



Мозговые механизмы узнавания человека по лицу и голосу

Ольга Королькова

Институт экспериментальной психологии МГППУ

olga.kurakova@gmail.com

Multisensory processing of faces and voices in person identity recognition

(Мультисенсорная переработка лиц и голосов при распознавании человека)



Lúcia Garrido
City, University of
London
(руководитель)

Nikolaus Kriegeskorte
Columbia University
(соруководитель)

Carolyn McGettigan
University College
London
(соруководитель)

Nadine Lavan
University
College London
(постдок)

Maria Tsantani
Birkbeck,
University of
London
(аспирант)

Проект поддержан фондом
LEVERHULME
TRUST



Brunel
University
London



Центральный Лондон



Brunel University London



Royal Holloway University London



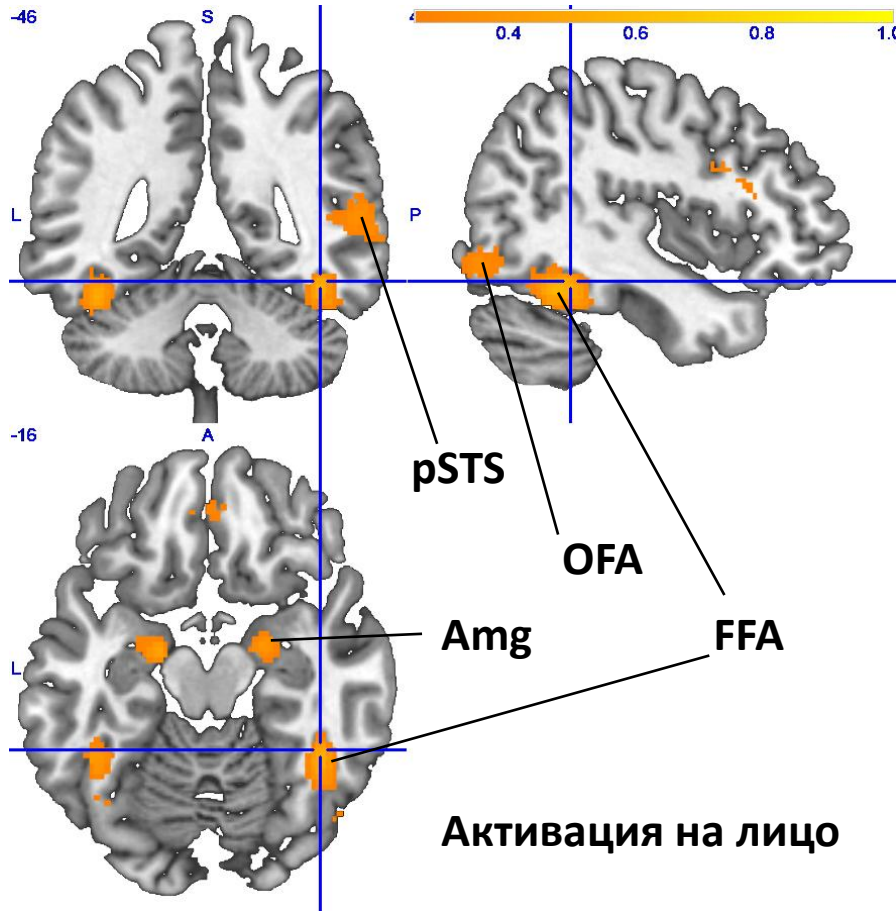
Brunel University London

Royal Holloway University with a parkland campus

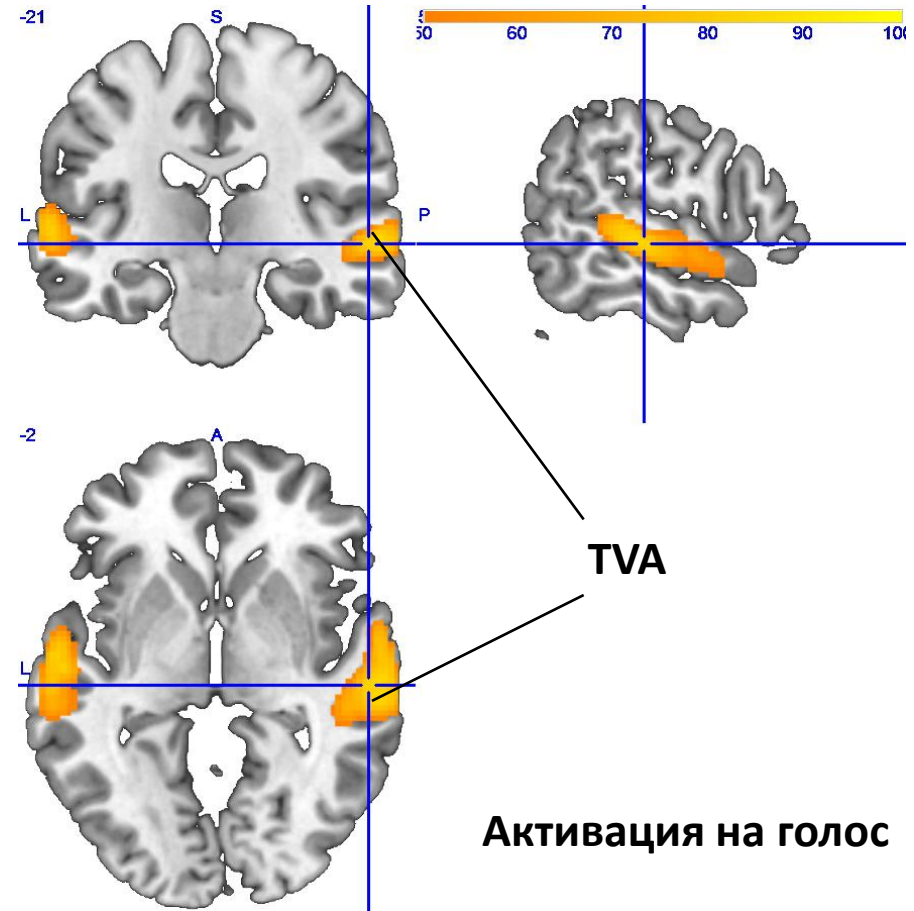


Проблема исследования

Описаны зоны мозга, избирательно активирующиеся на изображения лица, и зоны, активирующиеся на звуки голоса



Engell & McCarthy, 2013

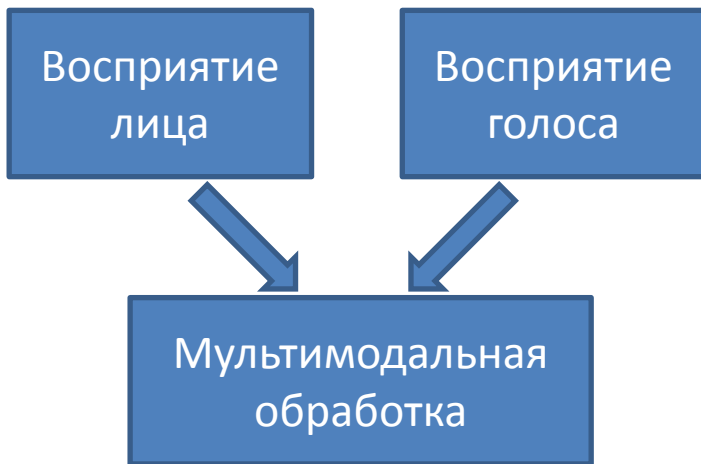


Pernet et al., 2015

Проблема исследования

Как в мозге происходит интеграция мультимодальной информации при узнавании человека по лицу и по голосу?

Гипотеза 1: существует специальная зона интеграции унимодальных ощущений в целостный образ (Bruce & Young, 1986; Campanella & Belin, 2007)



Гипотеза 2: интеграция обеспечивается непосредственными функциональными связями между зонами унимодальной обработки (von Kriegstein et al., 2005)



Цели проекта

Выявить и описать мозговые зоны переработки мультимодальной информации о человеке

Описать структуру репрезентаций в уни- и мультимодальных областях при восприятии лица и голоса

Сопоставить репрезентации лиц и голосов до и после «знакомства» с ранее неизвестными людьми

Соотнести активацию в этих областях с субъективным пространством восприятия знакомых людей и с индивидуальными особенностями распознавания и запоминания лиц и голосов

Исследование 1

Какие области мозга участвуют в распознавании лиц и голосов?

