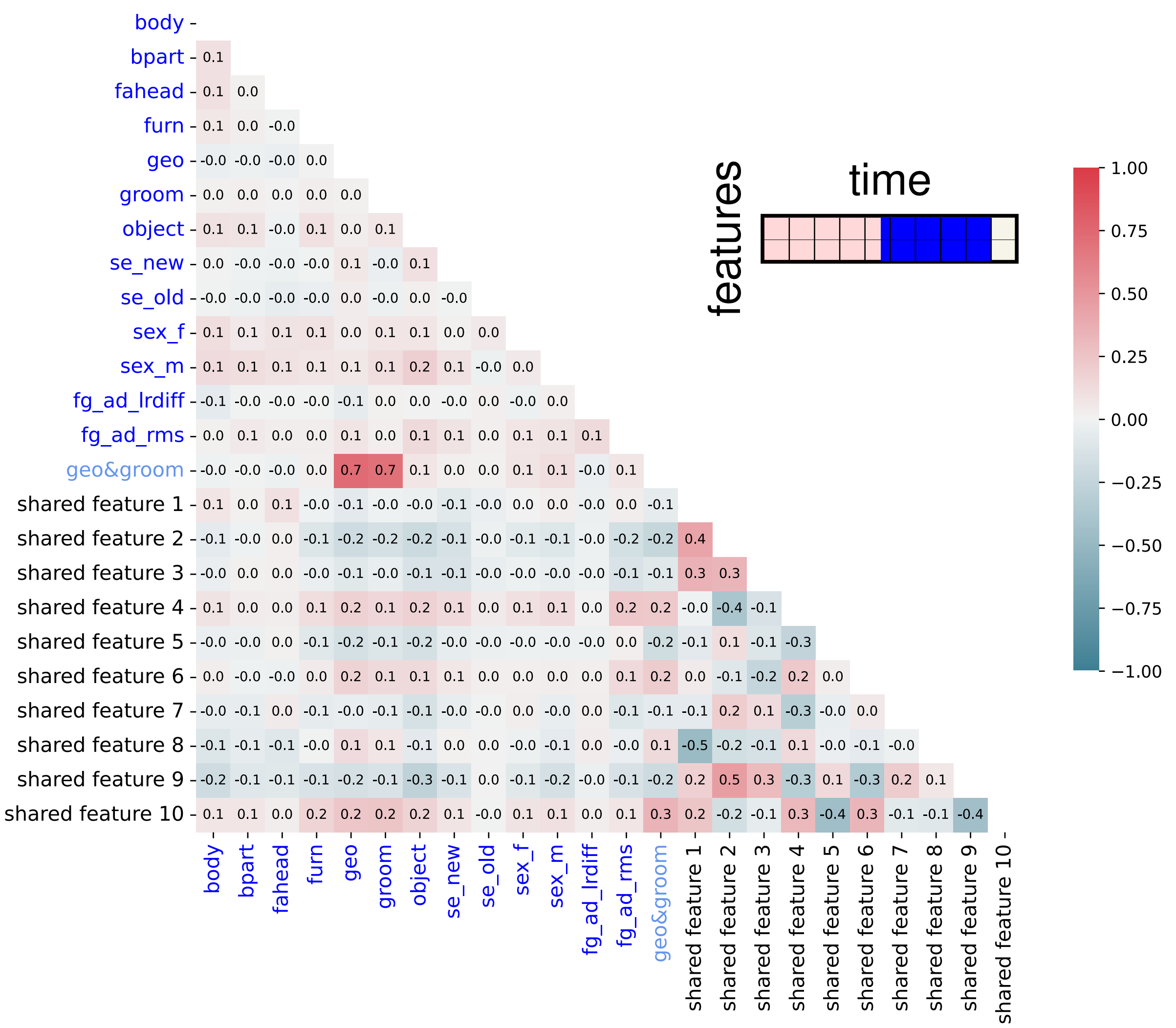


Fig. 5.4: **Similarity of hemodynamic responses modeled for the analysis of the movie in Häusler et al. (2022) and shared features calculated by the shared response model (SRM) in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients plotted in the figure, the time series of the shared features within the multi-paradigm CFS were trimmed to match the corresponding TRs of the movie (Hanke, Adelhöfer, et al., 2016). The modeled shared responses (i.e. regressors) `vse_new` to `vno_cut` are based on annotations of movie frames, whereas the regressors `fg_av_ger_lr` to `fg_av_ger_ud` represent low-level visual or auditory confounds (cf. Table 3 in Häusler et al., 2022). `vse_new`: change of the camera position to a setting not depicted before; `vse_old`: change of the camera position to a recurring setting; `vlo_ch`: change of the camera position to another locale within the same setting; `vpe_new`: change of the camera position within a locale not depicted before; `vpe_old`: change of the camera position within a recurring locale; `vno_cut`: a pseudo-randomly selected frames within a continuous movie shot; `fg_av_ger_lr`: left-right luminance difference; `fg_av_ger_lrdiff`: left-right volume difference; `fg_av_ger_ml`: mean luminance; `fg_av_ger_pd`: perceptual difference; `fg_av_ger_rms`: root mean square volume; `fg_av_ger_ud`: upper-lower luminance difference.



	body	bpart	fahead	furn	geo	groom	object	se_new	se_old	sex_f	sex_m	fg_ad_lrdiff	fg_ad_rms	geo&groom	shared feature 1	shared feature 2	shared feature 3	shared feature 4	shared feature 5	shared feature 6	shared feature 7	shared feature 8	shared feature 9	shared feature 10
body																								
bpart	0.1																							
fahead	0.1	0.0																						
furn	0.1	0.0	-0.0																					
geo	-0.0	-0.0	-0.0	0.0																				
groom	0.0	0.0	0.0	0.0	0.0																			
object	0.1	0.1	-0.0	0.1	0.0	0.1																		
se_new	0.0	-0.0	-0.0	-0.0	0.1	-0.0	0.1																	
se_old	-0.0	-0.0	-0.0	-0.0	0.0	-0.0	0.0	-0.0																
sex_f	0.1	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0															
sex_m	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	-0.0	0.0														
fg_ad_lrdiff	-0.1	-0.0	-0.0	-0.0	-0.1	0.0	0.0	-0.0	0.0	-0.0	0.0													
fg_ad_rms	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.1												
geo&groom	-0.0	-0.0	-0.0	0.0	0.7	0.7	0.1	0.0	0.0	0.1	0.1	-0.0	0.1											
shared feature 1	0.1	0.0	0.1	-0.0	-0.1	-0.0	-0.0	-0.1	-0.0	0.0	0.0	-0.0	0.0	-0.1										
shared feature 2	-0.1	-0.0	0.0	-0.1	-0.2	-0.2	-0.2	-0.1	-0.0	-0.1	-0.1	-0.0	-0.2	-0.2	0.4									
shared feature 3	-0.0	0.0	0.0	-0.0	-0.1	-0.0	-0.1	-0.1	-0.0	-0.0	-0.0	-0.0	-0.1	-0.1	0.3	0.3								
shared feature 4	0.1	0.0	0.0	0.1	0.2	0.1	0.2	0.1	0.0	0.1	0.1	0.0	0.2	0.2	-0.0	-0.4	-0.1							
shared feature 5	-0.0	-0.0	0.0	-0.1	-0.2	-0.1	-0.2	-0.0	-0.0	-0.0	-0.0	-0.0	0.0	-0.2	-0.1	0.1	-0.1	-0.3						
shared feature 6	0.0	-0.0	-0.0	0.0	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.0	-0.1	-0.2	0.2	0.0					
shared feature 7	-0.0	-0.1	0.0	-0.1	-0.0	-0.1	-0.1	-0.0	-0.0	0.0	-0.0	0.0	-0.1	-0.1	-0.1	0.2	0.1	-0.3	-0.0	0.0				
shared feature 8	-0.1	-0.1	-0.1	-0.0	0.1	0.1	-0.1	0.0	0.0	-0.0	-0.1	0.0	-0.0	0.1	-0.5	-0.2	-0.1	0.1	-0.0	-0.1	-0.0			
shared feature 9	-0.2	-0.1	-0.1	-0.1	-0.2	-0.1	-0.3	-0.1	0.0	-0.1	-0.2	-0.0	-0.1	-0.2	0.2	0.5	0.3	-0.3	0.1	-0.3	0.2	0.1		
shared feature 10	0.1	0.1	0.0	0.2	0.2	0.2	0.2	0.1	-0.0	0.1	0.1	0.0	0.1	0.3	0.2	-0.2	-0.1	0.3	-0.4	0.3	-0.1	-0.1	-0.4	

Fig. 5.5: **Similarity of hemodynamic responses modeled for the analysis of the audio-description in Häusler et al. (2022) and shared features calculated by the shared response model (SRM) in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients plotted in the figure, the time series of the shared features within the multi-paradigm CFS were trimmed to match the corresponding TRs of the audio-description (Hanke et al., 2014a). The modeled shared responses (i.e. regressors) `body` to `sex_m` are based on annotated categories of nouns spoken by the audio-description's narrator, whereas the regressors `fg_ad_ger_lrdiff` and `fg_ad_ger_rms` represent low-level auditory confounds (cf. Table 3 in Häusler et al., 2022). `body`: trunk of the body; overlaid clothes; `bpart`: limbs and trousers; `fahead`: (parts) of the face or head; `furn`: moveable furniture (insides & outsides); `geo`: immobile landmarks; `groom`: rooms & locales or geometry-defining elements; `object`: moveable and countable entities with firm boundaries; `se_new`: a setting occurring for the first time; `se_old`: a recurring setting; `sex_f`: female name, female person(s); `sex_m`: male name, male person(s); `fg_ad_lrdiff`: left-right volume difference; `fg_ad_rms`: root mean square volume. `geo&groom` is a combination of regressors as used on the positive side of the primary contrasts aimed to localize the PPA (cf. Table 5 in Häusler et al., 2022).

