**4 Оценка воспроизводимости показателей функциональной связности по данным когерентности ЭЭГ в условиях оперативного покоя**

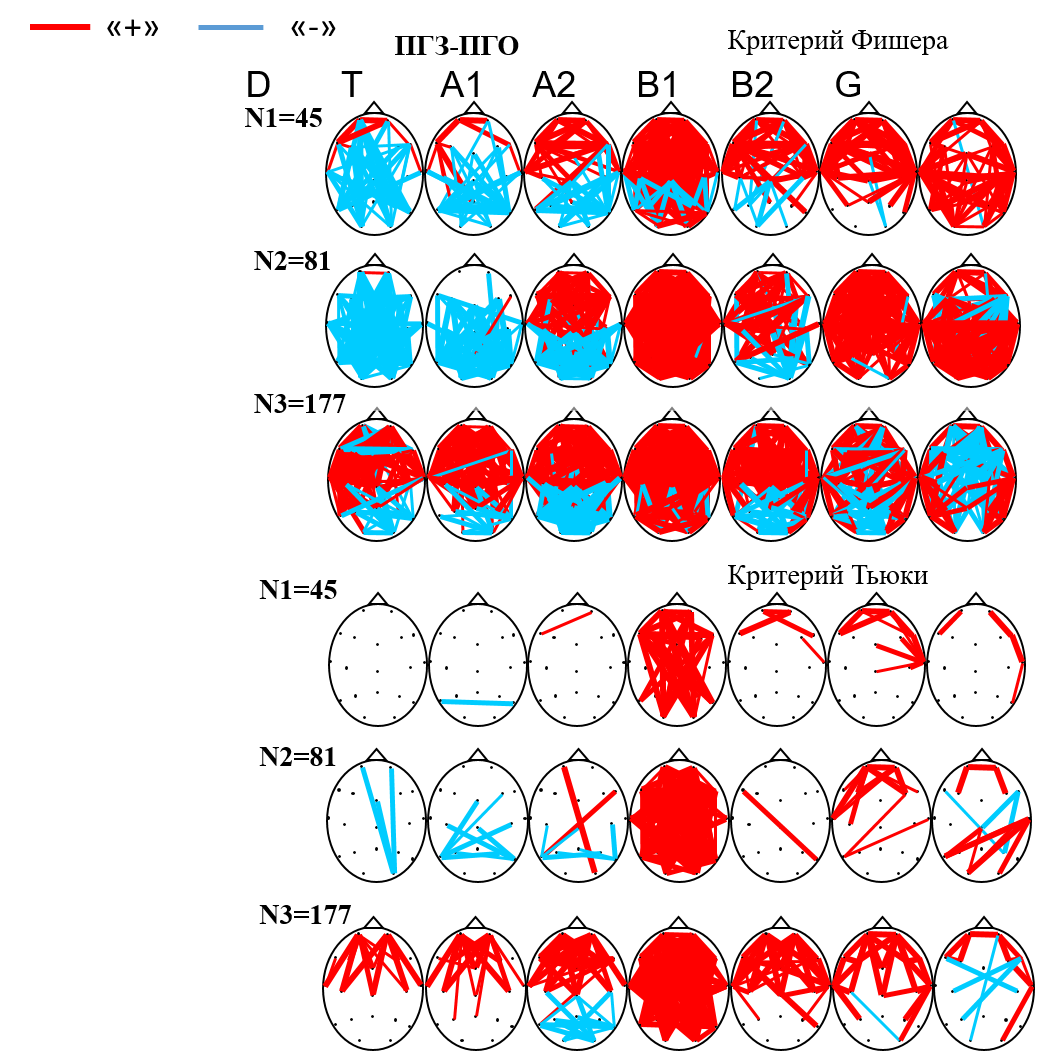
Текст для раздела отчета в раздел ВВЕДЕНИЕ