- Нарушения восприятия и лиц, и голосов у пациентов с повреждениями передней височной коры (Ellis, Young, & Critchley, 1989; Neuner & Schweinberger, 2000; Gainotti, 2011)
- Сходные паттерны активации на знакомые лица и голоса в левой передней височной коре при интракраниальной ЭЭГ (Abel et al., 2015)
- Мультисенсорные зоны в верхней височной извилине у макак (Perrodin et al., 2014)
- Связь повреждений в латеральной теменной коре с нарушениями распознавания лиц и голосов (Roswadowitz et al., 2018, повоксельное картирование повреждений мозга и поведенческих нарушений)
- фМРТ-активация на лица и на голоса в задней поясной коре, угловой извилине, гиппокампе, лобной коре, правой верхней височной извилине и борозде (Shah et al., 2011; Deen et al., 2015; Joassin et al., 2011; Watson et al., 2014; Tsantani et al., 2019)

Мультимодальные зоны восприятия людей

Области мозга, которые избирательно активируются при восприятии лиц, при восприятии голосов и при восприятии аудиовизуальных стимулов людей

- **Амодальные зоны:** одинаковые репрезентации людей независимо от модальности стимула (Jung, Larsen, & Walther, 2018)
- **Зоны мультисенсорной интеграции:** репрезентации мультимодальных стимулов отличаются от суммы унимодальных репрезентаций (Beauchamp, 2005; Calvert, Campbell, & Brammer, 2000; Laurienti et al., 2005; Noppeney, 2012; Perrodin et al., 2015; Stevenson et al., 2014)
- **Мозаика унимодальных зон** + недостаточное пространственное разрешение фМРТ (Beauchamp et al., 2004; Gentile et al., 2017)

$$AV = A = V$$

$$AV \neq (A + V)$$

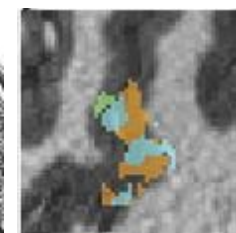
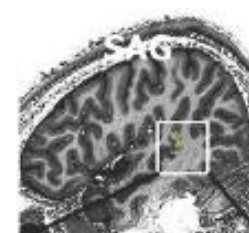
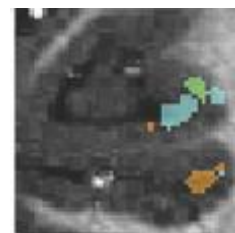
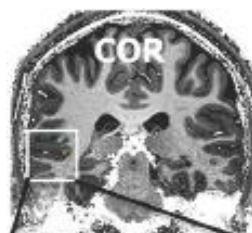
$$AV > (A + V)$$

$$AV < (A + V)$$

$$AV > \max(A, V)$$

$$AV > \text{mean}(A, V)$$

7 Tesla,
1.1×1.1×1.1 мм



Метод исследования

T2*-взвешенная фМРТ (3×3×3 мм, TR = 2000 мс)

T1-взвешенная структурная МРТ (1×1×1 мм)

2 группы участников: N1 = 30; N2 = 22; 18-27 лет

Функциональные локалайзеры:

- Зрительный (видео без звука)
- Аудиальный (звук без изображения)
- Аудиовизуальный (видео со звуком)



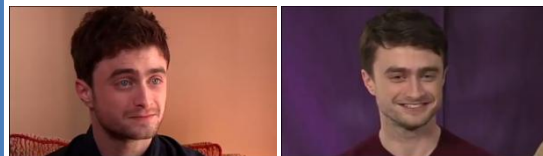
Siemens Tim Trio (3 Tesla)

Задача – нажимать на кнопку при повторе стимула

Стимулы: лица/голоса известных и незнакомых людей, видео/звуки сцен
(длительность 8 с)

2 подхода каждого локалайзера

+



Лица (2×8с)



Сцены (2×8с)