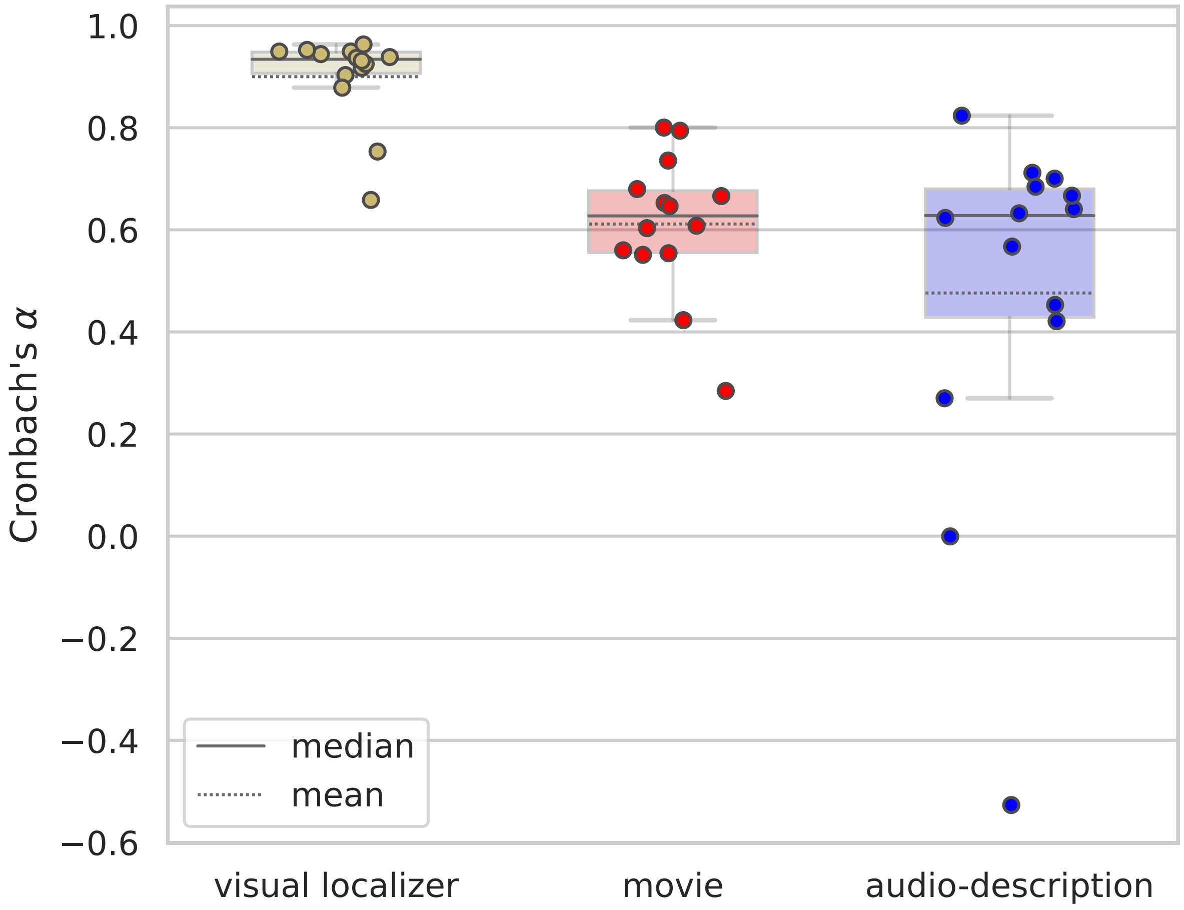


Fig. 5.6: **Reliability of the empirical Z-maps for each paradigm and subject.** Cronbach's α was calculated based on the Z-maps yielded by the first-level GLM analyses of the visual localizer (four runs; Sengupta et al., 2016) and naturalistic stimuli (eight runs; Häusler et al., 2022). The second-level GLM analyses across runs yielded the empirical Z-maps that were estimated in the present study.

5.2.4 Reliability of the empirical Z-maps

We calculated Cronbach's α as a measure of reliability and the amount of measurement error (Cortina, 1993; Cronbach, 1951) present in the empirical Z-maps of each paradigm and subject. Cronbach's α expresses the expected correlation between the currently used empirical Z-maps and an additional set of empirical Z-maps calculated based on data of a hypothetical independent dataset collected from the same paradigm and subjects (Jiahui et al., 2020; Jiahui, Feilong, Nastase, Haxby, & Gobbini, 2022). These expected correlations, represented by Cronbach's α, were calculated based on the first-level GLM Z-maps (four in case of the visual localizer; eight in case of the naturalistic stimuli) that were averaged in the second-level GLM analyses of Sengupta et al. (2016) and Häusler et al. (2022) respectively. Cronbach's α of the empirical (i.e. second-level) Z-maps for each subject and paradigm can be seen in Fig. 5.6, descriptive statistics across subjects for each paradigm can be seen in Table 5.6.

Table. 5.1: **Descriptive statistics of the reliability measure.** Cronbach's α was calculated based on the Z-maps yielded by the first-level GLM analyses of the visual localizer (four runs; Sengupta et al., 2016) and naturalistic stimuli (eight runs; Häusler et al., 2022). The second-level GLM analyses across runs yielded the empirical Z-maps that were estimated in the present study.

statistic	localizer	movie	audio-description
mean	0.90	0.61	0.48
std	0.09	0.14	0.36
min	0.66	0.28	-0.53
25%	0.91	0.56	0.43
50%	0.93	0.63	0.63
75%	0.95	0.68	0.68
max	0.96	0.80	0.82

5.3 Results

In order to assess the performance of the alignment procedures, we calculated the Pearson correlation coefficients between each individual's empirical Z-maps obtained from the previous analyses (Häusler et al., 2022; Sengupta et al., 2016) and their respective predicted Z-maps (s. Fig. 5.7). In general, the mean Pearson correlation coefficients vary depending on the paradigm being estimated (i.e. Z-maps of the visual localizer, movie, or audio-description), as well as the alignment procedure (anatomical vs. functional alignment). In the case of functional alignment, the quantity of the paradigm's data used as a predictor and to align a test subject to the CFS also affects the correlation coefficients. However, the functional alignment procedure consistently shows an increasing estimation performance as more data of a predictor is used to align the test subjects. In order to investigate potential differences between some conditions, we conducted 15 pairwise comparisons using Fisher z-transformed correlation values. These comparisons were not pre-planned, but rather were selected later as examples for further exploration. We used a Bonferroni correction for multiple comparisons to adjust the alpha level to an α of $0.05/15 = 0.00\overline{3}$.

When estimating the Z-maps of the visual localizer, the mean correlation between empirical Z-maps and Z-maps predicted using ≈15 min of the visual localizer (within-paradigm prediction) was significantly higher than the mean correlation between empirical Z-maps and Z-maps predicted via an anatomical alignment ($t(14) = 13.99$, $p < .0001$). Similarly, the mean correlation between empirical Z-maps and Z-maps predicted using ≈15 min of the movie (cross-paradigm prediction) was significantly higher than the correlations between empirical Z-maps and Z-maps predicted via anatomical alignment ($t(14) = 6.35$, $p < .0001$). A comparison between the prediction based on ≈15 min of the movie and ≈15 min of localizer data revealed a significantly lower performance of functional alignment using the movie ($t(14) = -11.64$, $p < .0001$). The prediction performance based on ≈30 min of the movie was significantly higher than the prediction performance based on ≈15 min of the movie ($t(14) = 5.49$, $p = .0001$), with no significant difference between ≈45 min and ≈30 min of the movie ($t(14) = 0.13$, $p = .8990$). Visual inspection indicated that the prediction performance based on ≈15 min of the audio-description was lower than the prediction performance based on anatomical alignment, functional alignment using ≈15 min of the localizer, or ≈15 min of the movie. However, the prediction performance of the audio-description increased monotonically the more

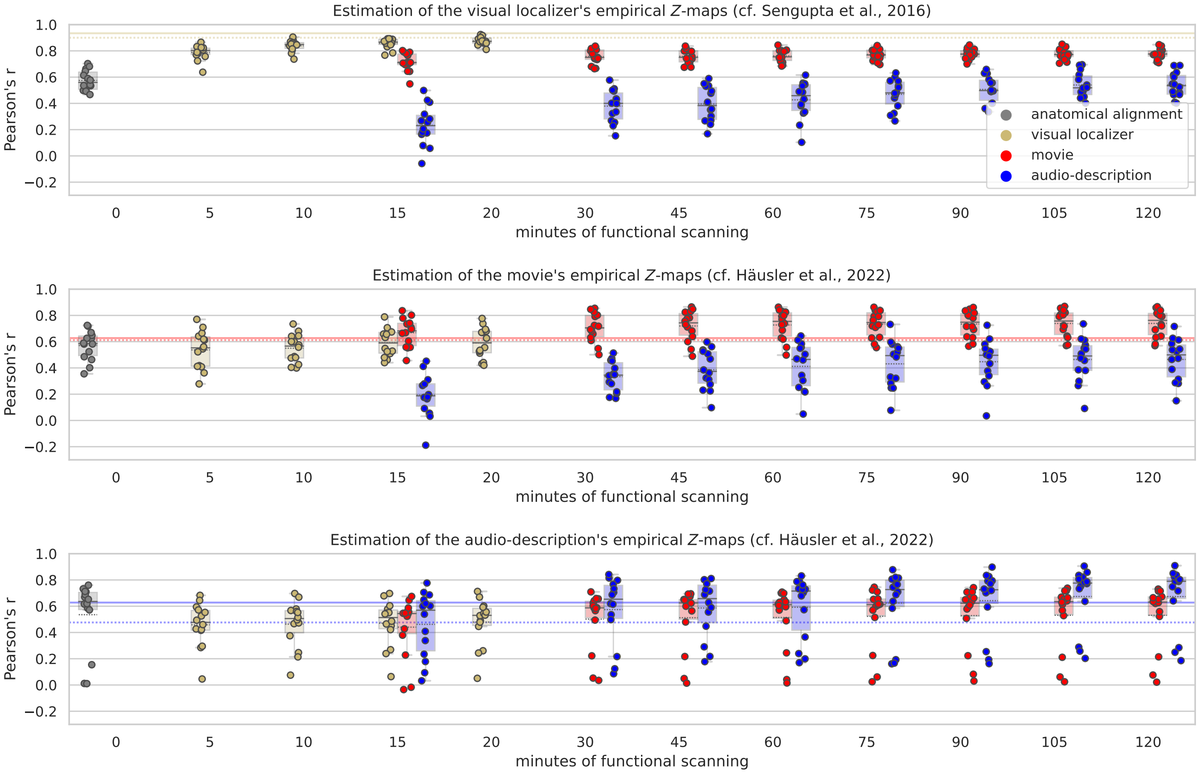


Fig. 5.7: **Similarity between empirical and predicted Z-maps for each subject and paradigm.** Functional alignment was performed based on an increasing amount of functional data used to align a test subject to the common functional space (CFS): each of the four runs of the visual localizer paradigm lasted ≈5 min (TR=2 s); each of the eight runs of the naturalistic stimuli lasted ≈15 min (TR=2 s). Solid horizontal lines: median of Cronbach's α across subjects the for empirical Z-maps of the respectively estimated paradigm (cf. Fig. 5.6). Dotted horizontal lines: mean of Cronbach's α across subjects for the empirical Z-maps of the respectively estimated paradigm (cf. Fig. 5.6). Grey dots: Pearson correlation coefficients between empirical Z-maps and an estimation using anatomical alignment. Green dots: correlations between empirical Z-map and an estimation using functional alignment based on transformations calculated from one up to four runs of the visual localizer. Red dots: correlations between empirical Z-map and an estimation using functional alignment based on transformations calculated from one up to eight runs of the movie. Blue dots: correlations between empirical Z-map and an estimation using functional alignment based on transformations calculated from one up to eight runs of the audio-description.

data were used to align the test subjects. A t-test comparing the prediction performance based on ≈120 min of the audio-description to anatomical alignment yielded no significant difference ($t(14) = -1.17$, $p = .2640$).

When estimating the Z-maps of the movie, the mean correlation between empirical Z-maps and predicted Z-maps using ≈15 min of the movie (within-paradigm prediction) were significantly higher than the correlations between empirical Z-maps and predicted Z-maps via an anatomical alignment ($t(14) = 5.78$, $p < .0001$). Comparing the within-paradigm prediction using ≈15 min of the movie to the cross-paradigm prediction using ≈15 min of the localizer revealed a higher prediction performance of the movie ($t(14) = 5.53$, $p < .0001$). There was no significant difference between the prediction using ≈15 min of the localizer and the prediction via anatomical alignment ($t(14) = 1.15$, $p = .2726$). The prediction performance based on ≈30 min of movie data was significantly higher than the prediction performance based on ≈15 min of the movie ($t(14) = 3.75$, $p = .0024$), whereas there was no significant difference between the performance based on ≈45 min and ≈30 min of the movie ($t(14) = 2.58$, $p = .0230$). As is evident by visual inspection, the prediction performance based on ≈15 min of the audio-description was lower than a prediction based on anatomical alignment, functional alignment using ≈15 min of the localizer, or functional alignment using ≈15 min of the movie. Here again, the prediction performance of the audio-description monotonically increased the more data were used to align the test subjects. A t-test comparing the prediction performance based on ≈120 min of the audio-description to a prediction via anatomical alignment yielded no significant difference ($t(14) = -2.40$, $p = .0318$).