Ускоренный рост накопления данных нейровизуализационных исследований об активности мозга человека, помимо своего положительного влияния на современные возможности изучения его работы за счет большого количества данных, характеризуется все более явно проявляющимся кризисом воспроизводимости. Практически в любом исследовательском направлении, можно найти противоречащие данные, что существенно влияет на прогресс в исследовании функций мозга. Причем, это касается не только относительно «новых методов» функциональной томографии, но и ставших классическими и широко-применяемыми, электрофизиологических методов. В относительно недавний период времени, наметился тренд в сторону использования состояния оперативного покоя, которое рассматривается как с физиологической, так и технической точек зрения, как менее вариабельное состояние (по сравнению с активационными исследованиями), которое позволяет относительно более надежно оценивать характеристики системной активности мозга в покое. В соответствии с этим трендом, именно состояние оперативного покоя используется для визуализации звеньев и связей между ними для изучаемых функциональных мозговых систем. Данная информация и используется для построения функциональных коннектомов. Одним из широко-используемых методов для анализа функциональной связности является когерентость суммарной электрической активности мозга (ЭЭГ). Несмотря на длительную историю исследований функций мозга с помощью когерентности ЭЭГ, проблема воспроизводимости весьма значимо проявляется не только при изучении когнитивной деятельности, но в таких классических условиях как при закрывании глаз – так называемый эффект Бергера. Таким образом, основной задачей данного этапа исследований была проверки воспроизводимости результатов оценки когерентности ЭЭГ при помощи непараметрических тестов на примере реакции закрывания глаз.

**Вводная часть к разделу**

Анализ когерентности в ЭЭГ является одной из наиболее широко используемых и прямолинейных методов, наряду с корреляционными подходами, методами анализа динамики осциляторной активности мозга [1]. Несмотря на широкое использование, данный параметр суммарной электрической активности мозга, критикуем из-за чувствительности к явлению объемной проводимости, что приводит к снижению как пространственной, так и временной разрешающей способности [2]. Кроме того, отмечается и подверженность к ложно положительным находкам за счет смещенности оценок величины функциональной связности областей коры, вследствие существования объемной проводимости мозга (аналогично, говориться и о влиянии выбора реферанта на оценку когерентности) [3,4]. В связи с этим, следует отметить, что в настоящее время не только разработаны подходы по альтернативным методам расчета когерентности, но и внедрены методы коррекции/учета объемной проводимости, которые позволяют относительно эффективно учитывать влияние данных фактором. Вместе с тем, использование этих методов не решает проблему воспроизводимости рассчитываемых параметров функциональной связности. Кроме того, данная проблема усугубляется еще и тем, что, обычно, в опубликованных работах встречаются выборки не превышающие 40-50 испытуемых (тенденция к увеличению выборки появилась относительно недавно и вес их относительно общего количества работ пока мал). При этом, существенная вариабельность может наблюдаться и при относительно больших объемах выборок. Проблему вариабельности получаемых результатов, демонстрирующая кризис воспроизводимости, можно проиллюстрировать следующим образом на материале базы экспериментальных данных накопленных за последние десятилетия в ИМЧ РАН.