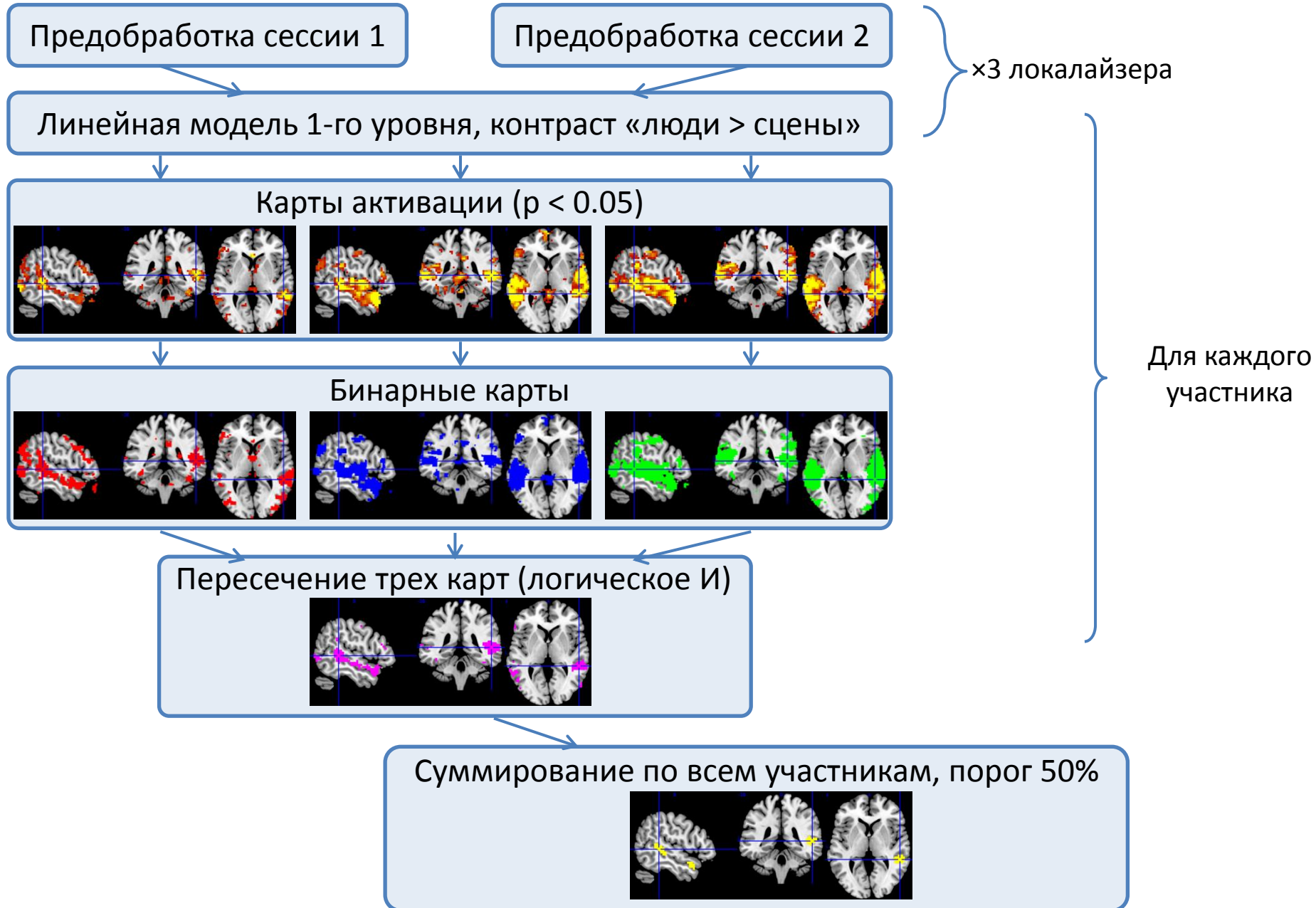
+

Фиксация 16с

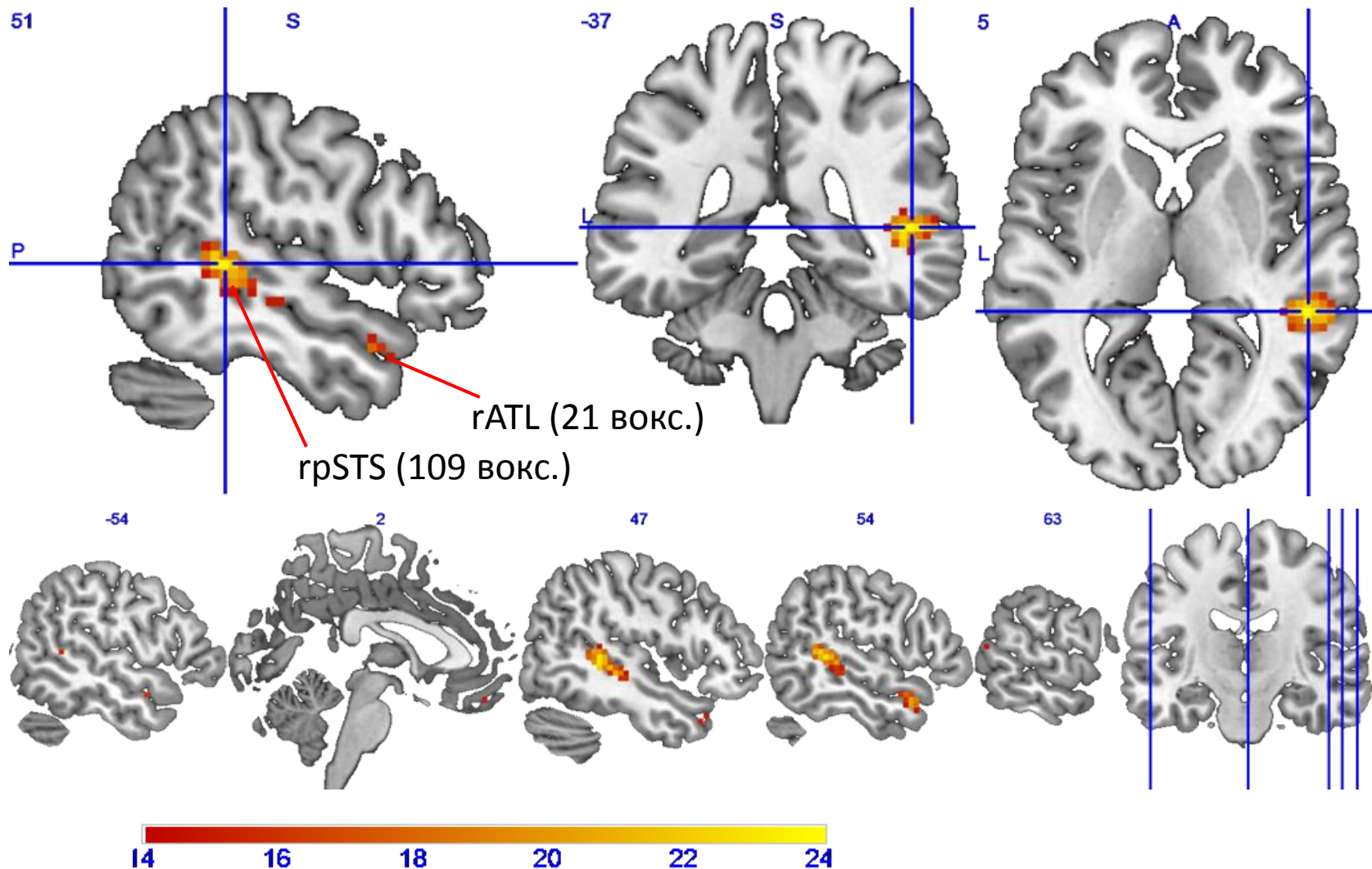
Фиксация 16с



Анализ данных: серия 1 (N = 30)



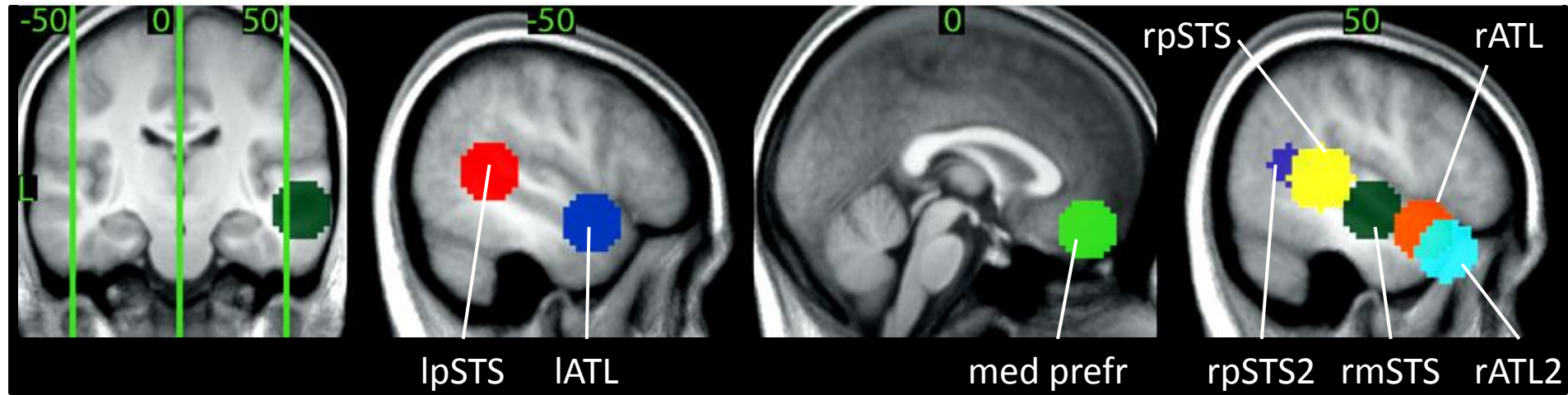
Результаты: области избирательной активации на лица/голоса



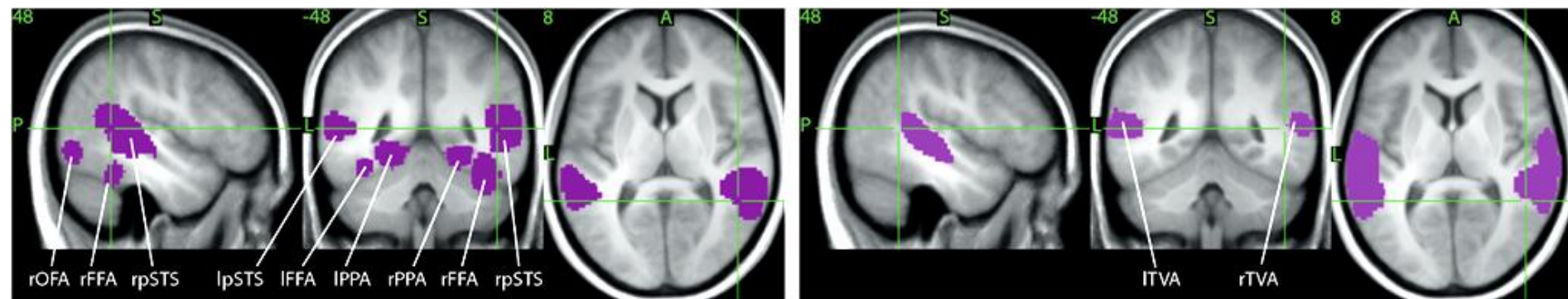
Число участников, имеющих значимую избирательную активацию на лица, голоса и аудиовизуальные стимулы людей