When estimating the Z-maps of the audio-description, the mean correlation between empirical Z-maps and predicted Z-maps via ≈15 min of the audio-description (within-paradigm prediction) was not significantly different to the mean correlation between empirical Z-maps and predicted Z-maps via anatomical alignment ($t(14) = -1.82$, $p = .0925$). Comparing the prediction based on ≈120 min of the audio-description to the anatomical alignment procedure yielded a significantly higher performance of the estimation via functional alignment ($t(14) = 6.56$, $p <= .0001$). Comparing the prediction based on functional alignment via ≈120 min of the movie to the anatomical alignment procedure yielded no significant difference ($t(14) = -0.76$, $p = .4625$).

5.4 Discussion

Block-design functional localizers are traditionally used to identify functional areas in individuals. However, these paradigms are typically limited to mapping a single domain of brain functions. Additionally, the diagnostic quality of functional localizers relies on an individual's comprehension of task instructions and compliance. In our study, we focused on the PPA as an example of a high-visual area. We estimated the results of t-contrasts (i.e. the empirical Z-maps), created in previous studies (Häusler et al., 2022; Sengupta et al., 2016) in order to localize the PPA, in an individual by leveraging data collected from a reference group. To address the challenge of functional-anatomical variability across individuals, we employed a functional alignment approach based on the shared response model (SRM) (Chen et al., 2015). Following an exhaustive leave-one-subject-out cross-validation, we computed a multi-paradigm common functional space (CFS) based on the training subjects' concatenated response time series from a visual localizer, a movie, and the movie's audio-description. Each test subject's response time series from one of the paradigms was separately used to functionally aligning the test subject with the corresponding paradigm's TRs within the CFS. Finally, we projected the empirical Z-maps of the reference group through the CFS into the test subject's voxel space to generate the predicted Z-maps for each test subject. Considering the challenges of acquiring functional data from a two-hour paradigm to functionally align an individual with a CFS, we also explored the relationship between the amount of data used for alignment and subsequent estimation performance. As a baseline comparison, we employed an anatomical alignment procedure in which the training subjects' Z-maps were projected via a non-linear transformation through the MNI space into the test subject's voxel space. Our findings demonstrate that an auditory narrative can be utilized to estimate the results of a visual localizer, although it requires a longer functional scanning session. Furthermore, we observed that employing ≈15 min of movie data sampled at 0.5Hz for volume-based functional alignment leads to more accurate estimations of the visual localizer results compared to a non-linear volume-based anatomical alignment approach. By leveraging data from a reference group, our procedure opens up the possibility of estimating results from many localizer paradigms using a naturalistic stimulus of similar duration to one traditional localizer.

5.4.1 Estimating the results of the visual localizer

We estimated the results of the visual localizer, which is the established method for identifying the PPA. Our results indicate that ≈15 min of localizer or movie data used for functional alignment with a CFS are sufficient to estimate the results of the visual localizer with higher fidelity compared to an estimation procedure based on anatomical alignment. When comparing the within-paradigm prediction based on the localizer data to the cross-paradigm prediction based on the movie data, the within-paradigm prediction showed a superior estimation performance. These results are in line with Haxby et al. (2011), who created two CFSs: one CFS based on a controlled paradigm that employed images of stimulus categories and another CFS based on a movie. They found that a cross-subject classification of the controlled paradigm's categories via the CFS and transformations calculated from the same paradigm's data outperformed the classification based on the movie data. Haxby et al. (2011) interpreted these results gained from two paradigms providing roughly the same amount of TRs to calculate the CFS and align a test subject as the controlled paradigm sampling the investigated brain states more extensively than the movie. In our study, the prediction performance based on movie data increased significantly when ≈30 min of the movie were used for alignment with the CFS. However, comparing ≈30 min to ≈45 min for aligning a test subject did not yield a significant difference. Overall, these results suggest that the estimation based on functional alignment using a movie approaches a performance limit, with diminishing benefits of longer scanning time (cf. Fig. 5.7).

We also explored a cross-paradigm prediction of the visual localizer's results using time series from the audio-description, lacking a visual stimulation. The prediction based on ≈15 min of audio-description showed the lowest performance among all alignment procedures, including anatomical alignment. A possible first explanation for the low performance might be that the audio-description offers a less diverse stimulation and has a sparser event structure that does not sample the targeted responses as extensively. Further investigations might explore alternative auditory paradigms that more extensively sample the targeted responses to achieve higher estimation performances. However, as more data were used, the prediction performance based on the audio-description improved while narrowing the gap to the prediction based on the same amount of movie data. When comparing the prediction based on ≈120 min of the audio-description to the prediction based on anatomical alignment, we found no statistical difference. These results suggest that an auditory paradigm could potentially substitute a visual paradigm for functional alignment and subsequent estimation of a visual paradigm's results. However, achieving a comparable estimation performance requires a lengthy functional scanning session. Therefore, a second explanation, not mutually exclusive to the first one, might be that neural responses in the PPA to an auditory paradigm differ from responses to a visual paradigm. This explanation would support findings of Häusler et al. (2022) and would be in line with Haxby et al. (2011). In Häusler et al. (2022), we observed hemodynamic activity during auditory stimulation that is restricted to the anterior part of the PPA. Nevertheless, further studies using controlled paradigms are needed to investigate responses of the PPA to auditory spatial information. Haxby et al. (2011) demonstrated that the general validity of their functional alignment procedure "based on the responses to the movie is not dependent on responses to stimuli that are in both the movie and the category perception experiments" (Haxby et al., 2011, p. 409). In light of our current findings, we hypothesize that a stimulus used for functional alignment does not necessarily need to sample hemodynamic responses identical to those evoked by the target paradigm, however with the disadvantage of a longer scanning time. Therefore, it would be interesting to explore whether functional alignment can be employed to estimate Z-maps from paradigms designed to elicit brain processes that are typically not sampled during passive movie watching, such as planning or decision-making.

5.4.2 Estimating the results of the movie & audio-description

We also estimated the results of Häusler et al. (2022), where we created t-contrast based on modeled hemodynamic activity during two naturalistic stimuli to localize the PPA under more ecologically conditions. Once again, the within-paradigm predictions achieved higher correlations between empirical and predicted Z-maps compared to the cross-paradigm predictions. Similar to the estimation of the visual localizer results, the estimation of the movie results based on the audio-description showed the lowest performance among all alignment procedures, including anatomical alignment. However, as more data from the audio-description were incorporated for alignment, the prediction performance improved. Moreover, similar to the estimation of the visual localizer, the gap between the estimation performance using the audio-description and the same amount of movie data narrowed. These results add evidence that an auditory paradigm can be used to estimate the results of a visual paradigm. However, it requires a long scan session.

Contrary to expectation, the within-paradigm prediction based on $\approx$15 min of the audio-description did not outperform the prediction via anatomical alignment. Given the exploratory nature of our study, it is challenging to untangle the individual contributions of the exclusively auditory stimulus, the multi-paradigm CFS, the transformation matrices of test subjects, and the mean reliability of the audio-description's Z-maps on the estimation performance. Notably, calculating Cronbach's α of the audio-description's Z-maps revealed mediocre to poor reliability in three participants. The three participants with low Cronbach's α also represent the three outliers exhibiting low correlations between the audio-description's empirical and predicted Z-maps independent of the amount of data used for functional alignment. Results suggest that these participants are not outliers showing a response in the PPA that is reliably deviant from the norm but rather exhibited more variable responses to auditory spatial information during the two-hour stimulation. Nevertheless, when estimating the Z-maps of the visual paradigms based on data of the audio-description, we did not identify the same participants as outliers. Therefore, the reliability of the audio-description's Z-maps might be low in these three participants, however results show that their responses to the audio-description could be used to acquire transformations that enabled the estimation of the visual paradigms' results. Moreover, the observation that the mean correlation between the audio-description's empirical and predicted Z-maps exceeds mean Cronbach's α as more data from the audio-description are used can be interpreted as the SRM denoising a test subject's data (cf. Chen et al., 2015). Therefore, it would be interesting to estimate the results of a controlled auditory experiment, such as a speech localizer, that yields more reliable results in all participants. To test the hypothesis that the SRM filters out noise, one could calculate more reliable second-level Z-maps from actual time-series, and time series with artificially added noise at different intensities. If the SRM filters out noise, the prediction based on an alignment using a test subject's time series with added noise should resemble the actual Z-maps more closely than the Z-maps calculated from noisy data.

Lastly, it is important to acknowledge that our analysis was restricted to voxels located in the union of individual PPAs, and thus, to voxels that exhibited significantly increased activity in at least one participant in Sengupta et al. (2016). However, the findings of Häusler et al. (2022) suggest that hemodynamic activity during auditory semantic stimulation is confined to the anterior part of the PPA. Consequently, the ROI introduced a bias against the reliability of the Z-maps derived from the audio-description. This is because the ROI includes voxels that did not show significantly increased activity in response to auditory spatial information in at least one participant, but rather exhibit random variation (or potentially significantly decreased activation, which would not affect the reliability measure).

5.4.3 Summary, shortcomings & future studies

Our analyses revealed that using ≈15 min of the audio-description for aligning a test subject with the CFS resulted in relatively low performance in the cross- and within-paradigm predictions. Nevertheless, results suggest that it is in principle feasible to substitute a visual paradigm with an auditory paradigm. An interesting avenue for further investigation would be to estimate the topography of language-related areas using a narrative for functional alignment and compare the results with an estimation based on a movie. Furthermore, our findings indicate that ≈15 min of movie data used for functional alignment are sufficient to estimate the results of the visual localizer with higher fidelity than an estimation based on an anatomical alignment. The prediction performance based on the movie significantly improves when ≈30 min of the movie are used. However, we found no significant difference in performance between aligning a test subject using ≈45 min compared to ≈30 min. Overall, our results suggest that the functional alignment based on movie data is approaching a performance limit, and longer scanning sessions than 30 min during audio-visual stimulation may not yield substantial benefits.

When estimating the results of the visual localizer, the cross-paradigm prediction based on the movie showed lower performance than the within-prediction based on the localizer. However, movies offer a more complex stimulation compared to controlled paradigms. While we focused on the PPA as a high-visual area, naturalistic stimuli have been successfully employed to investigate various domains of brain function, including vision, audition, language, emotions, or social cognition (s. Jääskeläinen et al., 2021, for a review). Given the ability of naturalistic stimuli to elicit responses across diverse domains, they potentially provide transformations that better generalize across a wider range of paradigms compared to transformations obtained from dedicated experiments such as a visual localizer. Previous studies that used an anatomical alignment procedure to estimate the most probable location of functional areas have "found a large variability in the degree to which functional areas respect macro-anatomical boundaries" (Frost & Goebel, 2012,

p. 1369). For instance, retinotopically defined regions of the early visual cortex exhibit low interindividual variability (Rosenke et al., 2021), while the spatial location of other functional areas, such as language areas, varies greatly across individuals (Frost & Goebel, 2012). Even within the domain of category-selective areas, interindividual variability varies across functional areas, with scene-selective regions showing larger variability in spatial topography compared to face-selective regions (Frost & Goebel, 2012; Zhen et al., 2017, 2015). Future studies should explore the performance of functional alignment in other domains than high-visual perception. In particular, we showed that an auditory paradigm can be used to estimate the results of a visual paradigm. Therefore, it would be interesting to explore whether functional alignment can be employed to estimate Z-maps from paradigms that were designed to elicit brain processes that are not sampled during passive movie watching such as planning or decision-making.