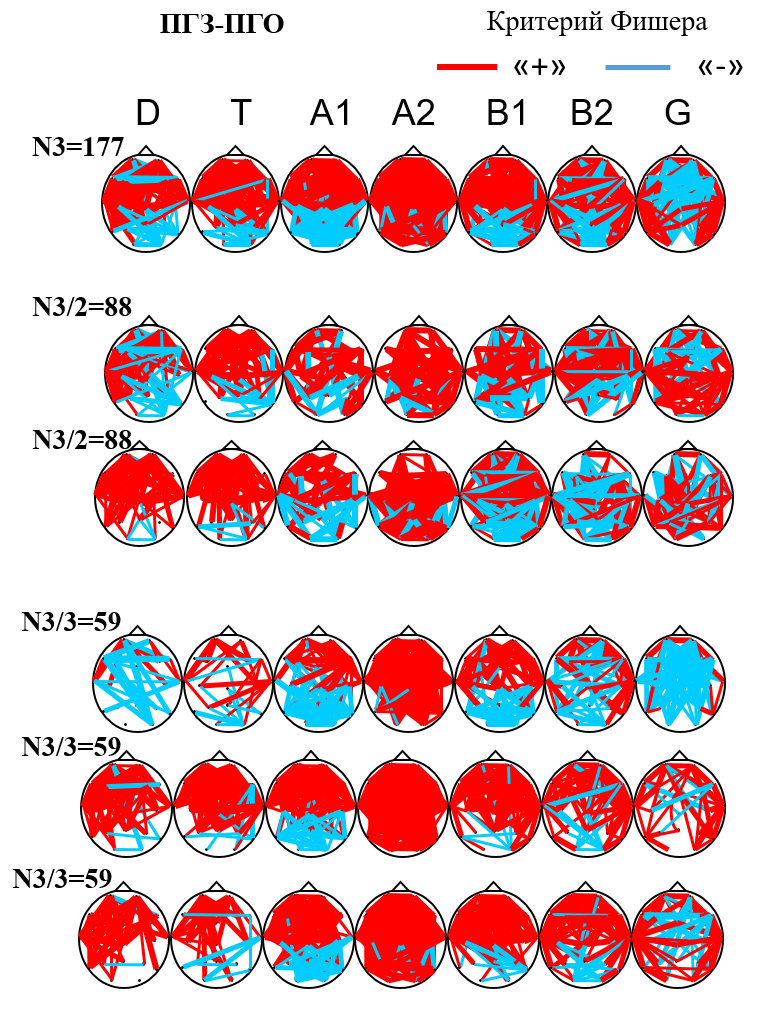
Так, воспроизводимость результатов оценки когерентности ЭЭГ может быть проиллюстрирована в сравнениях состояний закрытых и открытых глаз (ПГЗ-ПГО) для трёх независимых групп испытуемых разной численности (рисунок 4.1). При применении критерия Фишера во всех трёх группах воспроизводятся такие характеристики как доминирование увеличений когерентности в диапазоне альфа 2, а также преобладание увеличений в передних и проявление уменьшений в задних зонах кортекса в диапазонах альфа-1 и бета-1. В остальных диапазонах паттерны в группе 3 заметно отличаются от паттернов в группе 1 и группе 2. При применении критерия Тьюки в трёх группах воспроизводится только доминирование длинно-дистантных увеличений когерентности в диапазоне альфа-2. В других диапазонах паттерны в группах 1 и 2 заметно беднее и имеют мало общего как между собой, так и в сопоставлении с группой 3 наибольшей численности.



Примечания – ПГЗ (состояние покоя с закрытыми глазами), ПГО (состояние оперативного покоя с открытыми глазами), N (количество испытуемых в анализируемой выборке), «+» (обозначает большее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – увеличение когерентности при закрывании глаз), «-» (обозначает меньшее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – снижение когерентности при открывании глаз), D – дельта ритм, Т – тета ритм, А1 – альфа 1 ритм, А2 – альфа 2 ритм, В1 – бета-1 ритм, В2 – бета-2 ритм, G – гамма ритм