Определение зон интереса

Мультимодальные групповые маски



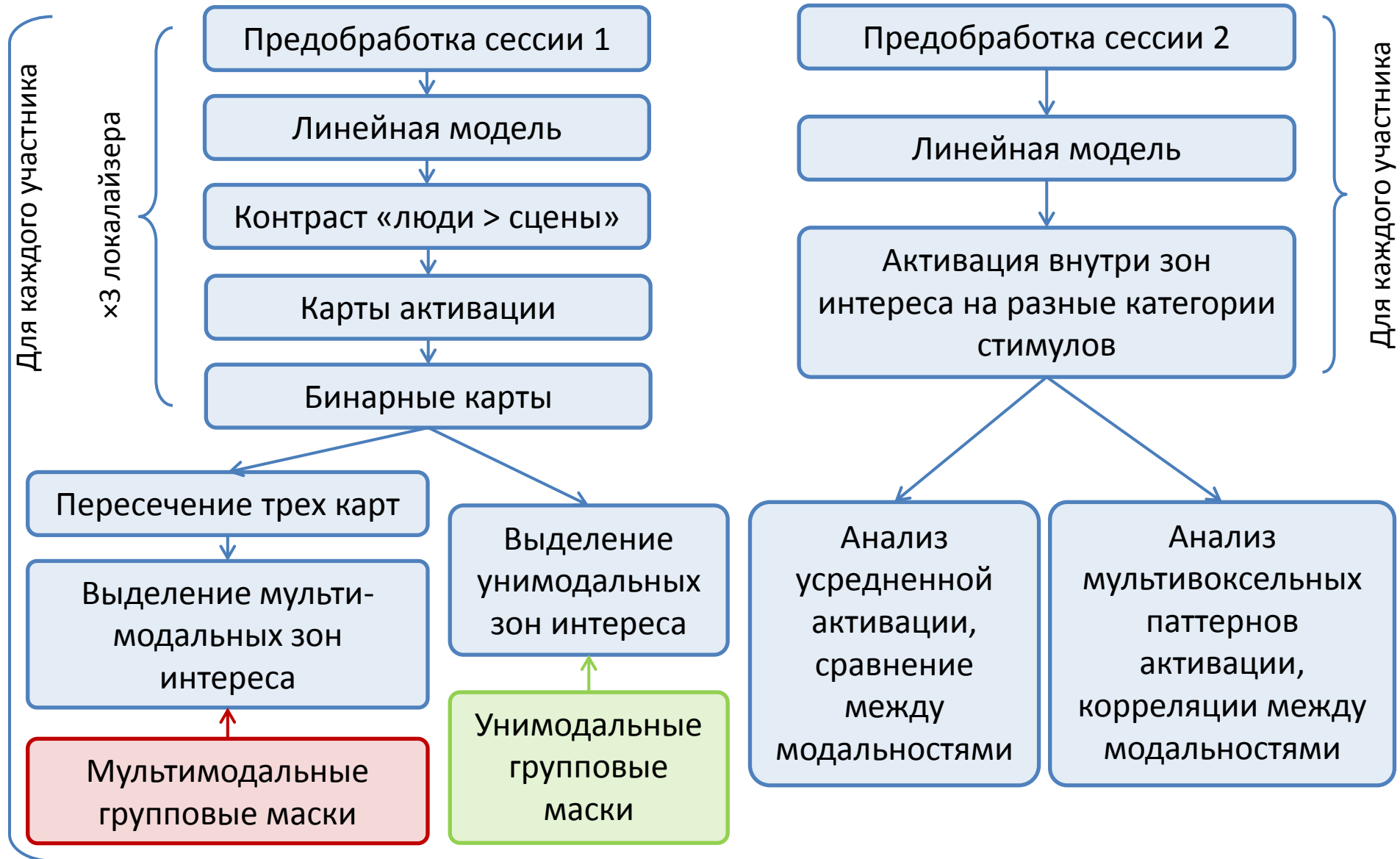
Унимодальные групповые маски



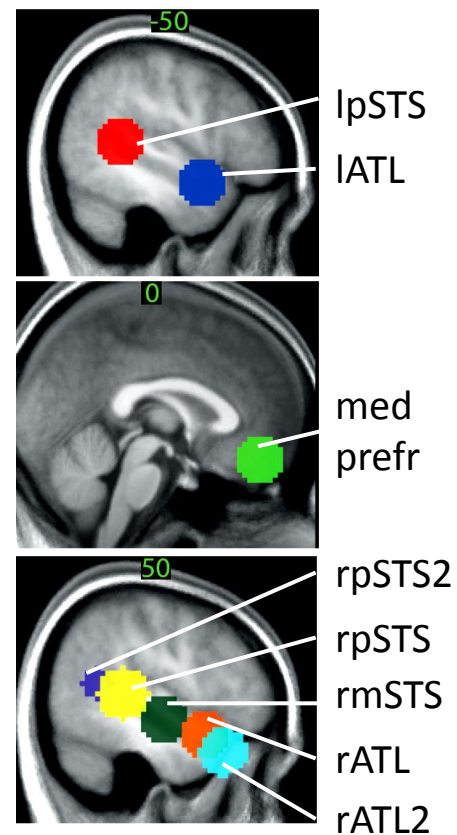
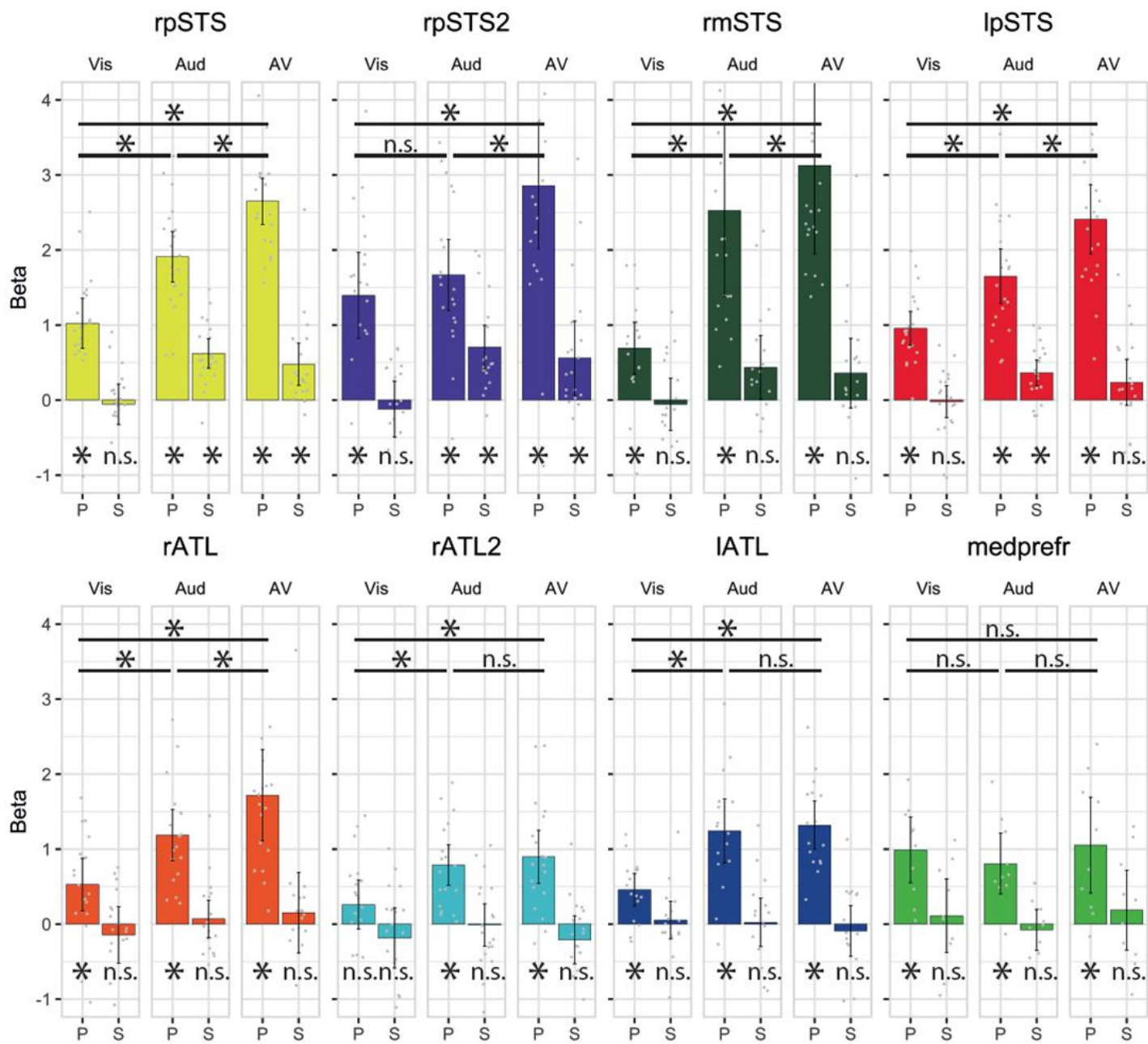
Зоны лица и зоны сцен (Julian et al., 2012)

Зоны голоса (Pernet et al., 2015)

Анализ данных: серия 2 (N = 22)