Lastly, we emphasize that our study is exploratory, based on a sample of 14 participants, and therefore provides preliminary results that warrant further in-depth investigations into aspects that are beyond of scope here. For example, we restricted our analysis to voxels in a ROI with an average of $\approx$1600 voxel per participant. The use of searchlight functional alignment (e.g., Guntupalli et al., 2016; Zhang et al., 2016) could cover the entire cortex but restricts the simultaneously aligned voxels to those within the searchlight's sphere. Future studies are needed to develop functional alignment algorithms that can align voxels across larger distances. Such algorithms would be especially beneficial for functional areas that show large interindividual variability in anatomical location, such as language areas, or for cases of atypical topography, e.g., resulting from cortical reorganization after brain injuries.

Our results suggest that 15 to 30 minutes of fMRI scanning during a naturalistic stimulus that captures a wide range of brain states could provide sufficient data for a functional *calibration scan*. This calibration scan could be employed to align a new subject to a CFS derived from extensive scans of a reference group. The reference group's scans would encompass data collected from both naturalistic paradigms and controlled paradigms. The controlled paradigms would include functional localizers specifically designed to reliably map perceptual or cognitive processes. Compared to a diagnostic run based on a controlled paradigm, a naturalistic stimulus would offer the additional benefits of higher engagement and better compliance (Eickhoff et al., 2020; Vanderwal et al., 2015), especially in children or patients. Once a new subject is aligned with the CFS, functional data collected from the reference group could be mapped through the CFS into the new subject's voxel space. This enables the estimation of patterns in the new subject that are common in the normative reference group when obtaining additional functional scans is not feasible due to scanner availability, time constraints, financial limitations, or compliance issues. Furthermore, this approach allows for quantifying the similarity (or dissimilarity) between a new subject's actual pattern and the common pattern estimated from the normative reference. For instance, Yates, Ellis, and Turk-Browne (2021) calculated a SRM to investigate the presence and localization of adult brain functions in children. The authors mapped hemodynamic responses of adults watching a movie through a CFS into the anatomical brain space of children that watched the same movie. They found reliable correlations between the predicted and actual fMRI activity of children, and the strength of these correlations in the precuneus, inferior frontal gyrus, and lateral occipital cortex could predict the children's age.

5.5 Conclusion

Our findings suggest that 15 minutes of functional scanning using an engaging movie can generate a sufficient amount of data to functionally align an individual with a common functional space and estimate brain patterns with higher fidelity than a procedure based on anatomical alignment. By leveraging data from a reference group, it becomes possible to estimate the results from many localizer paradigms using a naturalistic stimulus of similar duration to that of one traditional localizer. This procedure offers a more efficient and cost-effective method to acquire valuable insights into the functional organization of an individual when additional functional scans are not feasible due to various constraints. While our study focused on the PPA as an example of a high-visual area, future research should explore the performance of functional alignment in other domains of brain functions, especially those not directly sampled during passive movie watching.

Data Availability

All fMRI data and results are available as Datalad (Halchenko et al., 2021) datasets, published to or linked from the *G-Node GIN* repository (`gin.g-node.org/chaeusler/studyforrest-ppa-srm`). Raw data of the audio-description, movie and visual localizer were originally published on the *OpenfMRI* portal (`https://legacy.openfmri.org/dataset/ds000113`, Hanke et al., 2014b), (`https://legacy.openfmri.org/dataset/ds000113d`, Hanke, Kottke, et al., 2016). Results from the localization of higher visual areas are available as Datalad datasets at *GitHub* (`github.com/psychoinformatics-de/studyforrest-data-visualrois`). The realigned participant-specific time series that were used in the current analyses were derived from the raw data releases and are available as Datalad datasets at *GitHub* (`github.com/psychoinformatics-de/studyforrest-data-aligned`). The same data are available in a modified and merged form on OpenNeuro at `https://openneuro.org/datasets/ds000113`.

Code Availability

Scripts to generate the results as Datalad (Halchenko et al., 2021) datasets are available in a *G-Node GIN* repository (`gin.g-node.org/chaeusler/studyforrest-ppa-srm`).

6 General discussion

Traditionally, human brain mapping studies have averaged fMRI data across participants. To advance the field towards a clinical application, data need to be interpreted on the level of individuals. Functional localizers are an established method to describe the topography (i.e. the location, size, and shape) of functional areas in individuals. However, traditional localizer paradigms rely on selectively sampled, tightly controlled stimuli and participant compliance, and can usually map only one domain of brain functions. Naturalistic stimuli like movies and auditory narratives offer a time-locked event structure that samples a variety of brain functions ranging from low-level perception to high-level cognition. A localizer based on a naturalistic stimulus could offer a higher external validity and potentially map a variety of brain functions. Consequently, the purpose of this thesis was—while adhering to the principles of open, transparent, and reproducible science—to investigate whether a movie and the movie's audio-description may, in principle, replace a traditional localizer paradigm. As a proof of concept, we focused on the PPA, as an example of a high-level visual area. The PPA exhibits increased hemodynamic activity when participants view photos of landscapes, buildings or landmarks, compared to, for instance, photos of faces or tools (cf, Aminoff et al., 2013; R. A. Epstein & Vass, 2014, for reviews). Moreover, results of Aziz-Zadeh et al. (2008), who compared hemodynamic activity levels in the PPA correlated with different categories presented in spoken sentences, revealed that semantic scene-related information also modulates the PPA's activity level. We assessed the potential of both the movie and the audio-description to replace a visual localizer in two ways. As the first approach, we modified the procedure for the analysis of data from localizer paradigms to data from the two naturalistic stimuli. Hemodynamic responses correlating with the temporal structure of annotated stimulus features (cf. Häusler & Hanke, 2016, 2021) were modeled in order to create GLM t-contrasts that aimed to localize the PPA as identified in the same participants using a visual localizer in Sengupta et al. (2016). As the second approach, we applied a functional alignment procedure as a novel method in order to estimate results from the visual localizer, the movie and the audio-description in an individual from the results of individuals in a reference group.

6.1 Open Science

In recent years, there has been a growing movement towards open science, which seeks to make scientific research more accessible, transparent, and reproducible. Open science encompasses a range of practices, including open data, open-source software, and open access publishing. The overarching aim of this dissertation was to meet the standards of open, shared, accessible, and transparent science in general, as well as the standards of a reproducible and replicable research project specifically (cf. Fecher & Friesike, 2014; Watson, 2015). This aim included using open data and open-source software, and publishing data, materials, code, and results openly available.

6.1.1 Using open data and open-source software

The first objective in the context of open science was to use open data, open materials, and open-source software. To achieve this, the thesis utilized publicly available fMRI data (Hanke, Adelhöfer, et al., 2016; Hanke et al., 2014a; Sengupta et al., 2016), subject-specific ROIs (Sengupta et al., 2016) and stimulus annotations (Häusler & Hanke, 2016) that are part of the studyforrest project (`studyforrest.org`). The analyses were implemented in freely available and, where possible, open-source software to prevent creating an "artificial paywall" for running the analyses again on the initially openly accessible data. The thesis benefited from established free software such as Python and FSL (FMRIB's Software Library; S. M. Smith et al., 2004), which have been developed and debugged for years through collaborative efforts, but also from scientific software packages like DataLad (`datalad.org`; Halchenko et al., 2021) or BrainIAK (`brainiak.org`; Kumar et al., 2021, 2020) that emerged recently. The use of pre-existing software packages, data, and results from previous analyses enabled the project to shift time and resources from software development and data collection to subsequent stages of the project. The available data proved to be extremely valuable during the COVID-19 pandemic as acquiring new fMRI data from participants became impossible, which led to a need to revise the project plan.

A first issue with open data that is frequently overlooked is that openly accessible data do not exempt the data consumer from the responsibility of carefully scrutinizing the quality of the data and the underlying experimental paradigm (e.g., stimuli and code). It is too tempting to "simply push the data through an analysis pipeline" without carefully assessing the quality of the data first. Researchers must ensure that potential errors in the data are identified and addressed before proceeding with their analysis. Consumers of datasets must assume that anything not clearly specified in the dataset's description has not been taken into account. This is particularly important as the standards (e.g., quality, formats, parameters) and open sciences practices (e.g., documenting) may differ across scientific fields or even within a scientific field depending on a working group's expertise and rigor. Even if the data are provided by a reputable source, researchers who consider using third-party data should also consider themselves to be obliged to test and validate a dataset's quality as if it was created by themselves, and in accordance with their standards and particular use case. In this regard, open data, despite being collected to the best of knowledge and belief, can be compared to open-source software: data will contain noise, errors, or artifacts, just as software will contain bugs, but "given enough eyeballs, all bugs are shallow" (Raymond, 2001, p. 30).

A second issue of open data is that the decisions made during data collection, preprocessing, and further analyses may influence or even limit subsequently performed analyses. Researchers must carefully consider previous decisions when deciding whether to use pre-existing data, or to collect or preprocess data themselves. For instance, we and Sengupta et al. (2016) selected the PPA as one of several scene-selective areas because the PPA was the first scene-selective area to be discovered and is the most reliably activated region across studies that investigate visual scene perception. However, other areas, such as the retrosplenial complex (RSC) and occipital place area (OPA), have repeatedly been shown to be involved in visual spatial perception and navigation (Bettencourt & Xu, 2013; Chrastil, 2018; Dilks, Julian, Paunov, & Kanwisher, 2013; R. A. Epstein & Baker, 2019). Although we did not explicitly hypothesize it, we assumed that at least the analysis of the movie would likely reveal significant clusters in the medial parietal and lateral occipital cortex that may correspond to the functionally defined RSC and OPA respectively. Indeed, results revealed significantly increased activity in the medial parietal and lateral occipital cortex, and provide an incentive for further studies. However, in order to create the corresponding masks of participants of the studyforrest dataset, one would have to replicate the non-automatized procedure of Sengupta et al. (2016). This example shows how decisions made during data collection, preprocessing or preceding investigations, despite being state-of-the-art at the time of being published, are affecting subsequent studies. Hence, when considering using open data, researchers need to weigh the costs and benefits of one option (such as using preprocessed data as provided) relative to an alternative option (such as preprocessing raw data differently than provided), and then choose the option that will yield the highest net return. In summary, any previous step that required human intervention or was not fully automated influences the degree to which data or materials can be replicated, updated, or extended.

6.1.2 Publishing data, materials, code, and results

The second objective in the context of open science was to publish the data, materials, and results openly available. The data and custom code created for this dissertation are version-controlled, meaning that any changes made were logged and documented, to promote transparency. To ensure reproducibility, processing steps, ranging from downloading input data to plotting figures, are implemented in scripts that can be rerun from the command line. For example, the annotation of speech has been published freely accessible (Häusler & Hanke, 2021). Its content goes far beyond what was required to conduct the analyses in Häusler et al. (2022), serves as an extension of the studyforrest project and widens the “annotation bottleneck” (Aliko et al., 2020, p. 16) of two naturalistic stimuli. In Häusler et al. (2022), we used open fMRI data to investigate a research subject that was not anticipated when the data were initially made available. The results of Häusler et al. (2022) indicate that increased hemodynamic activity in the PPA generalizes from blocks of images to spatial information embedded in a movie and an auditory narrative. These results published in a peer-reviewed journal highlight the benefits of sharing and reusing data to explore unanticipated research questions and to generate new insights.