Рисунок 4.1. Когерентность ЭЭГ при закрывании глаз

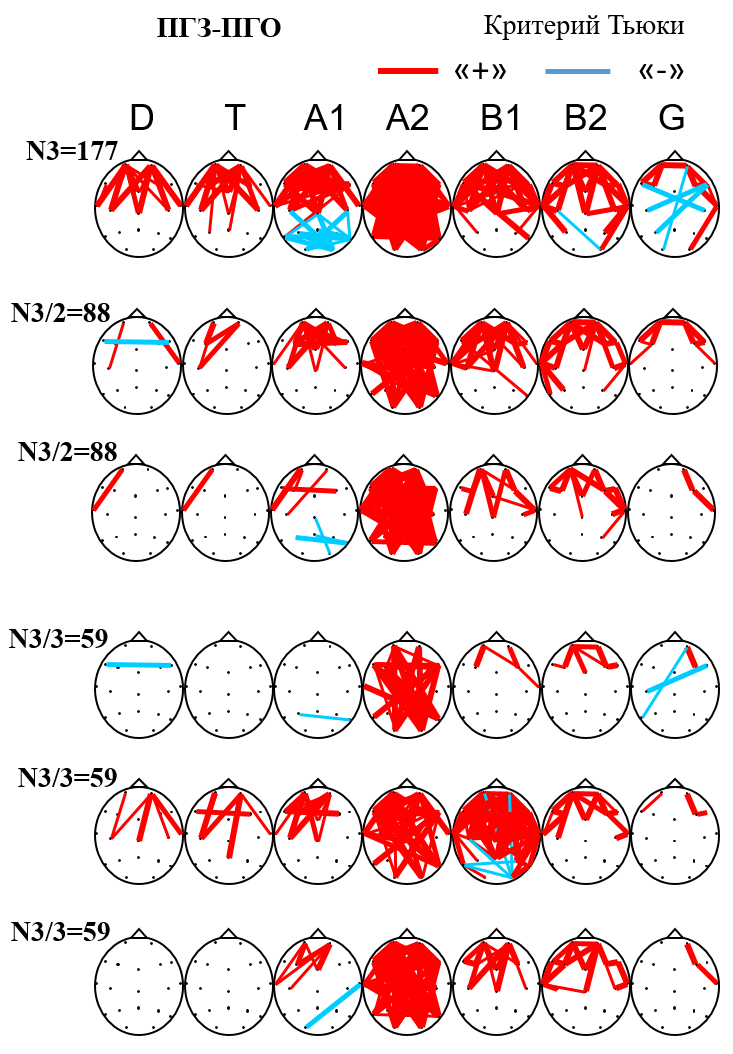
Воспроизводимость результатов оценки когерентности ЭЭГ может быть также проиллюстрирована в сравнениях ПГЗ>ПГО в подгруппах, получаемых посредством случайного деления на две или три подгруппы равной численности (рисунок 4.2). Уже в подгруппах численностью по 88 человек (обычно в одиночных психофизиологических ЭЭГ-исследованиях численность в разы меньше) заметна разница между результатами в подгруппах и различия с общей группой 177 человек даже в диапазоне альфа 2 с наиболее сильным эффектом. Ещё более выражены различия такого рода в подгруппах меньшей численности: например, паттерны в диапазонах дельта и гамма для подгрупп 1 и 2 численностью 59 человек (рисунок 4.2.).



Примечания – ПГЗ (состояние покоя с закрытыми глазами), ПГО (состояние оперативного покоя с открытыми глазами), N (количество испытуемых в анализируемой выборке), «+» (обозначает большее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – увеличение когерентности при закрывании глаз), «-» (обозначает меньшее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – снижение когерентности при открывании глаз), D – дельта ритм, Т – тета ритм, А1 – альфа 1 ритм, А2 – альфа 2 ритм, В1 – бета-1 ритм, В2 – бета-2 ритм, G – гамма ритм

Рисунок 4.2 Когерентность ЭЭГ при закрывании глаз при разных вариантах случайного разбиения группы из 177 испытуемых на подгруппы.

Использование более строгих критериев статистической оценки с использованием критерия Тьюки (рисунок 4.3) и в случае деления на подгруппы меньшей численности можно говорить о воспроизводимости хотя бы общего характера различий только в диапазоне альфа 2 (хотя частные несовпадения в отдельных парах отведений можно увидеть и в этом диапазоне). Если ориентироваться на результаты в группе 177 человек, как наиболее достоверные из доступных и считать их условно «истинными», то результаты в подгруппах в большинстве диапазонов дают только фрагментарные «намёки» вплоть до полного отсутствия результатов в том или ином диапазоне в подгруппах по 59 человек (рисунок 4.3).



Примечания – Примечания – ПГЗ (состояние покоя с закрытыми глазами), ПГО (состояние оперативного покоя с открытыми глазами), N (количество испытуемых в анализируемой выборке), «+» (обозначает большее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – увеличение когерентности при закрывании глаз), «-» (обозначает меньшее значение когерентности ЭЭГ в состоянии ПГЗ по сравнению с ПГО – снижение когерентности при открывании глаз), D – дельта ритм, Т – тета ритм, А1 – альфа 1 ритм, А2 – альфа 2 ритм, В1 – бета-1 ритм, В2 – бета-2 ритм, G – гамма ритм

Рисунок 4.3 Изменения когерентности ЭЭГ при закрывании глаз выявленные при более строгом критреии статистического оценивания Тьюки

Таким образом, учитывая вышесказанное на данном этапе исследований планировалось проверить воспроизводимость результатов оценки когерентности ЭЭГ при помощи методов непараметрической статистики

**4.2 Методика исследования**

КАК БЫЛИ ПОЛУЧЕНЫ ДАННЫЕ И СЧИТАЛАСЬ КОГЕРЕНТНОСТЬ

КАК ОТБИРАЛИСЬ ОТРЕЗКИ ЭЭГ и т.д.