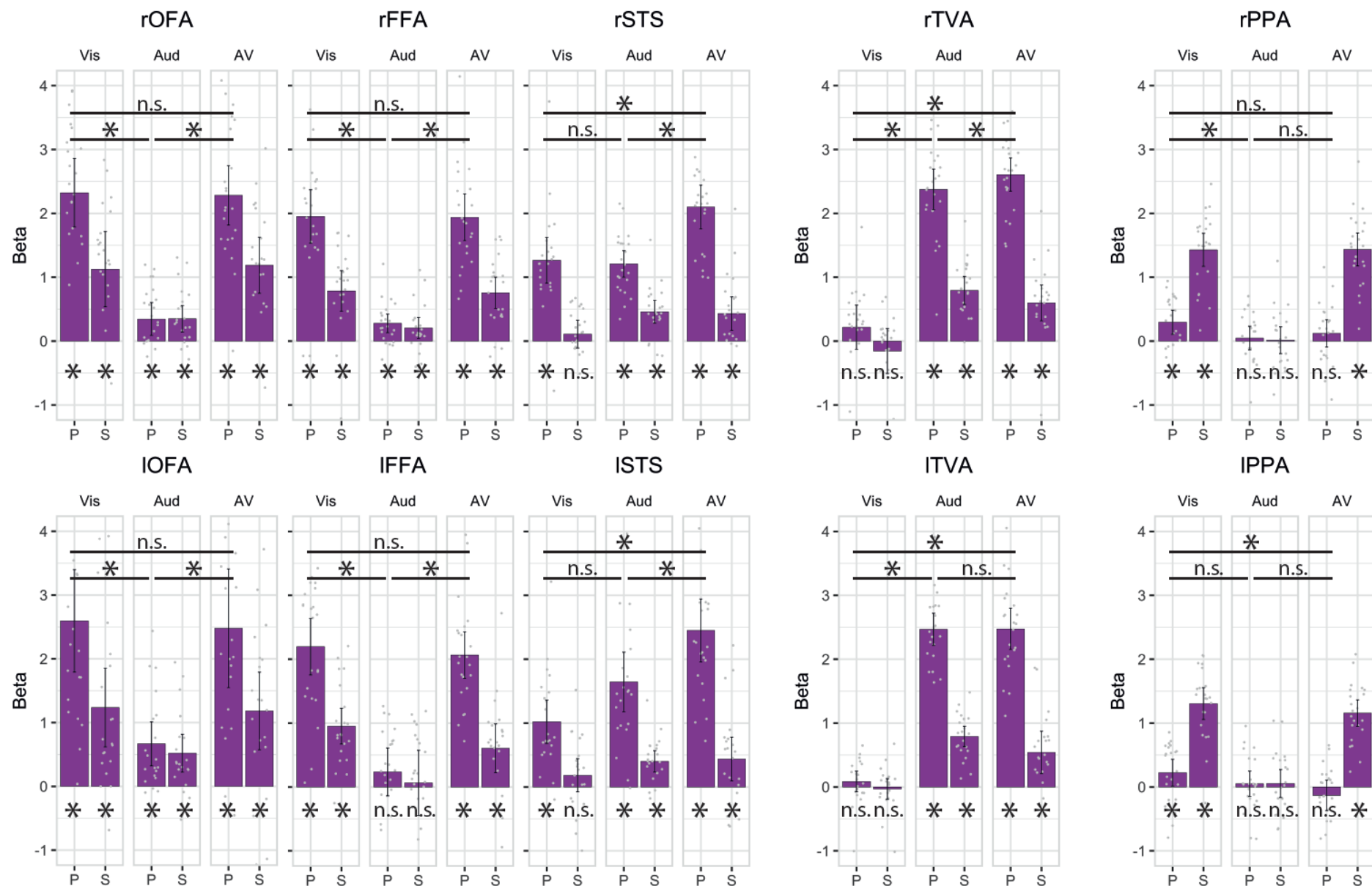


Результаты: средняя активация в зонах интереса



* t-тест: $p < 0.05$ с поправкой FDR (односторонний для сравнения с нулем; двусторонний парный для сравнения между модальностями)

Результаты: средняя активация в зонах интереса



* t-тест: $p < 0.05$ с поправкой FDR

Результаты: средняя активация в зонах интереса

Средняя активация на лица, голоса и аудиовизуальные стимулы людей значимо выше нуля во всех мультимодальных зонах интереса (исключение – rATL2, где нет активации на лица)

Отсутствует активация на зрительные сцены, однако есть активация на звуки (rpSTS, rpSTS2, lpSTS) и на аудиовизуальные сцены (rpSTS, rpSTS2) в областях, пересекающихся со слуховой корой

В ряде областей (rpSTS, rmSTS, lpSTS, rATL) активация на голоса выше, чем на лица: $A > V$. Активация на аудиовизуальные стимулы выше, чем на унимодальные: $AV > \max(A, V)$

В области rpSTS2: $AV > \max(A, V)$, но $A = V$

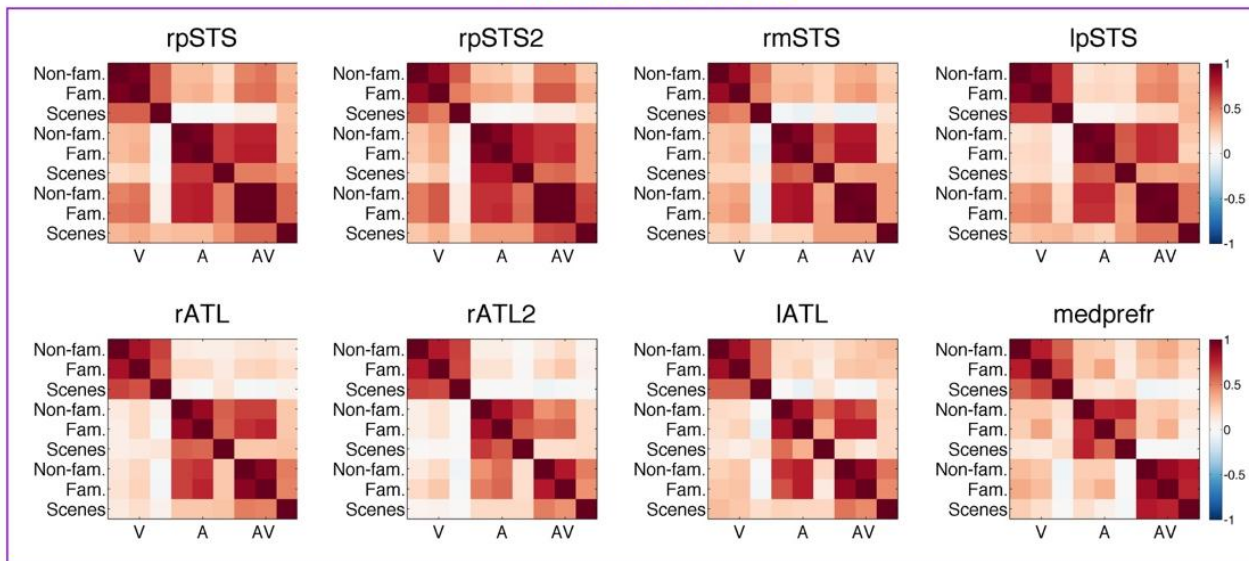
В областях rATL2 и lATL: $AV = A$; $AV > V$; $A > V$

В медиальной префронтальной коре различий между модальностями не выявлено: $AV = A = V$

Получена активация в унимодальных областях (избирательных к лицам, голосам и зрительным сценам) на релевантные категории стимулов

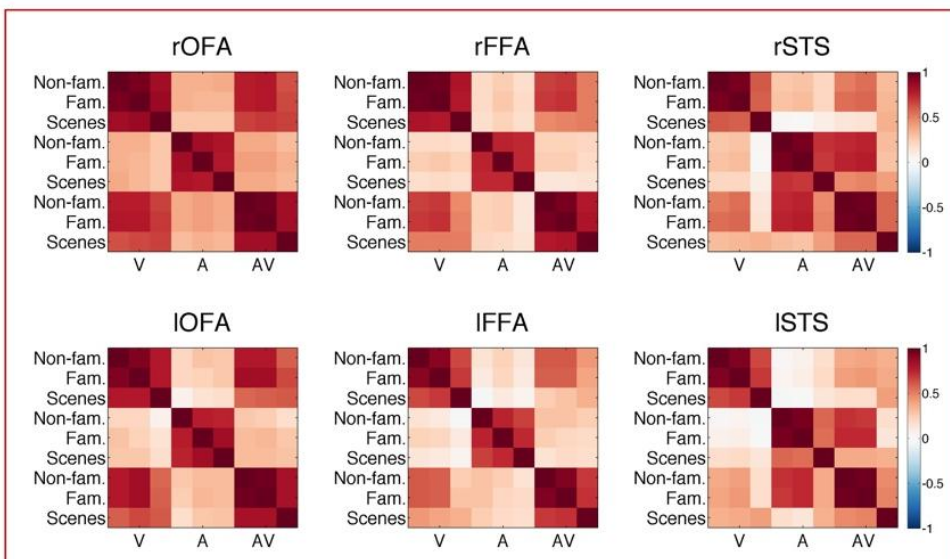
Результаты: паттерны активации в зонах интереса

Multimodal ROIs

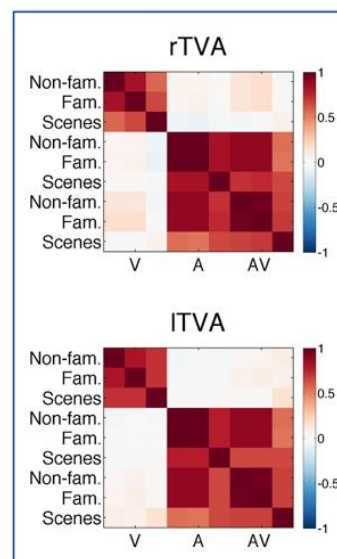


Попарные корреляции Пирсона между паттернами активации на незнакомые лица, известные лица и сцены в трех модальностях