From a negative perspective, creating data, materials, and code to be published requires a considerable amount of time and effort. To encourage third parties to reuse the data, dataset creators must anticipate potential use cases, collect the data with appropriate extent and rigor, convert the data into a standardized format (taking into account, for example, naming conventions and folder structure). Analyses pipelines need to be designed and tested in a way that they can reliably replicate every stage of a dataset. Additionally, dataset creators must take into account legal matters (such as intellectual property rights, use licenses, statements of agreement, and anonymization of participant data), facilitate discovery by humans and web bots (e.g., by including detailed descriptions and machine-readable metadata), and guarantee long-term curation and accessibility.

From a positive perspective, creating a dataset that is intended to be published has immediate benefits for a researcher. Meticulously recording each step and commenting the advantages and disadvantages of alternate procedural options leads to deeper understanding of the scientific area, its practices, and methods. The version-control of every step reduces the likelihood of a look-ahead bias. Tracking and extensively documenting each stage of the data and code from the beginning to the final results can also be thought of a lab protocol that comprises structured information for writing the corresponding scientific article. Hence, creating a dataset supposed to be published encourages precise work habits and good scientific practices in general.

6.1.3 Interim summary

The pursuit of conducting an open and reproducible research project was not mandatory for thesis submission but entailed significant additional effort and time. Since open science practices are not yet covered in graduate or PhD curricula, learning about principles and standards, as well as putting them into practice, relied on self-initiative and self-learning. The constantly emerging standards and principles, and the steadily developing software packages to apply these standards, made it difficult to put theoretical knowledge into practice. In my opinion, the time and effort needed for open science practices greatly exceed any short-term advantages. The lengthy procedures are not justified by "gambling" on being cited in the event that released data are reused, or by merely pursuing the "higher purpose" of addressing the replication crisis. Particularly, designing and testing fully automated analysis pipelines or scripts to plot complex figures without any manual finishing for the perceived sole purpose of reproducibility is out of proportion to the immediate benefit of easier bug tracking or simply "higher confidence in one's own work". On the contrary, PhD students that pursue a career in science may be concerned about exposing themselves to critique owing to a maximum of voluntary transparency and possibly (and blamelessly?) overlooked errors. Another concern is the potential for being "scooped", which refers to the risk that another working group is using the same data for a similar research question at the same time and eventually claiming priority to the research idea and its findings (cf. Laine, 2017). This risk is aggravated in case of early-career scientists that created and maintain a public dataset, pre-registered studies based on a public dataset, or have to adhere to inflexible project plans. Hence, undergraduate programs should teach the benefits and best practices, but also risks of open science, and provide practical training in related software packages. Postgraduate programs should create incentives to conduct open science projects. After all, open sciences is a suitable tool to (a) hold researchers responsible for collecting, storing, documenting, processing, and publishing data and materials in accordance with best practices, (b) increase the reproducibility of results and replicability of findings, (c) make knowledge and technologies widely accessible, (d) therefore increase the efficiency of the scientific progress and promote innovation, and (e) ultimately increase the public's trust in the scientific process and its findings.

6.2 Naturalistic stimuli for functional localization

The traditional method for identifying the PPA is to contrast hemodynamic responses to blocks of images of landscapes or landmarks with blocks of images of tools or faces. While the exact outline of the PPA varies depending on the type of stimuli, task, and contrast as well as statistical threshold, the traditional localizer approach can reliably delineate the PPA bilaterally in a large proportion of subjects (Zhen et al., 2017). Sengupta et al. (2016), for instance, were able to identify the left-hemispheric PPA in 12 of 14 subjects and right-hemispheric PPA in 14 of 14 subjects. In Häusler et al. (2022), we investigated whether the PPA, as previously identified in the same group of participants of Sengupta et al. (2016), could be localized using an audio-visual naturalistic stimulus and an exclusively auditory naturalistic stimulus. We adapted the traditional localizer approach and modeled hemodynamic responses to events in the two naturalistic stimuli to create t-contrasts that aimed to localize the PPA. For the statistical analysis (i.e. GLM) of the movie, we utilized an annotation of movie cuts and depicted locations (Häusler & Hanke, 2016). For the GLM of the audio-description, we extended the annotation of speech that we created and validated in Häusler and Hanke (2021) by further annotating nouns that the audio-description's narrator uses to describe the lacking visual content. On a group-average level, results demonstrate that increased hemodynamic activation in the PPA during the perception of static images generalizes to the perception of spatial information embedded in a movie and an auditory stimulus. We have shown that a model-driven analysis based on a naturalistic stimulus' annotation can replicate findings of studies that employed traditional paradigms. Our results provide further evidence (cf. Bartels & Zeki, 2004) that functional specialization of cortical areas is maintained during naturalistic stimulation. On an individual level, our analysis of the movie revealed bilateral clusters of increased hemodynamic activity in the PPA of five participants and a unilateral cluster in seven participants. The analysis of the audio-description revealed bilateral clusters in nine participants and a unilateral cluster in one participant. These findings suggest that a naturalistic stimulus, whether visual or auditory, could potentially replace a traditional localizer to assess brain functions in individuals.

6.2.1 Current challenges and limitations

The current thesis highlights obstacles in the pursuit of developing a multi-functional naturalistic localizer. Traditional experimental designs typically involve presenting participants with simple, well-controlled stimuli that are carefully designed to elicit clearly defined responses in targeted brain regions. This allows for modeling of hemodynamic responses that can predict and explain the neural activity observed in response to these stimuli. The currently dominant analysis approach is the mass-univariate GLM, which has its roots in positron emission tomography (PET) research, and is tailored to analyze data of parametric experimental designs that manipulate isolated experimental variables of interest. The GLM requires the researcher to specify which stimulus features are presumed to be correlated with the brain process under investigation. Then, the researcher needs to model a hypothesized hemodynamic time course that is fit to the data in order to predict the observed hemodynamic activity and contrast parameter estimates (Friston et al., 1998). Naturalistic stimuli, however, are continuous and complex, with a multitude of sensory features that can activate different brain regions simultaneously. Applying the traditional analysis approach is challenging and can lead to difficulties in isolating specific neural correlates of perceptual or cognitive processes. Moreover, properties of naturalistic stimuli stress physiological assumptions of the traditional GLM approach, such as cognitive subtraction (Friston et al., 1996), the consistency of hemodynamic responses across events (the rationale behind *trial-averaging*; cf. Dale & Buckner, 1997), and the linearity of hemodynamic responses (Boynton, Engel, Glover, & Heeger, 1996; Cohen, 1997; Dale, 1999). The properties of naturalistic stimuli also stress statistical assumptions such as the absence of collinearity among variables. The lack of experimental control, however, can be alleviated by detailed annotations that allow modeling and statistically controlling potentially confounding variables (e.g., Deniz, Nunez-Elizalde, Huth, & Gallant, 2019). Low-level visual features, such as brightness, and low-level auditory features, such as root-mean square power (i.e., volume), can be automatically extracted on a low temporal scale (e.g., per movie frame; cf. `github.com/psychoinformatics-de/studyforrest-data-confoundsannotation` for low-level annotations of the studyforrest project). Recent advances in machine learning have led to the creation tools, such as "pliers" (McNamara, de la Vega, & Yarkoni, 2017) implemented in the neuroscout platform (`neuroscout.org`; de la Vega et al., 2022), that can automatically extract high-level features like semantics or clearly defined object categories. Such tools can replace time-consuming manual annotations or provide a provisional scaffold that reduces manual labor for a growing number of stimulus features. However, variables that are difficult to define, hard to consistently label, have multiple interpretations, fluctuate on longer time scales, or subject-related variables, such as the level of engagement or felt emotions (cf. Lettieri et al., 2019; Saarimäki, 2021), still defy an automatic annotation. Hence, merely assessing the confound structure of a naturalistic stimulus as a possible candidate for a prospective model-driven analysis might become time-consuming. Moreover, a large amount of annotated features may lead to annotations that are "hard to use" (Richard et al., 2019, p. 2) or "high-dimensional, cumbersome models" (Richard et al., 2019, p. 2) eventually resulting in a lack of statistical power due to insufficient data samples.

6.2.2 Interim summary

Given the unsolved issues and current limitations, it is perhaps not surprising that, even 20 years after the group-level findings of Bartels and Zeki (2004), no functional localizer based on a movie or an auditory narrative exists that localizes high-level visual areas. The most ecologically valid visual localizers currently available are *dynamic localizers* (e.g., Fox, Iaria, & Barton, 2009; Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011) that use blocks of short videos (each video lasting ≈2-3 s) of scenes, faces, etc., making them well-suited for traditional modeling procedures.

Even naturalistic stimuli, despite being labeled as "naturalistic" and ubiquitously referred to as "more ecologically valid", are not strictly "naturalistic", but rather an approximation of real-life presented via a screen and headphones in the laboratory. Similar to traditional stimuli that have been carefully designed by researchers to probe specific brain processes, most naturalistic stimuli used in neuroscience have been carefully designed by professional media creators to appeal to their target audience. Film directors intentionally manipulate the viewers' attentional focus and mental states using a variety of techniques like camera-movement, composition, movie editing, or voice-overs (Brown, 2012; Dancyger, 2011; Katz, 1991; Mercado, 2011) that, when used correctly, largely occur unnoticed. For example, participants asked to spot movie cuts miss between 10% and 50% of them depending on the type of cut (T. J. Smith & Henderson, 2008). While these techniques reduce individual variation and lead to reliably synchronized spatial-temporal responses across subjects in a large part of the cortex (Hasson et al., 2008), they also introduce a confounding factor in the form of "naturalistic stimulus statistics".

Although naturalistic stimuli are intended to be inherently engaging and offer a task-free paradigm, sustained attention is still required by participants when "freely viewing" a movie or "freely listening" to an auditory narrative, which may become challenging over longer periods of time. The occasional disengagement of single subjects may have a negligible effect on group-average results, especially when statistics are calculated over long stimulus intervals. However, when a study aims to investigate a brain process that is expected to be universal across humans on an individual level, a varying engagement may impair the reliability of individual-level results across stimulus segments and potentially mask effects of interest. Nonetheless, a varying level of engagement may reflect a personal preference towards the presented stimulus or elicit, given the low demand characteristics (cf. Orne, 1962) of naturalistic paradigms, a more natural behavior reflecting a personal trait.

6.2.3 Estimation based on data of a reference group

Given the current limitations of modeling hemodynamic responses during naturalistic stimulation to reliably localize the PPA in all participants, the thesis aimed to explore a second approach to identify a functional area based on data of naturalistic stimuli. In

Chapter 5, we estimated the empirical Z-maps acquired from the analysis of the visual localizer in an individual by leveraging data from other individuals. Our results indicate that ≈15 min of movie data sampled at 0.5Hz used for functional alignment can estimate the results of the visual localizer more accurately than an estimation procedure based on anatomical alignment. Results suggest that a fMRI scanning session using a complex naturalistic stimulus lasting 15 to 30 minutes could serve as a functional calibration scan. This calibration scan can be used to align a new subject to a common functional space derived from data of a reference group. Once aligned with the common functional space, functional data of controlled and naturalistic paradigms can be mapped through the common model space into the new subject's voxel space. This approach enables the possibility of using a short movie to estimate the results of various localizers that are designed to provide reliable measures in all individuals. Just presenting an engaging short movie to a new subject would allow to estimate patterns that are observed in the normative reference group, when additional functional scans are not feasible due to various constraints, such as time, scanner availability, financial limitations, or compliance issues. Furthermore, this approach would allow for quantifying the similarity or dissimilarity between a new subject's actual pattern and the common pattern estimated from the normative reference. In particular, our results suggest that an auditory paradigm can be used to estimate the results of a visual paradigm. Future studies should explore the performance of functional alignment in other domains than high-visual perception. It would be intriguing to investigate whether functional alignment can be utilized to predict brain patterns that are correlated with paradigms specifically designed to elicit brain processes not captured during passive movie viewing, such as planning or decision-making. However, future developments are necessary to advance functional alignment to a clinical application. Increased sampling rates and improved functional alignment algorithms are required to reduce scanning time and the amount of data needed to functionally align a higher number of voxels simultaneously. Further developments that allow aligning a large number of voxels across larger distances would be particularly beneficial for functional areas that exhibit large interindividual variability in anatomical location, such as language areas, or in cases of atypical topography resulting from cortical reorganization after brain injuries.