Для решения задачи оценки воспроизводимости результатов и степени достоверности полученных оценок были проведены многочисленные симуляционные эксперименты основанные на совокупности непараметрических тестов. В частности ставились следующие вопросы :

- насколько устойчиво воспроизводится эффект в подгруппах фиксированного размера.

- есть ли связь между найденным уровнем значимости и частотой воспроизводимости эффекта.

- чем руководствоваться при выборе порогов для уровней значимости и как корректировать на множественные сравнения

Для ответов на эти вопросы был реализован набор непараметрических тестов.

**Анализ межканальных различий с помощью теста на перемешивание**

Первым шагом является индивидуальный поиск межканальных различий. Обычно это делается с помощью стандартого t-test а, но нами была использована его непараметрическая альтернатива. Результаты, полученные с помощью обоих подходов очень близки, но в рамках непараметрического подхода не накладывается никаких ограничений на распределение изучаемых данных, что играет особенно большую роль в случае малых выборок.

При использовании перестановок (пермутация), основываясь на имеющихся экспериментальных данных делается предположение о том, как бы выглядело распределение тестовой статистики, если бы нулевая гипотеза была верна (Cohen 2014). Более подробно, используемая в данном случае процедура тестирования описана в статье (Maris, Schoffelen, and Fries 2007). В общем виде нулевую гипотезу теста на перестановки можно сформулировать следующим образом – при перемешивании двух состояний А и В не будет значительной разницы в тестовой статистике до перемешивания и после. Цель такого тестирования с помощью перемешиваний – определить вероятность того, что наблюдаемая тестовая статистика могла бы быть получена при верной нулевой гипотезе. Альтернативная гипотеза: наблюдаемая тестовая статистика не случайна, обнаружена не из-за случайных выбросов (outliers), а эффект действительно присутствует – т.е. не случаен и статистически значим.

Чтобы получить распределение статистической величины при верной нулевой гипотезе, мы должны повторить процедуру перемешивания многократно. Далее полученная эмпирическая тестовая статистика сравнивается со статистикой, полученной при перемешивании. Если эмпирическая тестовая статистика располагается внутри распределения при верной нулевой гипотезе, то нулевая гипотеза не может быть отвергнута. Это может значить, что реальной разницы между срединми значениями сравниваемых переменных нет, а метки испытаний А и В могли быть получены случайным образом.

Если же эмпирическая статистика достаточно далека от распределения при верной нулевой гипотезе, нулевая гипотеза отвергается, а эффект можно считать статистически значимым.

В случае данного исследования, для каждого ритма осциляторной активности мозга и для каждой пары каналов ЭЭГ считается разница между значениями когерентности в состояни оперативного покоя с открытыми глазами и состоянием оперативного покоя с закрытыми глазами. Для всей совокупности этих данных значения для каждого испытуемого случайным образом осуществляется выбор между двумя типами состояний состояние: закрытые глаза – открытые глаза. Далее для таких «назначенных» состояний рассчитывается разница между параметрами когерентности. Данная процедруа осуществляется для каждой из анализируемых пар из 16 каналов ЭЭГ когерентности . На следующем этапе рассчитывались средние значения когерентности для таких пар каналов. Если, в действительности, эффекта разницы в параметрах когерентности ЭЭГ между сравниваемыми состояниями нет, то при вышеописанном перемешивании («назначении» состояний случайным образом, вне зависимости от того, какое именно было состояние – открытые или закрытые глаза), то разница между эмпирическим средним значением переменной и средним значением переменной полученным с помощью таких перемешиваний не будет значима (будет отсутствовать) .