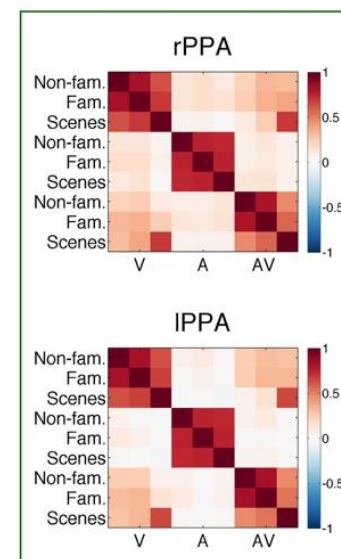
Face ROIs



Voice ROIs



Scene ROIs



Результаты: паттерны активации в зонах интереса

В большинстве мультимодальных зон интереса корреляции между паттернами активации значимо больше нуля

Величина корреляции между унимодальными репрезентациями низкая или умеренная (0.13-0.33); между мультимодальными и унимодальными – от низкой до высокой (0.18-0.82), что говорит о частичном сходстве паттернов

В большинстве мультимодальных зон корреляции между аудиальными и аудиовизуальными паттернами выше, чем между унимодальными, и выше, чем между зрительными и аудиовизуальными ($A/AV > A/V$; $A/AV > V/AV$)

В унимодальных зонах корреляции между аудиовизуальными паттернами и релевантными унимодальными значимо выше корреляций между двумя унимодальными паттернами

Большее сходство аудиовизуальных паттернов с аудиальными, чем со зрительными, в передней височной коре и верхней височной извилине может быть связано с тем, что данные зоны пересекаются со слуховой корой

Выделенные зоны, предположительно, содержат как мульти-, так и унимодальные нейроны

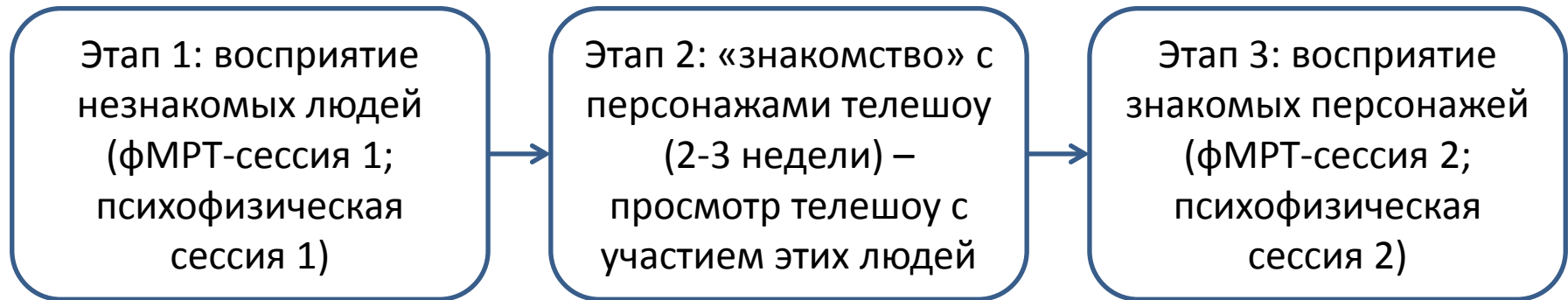
Исследование 1: выводы

Области мозга, участвующие в распознавании лиц и голосов

- Выявлены мультимодальные зоны, избирательные к лицам/голосам, самые крупные из которых расположены в правой височной коре (задняя часть верхней височной борозды, передняя височная кора)
- Результаты на независимой выборке подтвердили расположение этих зон
- Получена значимая активация на лица, голоса и аудиовизуальные стимулы в этих зонах (на независимых данных)
- В ряде зон активация на аудиовизуальные стимулы выше, чем активация на унимодальные – согласно одному из критериев ($AV > \max(A, V)$), это может свидетельствовать о мультимодальной интеграции информации
- В мультимодальных зонах паттерны активации на стимулы разных модальностей значимо коррелируют – вероятно, популяции нейронов, включенные в анализ зрительной, аудиальной и аудиовизуальной информации о человеке, частично пересекаются

Исследование 2

Влияет ли знакомство с людьми на сходство паттернов активации на лица и голоса?



Гипотеза 1: после «знакомства» с ранее неизвестными людьми паттерны активации на их лица и голоса будут более сходны между собой, чем паттерны активации до «знакомства»

Гипотеза 2: в унимодальных зонах паттерны активации на стимулы одной и той же модальности до и после «знакомства» будут более сходны, чем паттерны активации на разные модальности

Гипотеза 3: различия между паттернами активации после «знакомства» с персонажами в одной модальности можно предсказать по различиям в другой

Гипотеза 4: в унимодальных зонах возможно различие паттернов активации на разных «персонажей»

Метод исследования

Участники: N = 29; 18-27 лет, английский язык родной

T2*-взвешенная фМРТ (3×3×3 мм, TR = 2000 мс + 1500 мс для стимула)

T1-взвешенная структурная МРТ (1×1×1 мм)

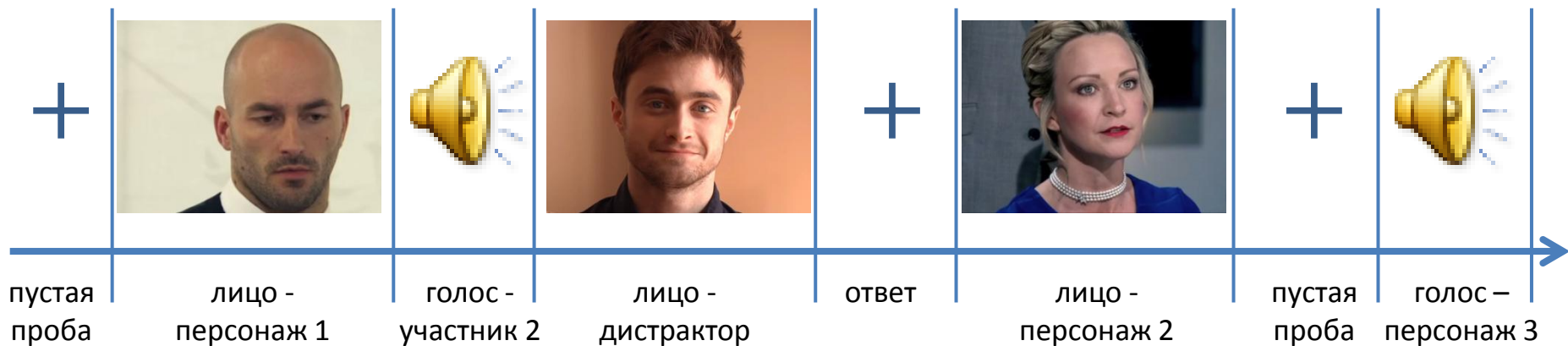
Стимулы: 8 персонажей телешоу (5 муж., 3 жен.), 3 видео- и 3 аудиозаписи каждого персонажа, по 4 повторения каждого стимула; 12 дистракторов – лица и голоса известных людей

Длительность каждого стимула или дистрактора – 1.2 с

Задача – нажимать на кнопку при появлении дистрактора

Три подхода по 15.5 минут в каждой из 2 сессий (до и после «знакомства»)