6.3 Conclusions

Traditional localizer paradigms are designed to selectively elicit certain perceptual or cognitive processes in the majority of study participants. In contrast, naturalistic stimuli more closely resemble the complexity of real-world experiences, leading to responses in multiple brain regions simultaneously. As a proof of concept, the thesis focused on the PPA, a high-level visual area, to investigate the potential of a movie and its audio-description to replace a traditional visual localizer. We found that a model-driven analysis based on annotations of stimulus features embedded in the naturalistic stimuli can yield results that are consistent with previous studies using a visual localizer. A second, model-driven approach revealed that 15 minutes of functional scanning during movie watching can generate a sufficient amount of data to functionally align a subject with a common functional space and estimate brain patterns more accurately than a procedure based on anatomical alignment. Further studies are needed to investigate whether findings of this thesis generalize to other high-level visual areas and other domains of brain functions.

The thesis also highlights challenges of naturalistic stimuli. Researchers should familiarize themselves with media techniques to make informed decisions when choosing potential stimulus candidates. Moreover, the creation of annotations is crucial for a couple of reasons. First, annotations allow to quantify the confound structure of potential candidates. Second, annotations enable model-driven analyses that may reveal which stimulus features and brain responses are correlated, enabling future investigations to more effectively choose an appropriate naturalistic stimulus for a given research question or population. Last, annotations allow to test physiological assumptions, address statistical issues, and eventually adapt traditional analysis approaches and foster the development of new analysis methods.

To address the challenge of individually varying levels of engagement due to a lack of a task, fMRI scanning sessions should be accompanied with EEG, eye-tracking, physiological recordings, such as skin conductance response and heart rate, and followed up with self-reports to evaluate a participant's alertness and audio track audibility within the noisy scanner. These additional recordings may increase expenditures but also the number of use cases of a dataset. Given the versatility naturalistic stimuli, researchers who consider running an experiment that employs a naturalistic stimulus, should also consider eventually sharing the data with the scientific community. Creating a dataset with the intention of sharing it promotes best practices of collecting, documenting, storing, and processing data, and ultimately promotes a collaborative effort to advance the field.

With further investigations and refinements of analysis methods, it may be possible to use only one naturalistic stimulus to assess multiple domains of brain functions on an individual level under more ecological conditions. It is important to note that naturalistic stimuli should not necessarily be expected to yield identical results to traditional paradigms. Traditional localizers aim to minimize inter-subject variability and reliably localize functional areas in all healthy individuals. The lack of demand characteristics of naturalistic stimuli presents challenges when investigating brain responses assumed to be universal across all humans. However, naturalistic paradigms allow participants to behave more genuinely during an experiment, potentially revealing individual differences in perception and cognition that may also be correlated with genuine behavior in the real world. Therefore, naturalistic stimuli should not seek to replace traditional paradigms and it is important for researchers to acknowledge the strengths and limitations of both paradigms.

7 References

Alday, P. M. (2019). M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457–473.

Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. I. (2020). A "Naturalistic Neuroimaging Database" for understanding the brain using ecological stimuli. *Scientific Data*, *7*(1), 1–21.

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–390.

Anderson, M. J., Capota, M., Turek, J. S., Zhu, X., Willke, T. L., Wang, Y., ... Norman, K. A. (2016). Enabling factor analysis on thousand-subject neuroimaging datasets. In *2016 IEEE International Conference on Big Data* (p. 1151–1160).

Aziz-Zadeh, L., Fiebach, C. J., Naranayan, S., Feldman, J., Dodge, E., & Ivry, R. B. (2008). Modulation of the FFA and PPA by language related to faces and places. *Social Neuroscience*, *3*(3-4), 229–238.

Baker, M. (2016). Reproducibility crisis. *Nature*, *533*(26), 353–66.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., ... others (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Bartels, A., & Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping*, *21*(2), 75–85.

Bazeille, T., Dupre, E., Richard, H., Poline, J.-B., & Thirion, B. (2021). An empirical evaluation of functional alignment using inter-subject decoding. *NeuroImage*, *245*, 1–11.

Benson, N. C., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2014). Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. *PLoS Computational Biology*, *10*(3), 1–9.

Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *Journal of Cognitive Neuroscience*, *25*(10), 1711–1722.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*(13), 4207–4221.

Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American educational research journal*, *5*(4), 437–474.

Brown, B. (2012). *Cinematography. Theory and Practice. Image Making for Cinematographers and Directors* (2nd ed.). Oxford: Focal Press.

Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A reduced-dimension fMRI shared response model. *Advances in Neural Information Processing Systems*, *28*, 1–9.

Chrastil, E. R. (2018). Heterogeneity in human retrosplenial cortex: A review of function and connectivity. *Behavioral neuroscience*, *132*(5), 317–338.

Coalson, T. S., Van Essen, D. C., & Glasser, M. F. (2018). The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proceedings of the National Academy of Sciences*, *115*(27), 1–10.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *Neuroimage*, *6*(2), 93–103.

Cohen, M. S., & Bookheimer, S. Y. (1994). Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences*, *17*(7), 268–277.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

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

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human brain mapping*, *5*(5), 329–340.

Dancyger, K. (2011). *The Technique of Film and Video Editing: History, Theory, and Practice.* Oxford: Focal Press.

de la Vega, A., Rocca, R., Blair, R. W., Markiewicz, C. J., Mentch, J., Kent, J. D., … Yarkoni, T. (2022). Neuroscout, a unified platform for generalizable and reproducible fMRI research. *Elife*, *11*, e79277.

Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *Journal of Neuroscience*, 0675–19.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*(4), 1331–1336.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Drobyshevsky, A., Baumann, S. B., & Schneider, W. (2006). A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. *NeuroImage*, *31*(2), 732–744.

Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443.

DuPre, E., Hanke, M., & Poline, J.-B. (2020). Nature abhors a paywall: How open science can realize the potential of naturalistic stimuli. *NeuroImage*, *75*, 116330.

Eglen, S. J., Marwick, B., Halchenko, Y. O., Hanke, M., Sufi, S., Gleeson, P., … others (2017). Toward standard practices for sharing computer code and programs in neuroscience. *Nature Neuroscience*, *20*(6), 770–773.

Eickhoff, S. B., Constable, R. T., & Yeo, B. T. (2018). Topographic organization of the cerebral cortex and brain cartography. *NeuroImage*, *170*, 332–347.

Eickhoff, S. B., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *NeuroImage*, *217*, 116860.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. *Annual Review of Vision Science*, *5*, 373–397.

Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Epstein, R. A., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1635), 20120533.

Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. In *Opening Science* (p. 17–47). Springer, Cham.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of neurophysiology*, *104*(2), 1177–1194.

Feilong, M., Nastase, S. A., Guntupalli, J. S., & Haxby, J. V. (2018). Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage*, *183*, 375–386.

Fernández, G., de Greiff, A., von Oertzen, J., Reuber, M., Lun, S., Klaver, P., … Elger, C. E. (2001). Language mapping in less than 15 minutes: real-time functional MRI during routine clinical investigation. *NeuroImage*, *14*(3), 585–594.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., … others (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327.

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Human brain mapping*, *30*(5), 1637–1651.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, *7*(1), 30–40.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*(2), 97–104.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*(4), 1077–1087.

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369–1381.

Guntupalli, J. S., Feilong, M., & Haxby, J. V. (2018). A computational model of shared fine-scale structure in the human connectome. *PLoS Computational Biology*, *14*(4), e1006120.

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cerebral Cortex*, *26*(6), 2919–2934.

Halchenko, Y. O., Meyer, K., Poldrack, B., Solanky, D. S., Wagner, A. S., Gors, J., … Hanke, M. (2021). DataLad: distributed system for joint management of code, data, and their relationship. *Journal of Open Source Software*, *6*(63), 3262. doi: 10.21105/joss.03262

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582.

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., … Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific Data*, *3*(160092), 1–15.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., … Stadler, J. (2014a). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific Data*, *1*(1), 1–18.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., … Stadler, J. (2014b). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *OpenfMRI*. Retrieved from `https://legacy.openfmri.org/dataset/ds000113`

Hanke, M., Kottke, D., Iacovella, V., Hoffmann, M. B., Sengupta, A., Kaule, F. R., … Stadler, J. (2016). Simultaneous fMRI/eyetracking while movie watching, plus visual localizers. *OpenfMRI*. Retrieved from `https://legacy.openfmri.org/dataset/ds000113d`

Hanke, M., Pestilli, F., Wagner, A. S., Markiewicz, C. J., Poline, J.-B., & Halchenko, Y. O. (2021). In defense of decentralized research data management. *Neuroforum*, *27*(1), 17–25.

Hasson, U., & Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272–1278.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, *2*(1), 1–26.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, 116561.

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

Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, *9*, e56601.

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

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *Journal of Neuroscience*, *32*(44), 15277–15283.

Häusler, C. O., Eickhoff, S. B., & Hanke, M. (2022). Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area". *Scientific Data*, *9*(1). doi: 10.1038/s41597-022-01250-4

Häusler, C. O., & Hanke, M. (2016). An annotation of cuts, depicted locations, and temporal progression in the motion picture “Forrest Gump”. *F1000Research*, *5*(2273), 1–5.

Häusler, C. O., & Hanke, M. (2021). A studyforrest extension, an annotation of spoken language in the German dubbed movie “Forrest Gump” and its audio-description. *F1000Research*, *10*(54). doi: 10.12688/f1000research.27621.1

Jiahui, G., Feilong, M., di Castello, M. V. O., Guntupalli, J. S., Chauhan, V., Haxby, J. V., & Gobbini, M. I. (2020). Predicting individual face-selective topography using naturalistic stimuli. *NeuroImage*, *216*, 116458.

Jiahui, G., Feilong, M., Nastase, S. A., Haxby, J. V., & Gobbini, M. I. (2022). Cross-movie prediction of individualized functional topography. *bioRxiv*, 2022–11.

Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, *224*, 117445.

Kandylaki, K. D., & Bornkessel-Schlesewsky, I. (2019). *From story comprehension to the neurobiology of language* (Vol. 34) (No. 4). Taylor & Francis.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128.

Katz, S. D. (1991). *Film Directing Shot by Shot: Visualizing from Concept to Screen* (1st ed.). Study City: Michael Wiese Productions.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., ... others (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, *46*(3), 786–802.

Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., ... others (2021). BrainIAK: The brain imaging analysis kit. *Aperture Neuro*, *1*(4), 19.

Kumar, M., Ellis, C. T., Lu, Q., Zhang, H., Capotă, M., Willke, T. L., ... Norman, K. A. (2020). BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLoS Computational Biology*, *16*(1), e1007549.

Labs, A., Reich, T., Schulenburg, H., Boennen, M., Mareike, G., Golz, M., ... others (2015). Portrayed emotions in the movie “Forrest Gump” [version 1; referees: 2 approved]. *F1000Research*, *4*(92), 1–14.

Laine, H. (2017). Afraid of scooping: Case study on researcher strategies against fear of scooping in the context of open science. *Data Science Journal*.

Langers, D. R. (2014). Assessment of tonotopically organised subdivisions in human auditory cortex using volumetric and surface-based cortical alignments. *Human Brain Mapping*, *35*(4), 1544–1561.

Leek, J. T., & Jager, L. R. (2017). Is most published research really false? *Annual Review of Statistics and Its Application*, *4*, 109–122.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906–2915.

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., ... Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, *10*(1), 1–13.

McNamara, Q., de la Vega, A., & Yarkoni, T. (2017). Developing a comprehensive framework for multimodal feature extraction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (p. 1567–1574).

Mercado, G. (2011). *The Filmmaker’s Eye: Learning (and Breaking) the Rules of Cinematic Composition* (1st ed.). Focal Press.

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). *Measuring shared responses across subjects using intersubject correlation* (Vol. 14) (No. 6). Oxford University Press.

Nastase, S. A., Liu, Y.-F., Hillman, H., Norman, K. A., & Hasson, U. (2019). Leveraging shared connectivity to aggregate heterogeneous datasets into a common response space. *BioRxiv*, *217*, 741975.

Natu, V. S., Arcaro, M. J., Barnett, M. A., Gomez, J., Livingstone, M., Grill-Spector, K., & Weiner, K. S. (2021). Sulcal depth in the medial ventral temporal cortex predicts the location of a place-selective region in macaques, children, and adults. *Cerebral Cortex*, *31*(1), 48–61.

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., … others (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3), 299–303.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., … others (2022). Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology*, *73*, 719–748.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123–135.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783.

Pernet, C., & Poline, J.-B. (2015). Improving functional magnetic resonance imaging reproducibility. *Gigascience*, *4*(1), 1–8.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., … Dehaene, S. (2007). Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC neuroscience*, *8*(1), 91.

Pinho, A. L., Amadon, A., Gauthier, B., Clairis, N., Knops, A., Genon, S., … others (2020). Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping. *Scientific Data*, *7*(1), 1–16.

Pinho, A. L., Amadon, A., Ruest, T., Fabre, M., Dohmatob, E., Denghien, I., … others (2018). Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Scientific Data*, *5*(1), 1–15.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage*, *56*(4), 2356–2363.

Plesser, H. E. (2018). Reproducibility vs. replicability: a brief history of a confused terminology. *Frontiers in neuroinformatics*, *11*, 76.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., … Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2019). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA psychiatry*, *77*(5), 534–540.

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

Raichle, M. E. (2009). A brief history of human brain mapping. *Trends in Neurosciences*, *32*(2), 118–126.

Raymond, E. S. (2001). *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (Hardcover ed.). O'Reilly.

Richard, H., Martin, L., Pinho, A. L., Pillow, J., & Thirion, B. (2019). Fast shared response model for fMRI data. *arXiv preprint arXiv:1909.12537*, 1–26.

Rosenke, M., van Hoof, R., van den Hurk, J., Grill-Spector, K., & Goebel, R. (2021). A probabilistic functional atlas of human occipito-temporal visual cortex. *Cerebral Cortex*, *31*(1), 603–619.

Saarimäki, H. (2021). Naturalistic stimuli in affective neuroimaging: a review. *Frontiers in human neuroscience*, *15*, 1–25.

Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., & Haxby, J. V. (2010). Function-based intersubject alignment of human cortical anatomy. *Cerebral cortex*, *20*(1), 130–140.

Savoy, R. L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta psychologica*, *107*(1-3), 9–42.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: a defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419–436.

Sengupta, A., Kaule, F. R., Guntupalli, J. S., Hoffmann, M. B., Häusler, C., Stadler, J., & Hanke, M. (2016). A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Scientific Data*, *3*(160093), 1–14.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696.

Silva, M. A., See, A. P., Essayed, W. I., Golby, A. J., & Tie, Y. (2018). Challenges and techniques for presurgical brain mapping with functional MRI. *NeuroImage: Clinical*, *17*, 794–803.

Simony, E., & Chang, C. (2020). Analysis of stimulus-induced brain dynamics during naturalistic paradigms. *NeuroImage*, *216*, 116461.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., … others (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, 208–219.

Smith, T. J., & Henderson, J. M. (2008). Edit blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, *2*(2), 1–17.

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, *99*, 301–311.

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience*, *35*(36), 12412–12424.

Stupple, A., Singerman, D., & Celi, L. A. (2019). The reproducibility crisis in the age of digital medicine. *NPJ digital medicine*, *2*(1), 1–3.

Szaflarski, J. P., Gloss, D., Binder, J. R., Gaillard, W. D., Golby, A. J., Holland, S. K., … others (2017). Practice guideline summary: Use of fMRI in the presurgical evaluation of patients with epilepsy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*, *88*(4), 395–402.

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, *36*, 100600.

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral cortex*, *25*(10), 3911–3931.

Watson, M. (2015). When will 'open science'become simply 'science'? *Genome biology*, *16*(1), 1–3.

Wegrzyn, M., Aust, J., Barnstorf, L., Gippert, M., Harms, M., Hautum, A., ... others (2018). Thought experiment: Decoding cognitive processes from the fMRI data of one individual. *PLoS One*, *13*(9), e0204338.

Weiner, K. S., Barnett, M. A., Witthoft, N., Golarai, G., Stigliani, A., Kay, K. N., ... others (2018). Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *NeuroImage*, *170*, 373–384.

Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., ... Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage*, *84*, 453–465.

Westfall, J., Nichols, T. E., & Yarkoni, T. (2016). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*, 1–24. doi: 10.12688/wellcomeopenres.10298.2

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... others (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1–9.

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., ... others (2014). Best practices for scientific computing. *PLoS biology*, *12*(1), e1001745.

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230–242.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386.

Yates, T. S., Ellis, C. T., & Turk-Browne, N. B. (2021). Emergence and organization of adult brain function throughout child development. *NeuroImage*, *226*, 117606.

Yeo, B. T., Sabuncu, M. R., Vercauteren, T., Ayache, N., Fischl, B., & Golland, P. (2009). Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*, *29*(3), 650–668.

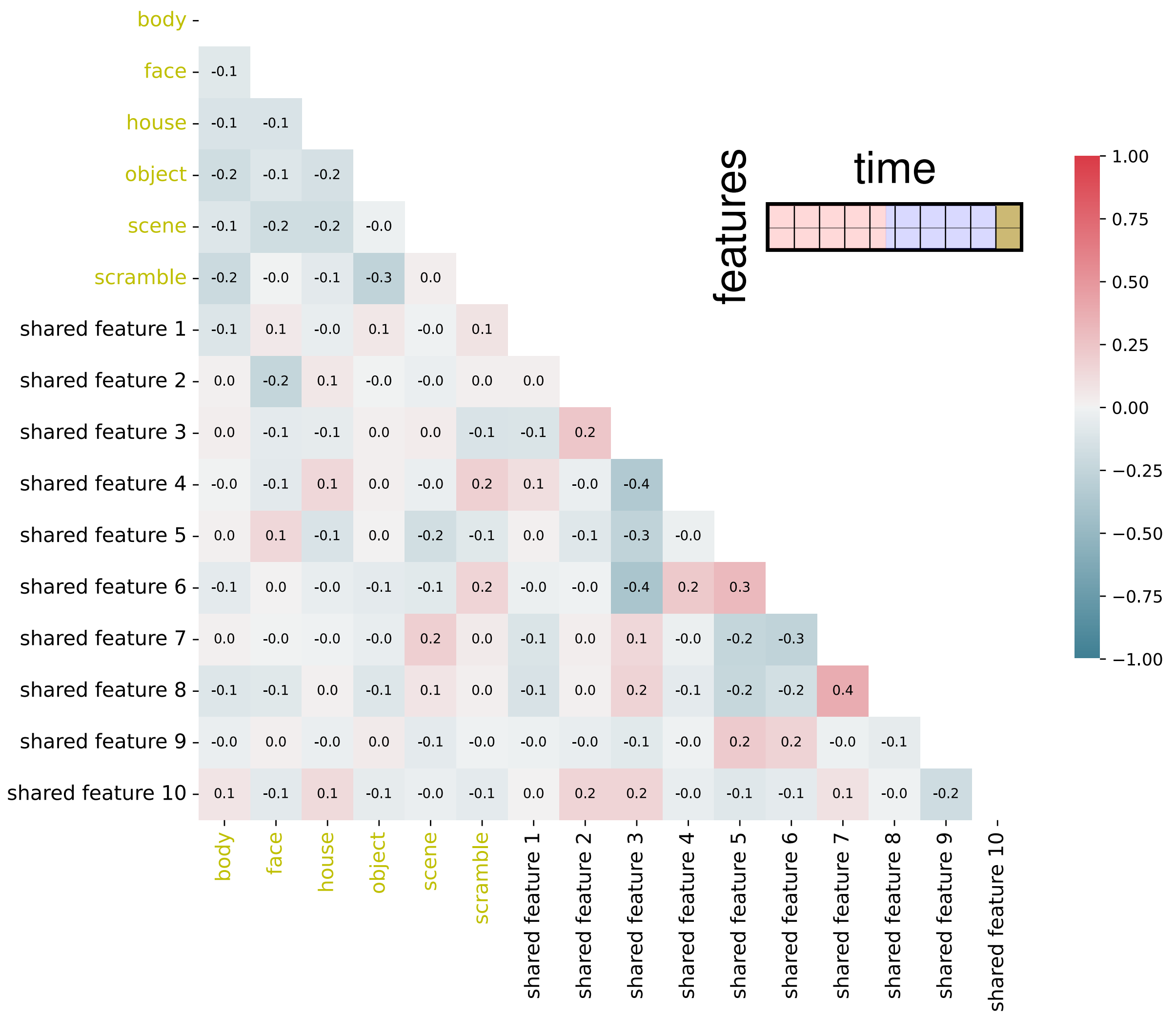
Zemeckis, R. (1994). *Forrest Gump [motion picture].* United States, Paramount Pictures.

Zhang, H., Chen, P.-H., Chen, J., Zhu, X., Turek, J. S., Willke, T. L., ... Ramadge, P. J. (2016). A searchlight factor model approach for locating shared information in multi-subject fMRI analysis. *arXiv preprint arXiv:1609.09432*.

Zhen, Z., Kong, X.-Z., Huang, L., Yang, Z., Wang, X., Hao, X., ... Liu, J. (2017). Quantifying the variability of scene-selective regions: Interindividual, interhemispheric, and sex differences. *Human Brain Mapping*, *38*(4), 2260–2275.

Zhen, Z., Yang, Z., Huang, L., zhen Kong, X., Wang, X., Dang, X., ... Liu, J. (2015). Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. *NeuroImage*, *113*, 13–25.