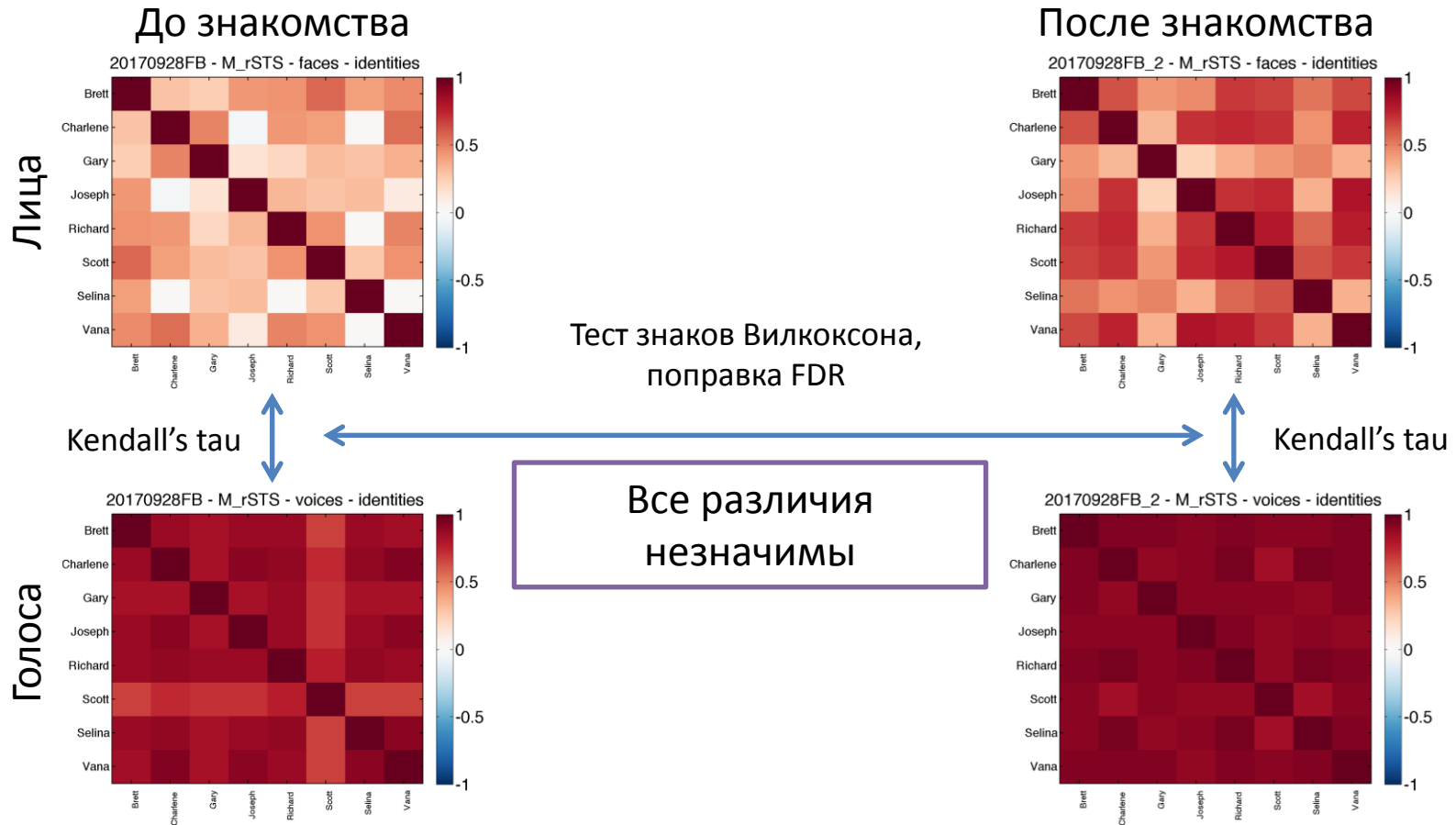
Зоны интереса – из исследования 1 (данные второй группы участников)



Меняются ли репрезентации лиц и голосов после «знакомства»?

Гипотеза 1: в уни- и мультимодальных зонах структуры репрезентаций лиц и голосов после «знакомства» более сходны, чем до «знакомства»

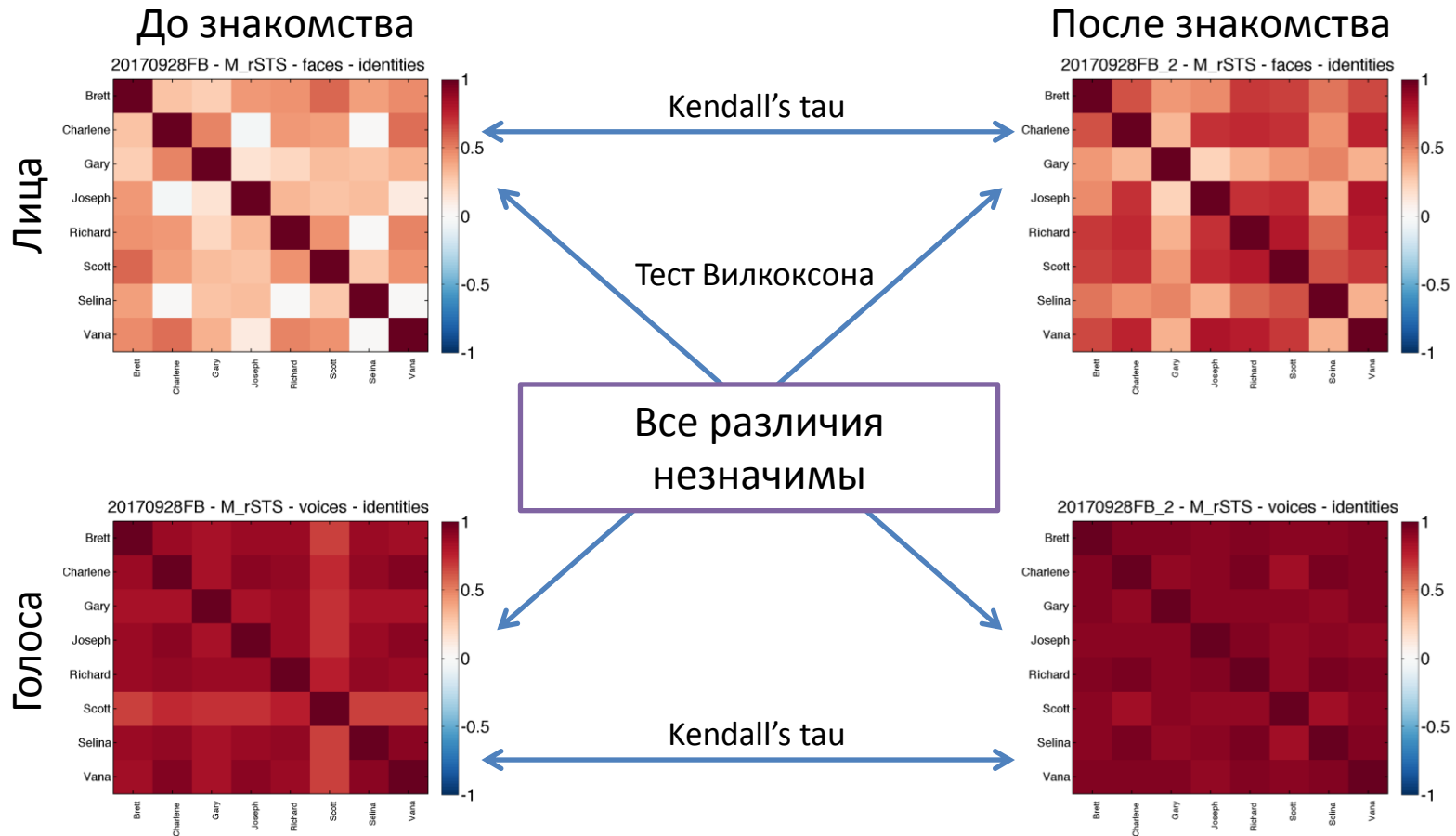
Матрицы корреляций Пирсона между паттернами активации на 8 персонажей



Насколько стабильны репрезентации в каждой модальности?

Гипотеза 2: в унимодальных зонах структуры репрезентаций стимулов в релевантной модальности более сходны, чем между модальностями

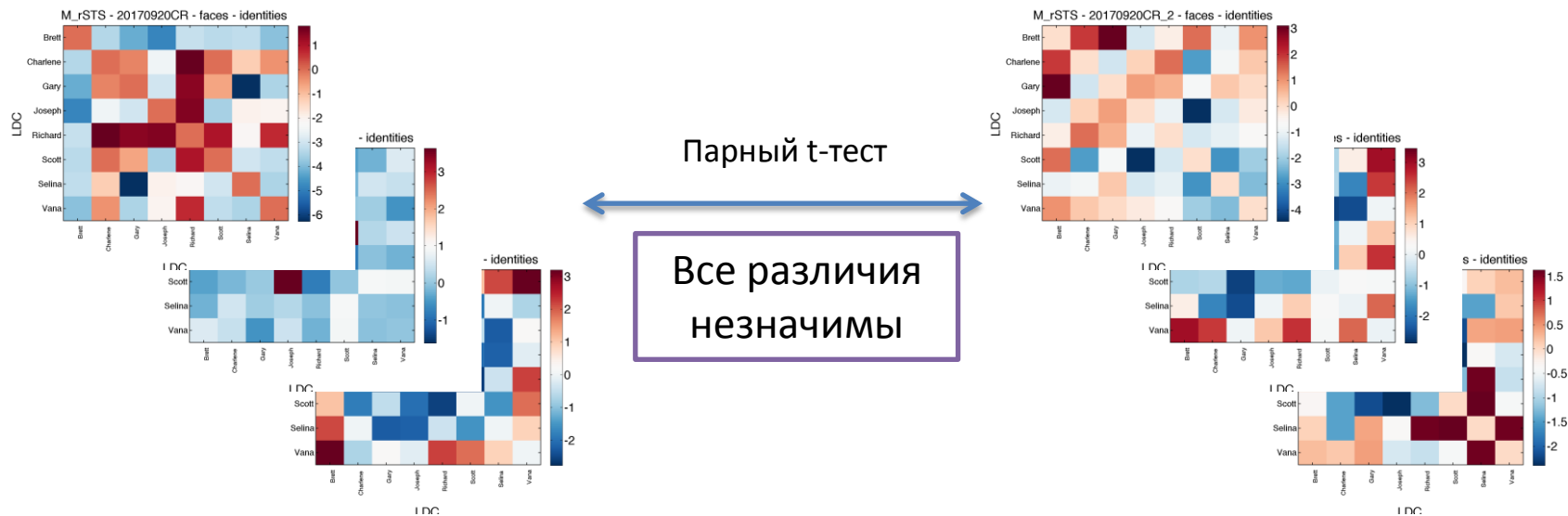
Матрицы корреляций Пирсона между паттернами активации на 8 персонажей



Можно ли предсказать структуру репрезентаций в одной модальности по данным другой?