Appendix



	body	face	house	object	scene	scramble	shared feature 1	shared feature 2	shared feature 3	shared feature 4	shared feature 5	shared feature 6	shared feature 7	shared feature 8	shared feature 9	shared feature 10
body																
face	-0.1															
house	-0.1	-0.1														
object	-0.2	-0.1	-0.2													
scene	-0.1	-0.2	-0.2	-0.0												
scramble	-0.2	-0.0	-0.1	-0.3	0.0											
shared feature 1	-0.1	0.1	-0.0	0.1	-0.0	0.1										
shared feature 2	0.0	-0.2	0.1	-0.0	-0.0	0.0	0.0									
shared feature 3	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	0.2								
shared feature 4	-0.0	-0.1	0.1	0.0	-0.0	0.2	0.1	-0.0	-0.4							
shared feature 5	0.0	0.1	-0.1	0.0	-0.2	-0.1	0.0	-0.1	-0.3	-0.0						
shared feature 6	-0.1	0.0	-0.0	-0.1	-0.1	0.2	-0.0	-0.0	-0.4	0.2	0.3					
shared feature 7	0.0	-0.0	-0.0	-0.0	0.2	0.0	-0.1	0.0	0.1	-0.0	-0.2	-0.3				
shared feature 8	-0.1	-0.1	0.0	-0.1	0.1	0.0	-0.1	0.0	0.2	-0.1	-0.2	-0.2	0.4			
shared feature 9	-0.0	0.0	-0.0	0.0	-0.1	-0.0	-0.0	-0.0	-0.1	-0.0	0.2	0.2	-0.0	-0.1		
shared feature 10	0.1	-0.1	0.1	-0.1	-0.0	-0.1	0.0	0.2	0.2	-0.0	-0.1	-0.1	0.1	-0.0	-0.2	

Fig. A.1: **Mean similarity of hemodynamic responses modeled for the analysis of the visual localizer in Sengupta et al. (2016) and shared features in 1000 shared response models based on randomly shuffled runs in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients between modeled responses and shared responses within one of the multi-paradigm CFSs, the time series of the respective CFS's shared features were trimmed to match the corresponding TRs of the visual localizer paradigm (Sengupta et al., 2016). The modeled hemodynamic responses (i.e. regressors) represent predicted responses to the six categories of pictures that were presented in blocks.

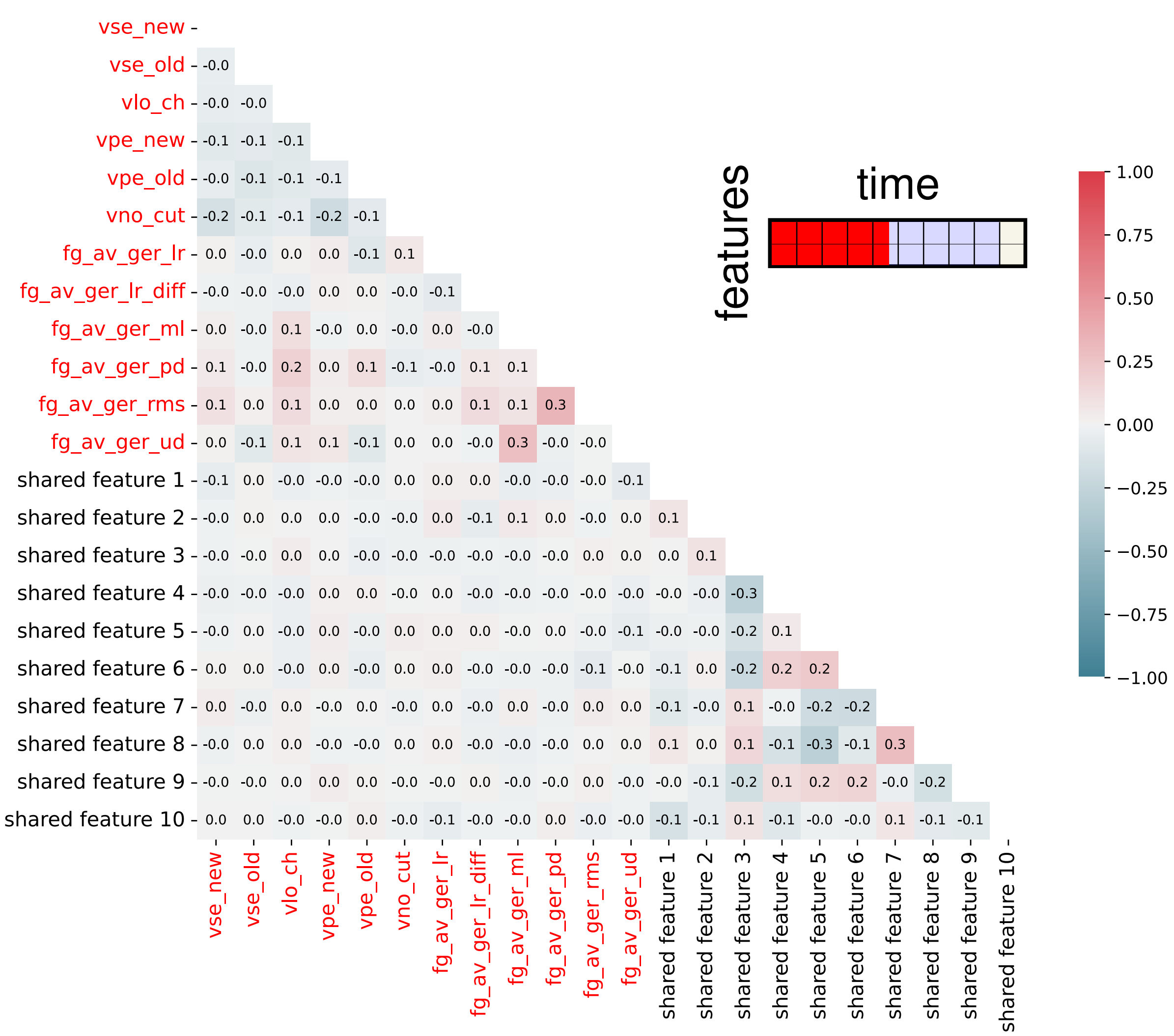


Fig. A.2: **Mean similarity of hemodynamic responses modeled for the analysis of the movie in Häusler et al. (2022) and shared features in 1000 shared response models based on randomly shuffled runs in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients between modeled responses and shared responses within one of the multi-paradigm CFSs, the time series of the respective CFS's shared features were trimmed to match the corresponding TRs of the movie (Hanke, Adelhöfer, et al., 2016). The modeled shared responses (i.e. regressors) `vse_new` to `vno_cut` are based on annotations of movie frames, whereas the regressors `fg_av_ger_lr` to `fg_av_ger_ud` represent low-level visual or auditory confounds (cf. Table 3 in Häusler et al., 2022). `vse_new`: change of the camera position to a setting not depicted before; `vse_old`: change of the camera position to a recurring setting; `vlo_ch`: change of the camera position to another locale within the same setting; `vpe_new`: change of the camera position within a locale not depicted before; `vpe_old`: change of the camera position within a recurring locale; `vno_cut`: a pseudorandomly selected frames within a continuous movie shot; `fg_av_ger_lr`: left-right luminance difference; `fg_av_ger_lrdiff`: left-right volume difference; `fg_av_ger_ml`: mean luminance; `fg_av_ger_pd`: perceptual difference; `fg_av_ger_rms`: root mean square volume; `fg_av_ger_ud`: upper-lower luminance difference.

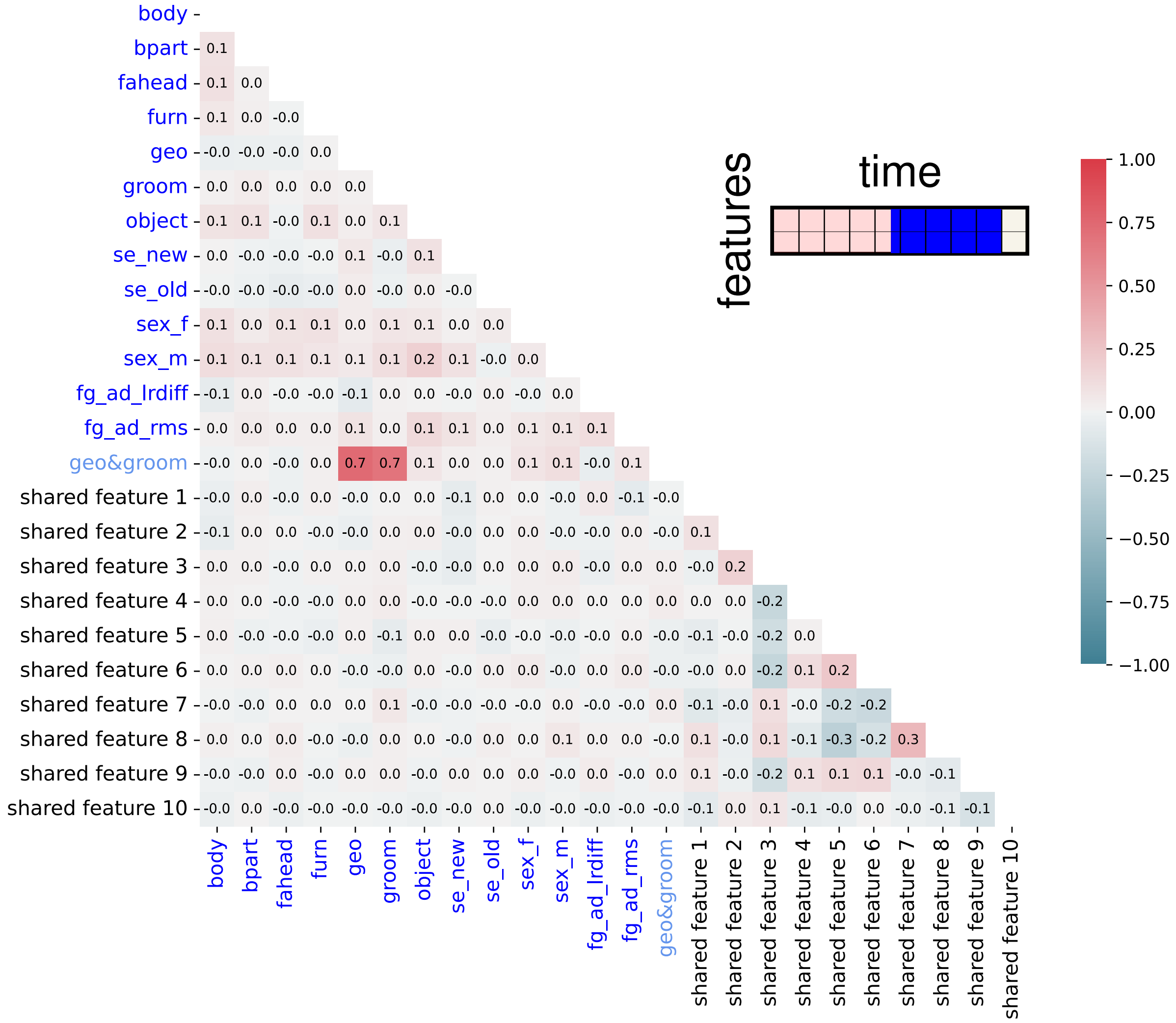


Fig. A.3: **Mean similarity of hemodynamic responses modeled for the analysis of the audio-description in Häusler et al. (2022) and shared features in 1000 shared response models based on randomly shuffled runs in the first fold of the cross-validation.** Before calculating the Pearson correlation coefficients between modeled responses and shared responses within one of the multi-paradigm CFSs, the time series of the respective CFS's shared features were trimmed to match the corresponding TRs of the visual localizer paradigm (Sengupta et al., 2016). The modeled shared responses (i.e. regressors) `body` to `sex_m` are based on annotated categories of nouns spoken by the audio-description's narrator, whereas the regressors `fg_ad_ger_lrdiff` and `fg_ad_ger_rms` represent low-level auditory confounds (cf. Table 3 in Häusler et al., 2022). `body`: trunk of the body; overlaid clothes; `bpart`: limbs and trousers; `fahead`: (parts) of the face or head; `furn`: moveable furniture (insides & outsides); `geo`: immobile landmarks; `groom`: rooms & locales or geometry-defining elements; `object`: moveable and countable entities with firm boundaries; `se_new`: a setting occurring for the first time; `se_old`: a recurring setting; `sex_f`: female name, female person(s); `sex_m`: male name, male person(s); `fg_ad_lrdiff`: left-right volume difference; `fg_ad_rms`: root mean square volume. `geo&groom` is a combination of regressors as used on the positive side of the primary contrasts aimed to localize the PPA (cf. Table 5 in Häusler et al., 2022).