Гипотеза 3: в уни- и мультимодальных зонах различия между паттернами активации после «знакомства» с персонажами в одной модальности можно предсказать по различиям в другой

- У каждого испытуемого данные каждой сессии делятся на 2 части (2 подхода и 1 подход)
- В каждой части выделяются мультивоксельные паттерны активации на лицо и на голос каждого персонажа
- Рассчитывается линейный дискриминант между каждой парой паттернов по данным 2 подходов в одной модальности и тестируется на данных 1 подхода в другой модальности (Nili, 2014; Walther et al., 2016)
- Анализ повторяется для всех комбинаций из трех подходов, для каждой сессии
- Результаты в каждой сессии усредняются между персонажами и модальностями для каждого испытуемого
- Парный t-тест между сессиями (до и после «знакомства»), поправка FDR



Можно ли предсказать структуру репрезентаций в одной сессии по данным другой?

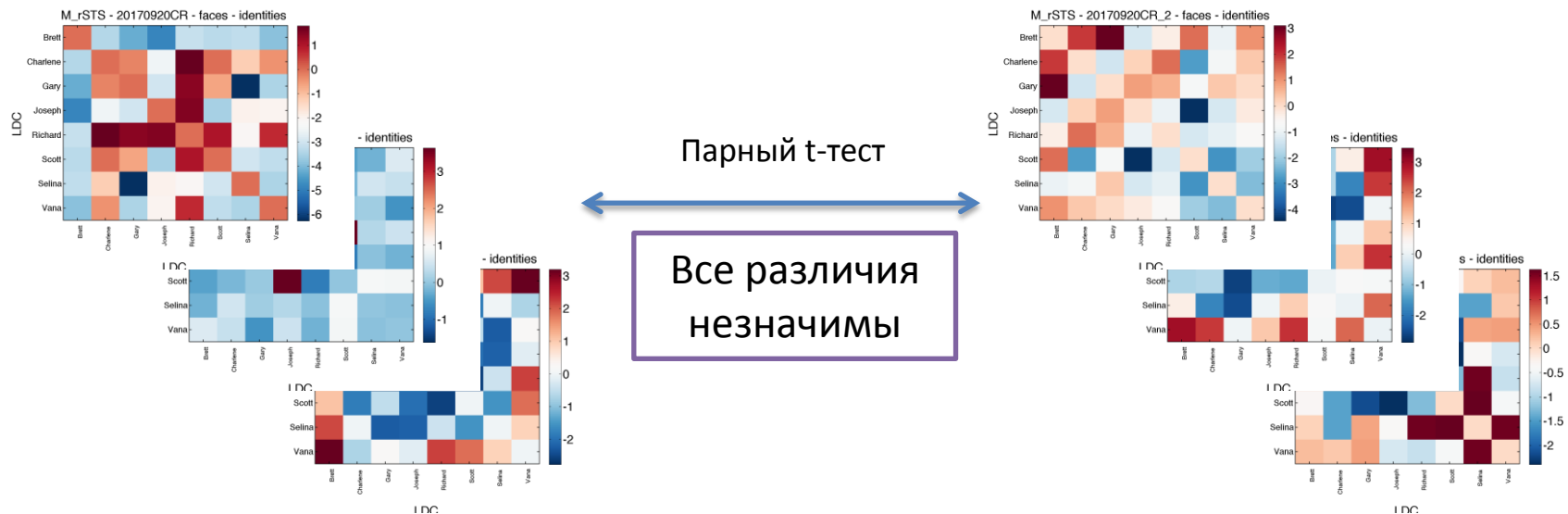
Гипотеза 4: в унимодальных зонах линейный дискриминант значимо выше нуля, и после «знакомства» он выше, чем до «знакомства»

- Линейный дискриминант рассчитывается между каждой парой паттернов по данным 2 подходов в релевантной модальности и тестируется на данных оставшегося подхода в этой же модальности

- Одновыборочный t-тест для каждой модальности, поправка FDR:

- На лица в rOFA, rSTS и ISTS дискриминант значимо больше нуля в обеих сессиях,
- В rFFA и IFFA – только в первой сессии;
- На голоса в rTVA и ITVA дискриминант значимо больше нуля в обеих сессиях.

- Парный t-тест между сессиями (до и после «знакомства»), поправка FDR



Исследование 2: выводы

- Не удалось выявить изменений структуры паттернов активации после «знакомства» с ранее неизвестными персонажами (паттерны активации не становятся более похожими)
- Структуры паттернов в одной и той же модальности не являются более похожими, чем в разных модальностях
- Не удалось предсказать структуру репрезентаций в одной модальности по данным другой
- В унимодальных зонах линейный дискриминант позволяет различать паттерны активации на лица либо голоса разных персонажей (в рамках одной сессии сканирования)
- После «знакомства» с персонажами точность различения не увеличивается

Исследование 3

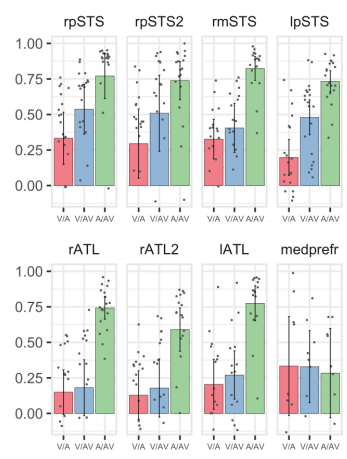
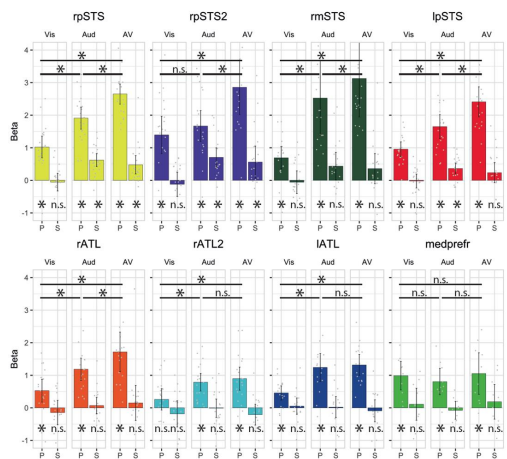
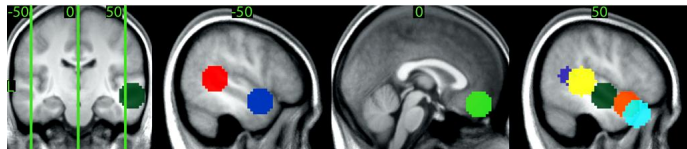
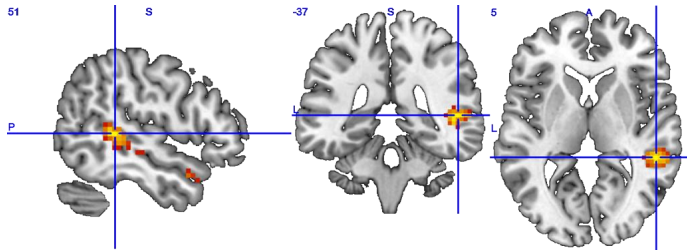
Психофизические данные

- Парное сравнение сходства всех видеозаписей по шкале от 1 до 7
- Парное сравнение сходства всех аудиозаписей по шкале от 1 до 7
- Оценки всех видео- и аудиозаписей по шкалам «симпатия», «надежность», «доминирование» и «привлекательность» (likeability, trustworthiness, dominance, attractiveness) от 1 до 7
- Узнавание участников шоу (выбор имени участника из списка)

Индивидуальные особенности восприятия лиц и голосов

- BVMT (Bangor Voice Matching Test) – различение людей по голосу
- CFMT (Cambridge Memory Test for Faces) – память на лица
- GVMТ (Glasgow Voice Memory Test) – память на голоса
- GFMT (Glasgow Face Matching Test) – различение лиц

... to be continued



Спасибо за внимание!