Далее, повторяя такую процедуру многократно, осуществляется расчет p-value, которое вычислялось как число раз при котором значения статистики при нулевой гипотезе (среднее после перемешивания) получилось более экстремальными, то есть большими по абсолютной величине, чем значение эмпирической статистики, деленное на количество итераций перемешивания. Чем больше таких итераций, тем точнее значение p-value. Далее для каждого ритма для любой подвыборки можно получить карту различий в уровнях межканальных синхронизаций.

Следующий вопрос на который мы отвечали - это устойчивость полученных различий для подгруппы фиксированного размера. Для этого использовалась техника бутстрепа, а именно из исходной большой группы размером 177 человек случайным образом выбирались подвыборки фиксированного размера: 30, 40, 60, 120, 140, 160 и для каждой подвыборки проводился анализ межканальных различий описанных выше. Процедура повторялась многократно (100 раз), после чего появилась возможность анализа результатов для каждой подгруппы. Мы далее вычисляли такую характеристику как частоту воспроизведения конкретного различия при заданном уровне значимости. В качестве уровеня значимости мы взяли уровень, который обычно используют для анализа индивидуальных различий 0.05 и более строгий уровень с учетом множественных сравнений равный 0.001. Для обоих уровней для подгруп разного размера были построены так называемые карты возспроизводимости различий. Похожий подход, но для оценки воспроизводимости был использован для анализа фМРТ данных (Cremers, Wager, and Yarkoni 2017)

Последний эксперимент, который был проведен - это анализ связи уровня значимости эффекта и его воспроизводимости. Очень часто более высокий уровень значимости (что соответствует более низкому p-value) воспринимается как более надежное. Для проверки этого эффекта для данной задачи мы провели следующий эксперимент: для подгруппы фиксированного размера оценили индивидуальные различия (с помощью непараметрического теста описанного в начале главы), а затем отсортировали их по убыванию значимости и далее с помощью методов повторных подвыборок провели для каждой из них анализ индивидуальных различий и построили распределение p-value для каждой пары каналов. В результате такого эксперимента можно было проверить является ли более высокий уровень значимости свидетельством в пользу того что этот эффект будет более значимым и при повторных экспериментах.

Пошагово процедура анализа заключается в следующем:

* Для фиксированной подвыборки делается индивидуальный анализ связей и для каждой связи вычисляется p\_value. Далее, для удобства вводится параметр значимости (1-p\_value) и все ребра сортируются по убыванию этого параметра. В этом случае максимальное значение равно 1 (наиболее достоверная разница), а минимальное 0 (нет различий).
* Далее делается N повторных экспериментов для разных подвыборок, в которые не входят испытуемые из первого эксперимента и для каждого ребра вычисляются N значений p\_value.
* Полученного набор значений 1-p\_val представляет собой распределение значимостей, для которых можно оценить квантили и представить любым удобным способом, например с помощью бокс плотов.

**4.3 Результаты исследований**

**Непараметрический поиск межканальных различий.**

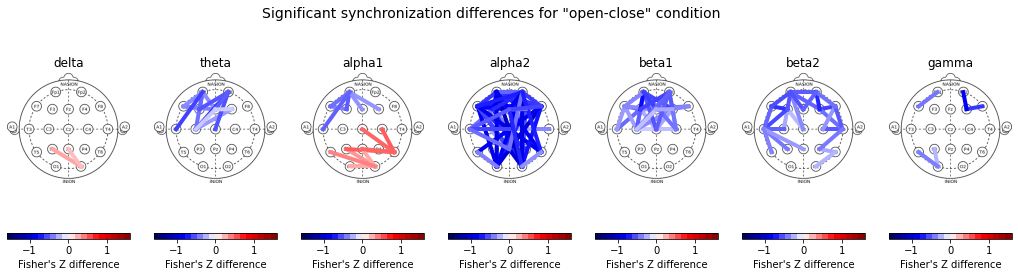
Для каждого ритма для каждой пары каналов был проведен непараметрический тест поиска различий с помощью перемешивания, описанный выше. На Рис 4.4 показано распределение разниц между состояниями для перемешанных данных, то есть при верности нулевой гипотезы об отсутствии разница и наблюдаемая статистика для ситуации когда разница отсутствует и наблюдается. В результате 10000 итераций процедуры поиска различий с помощью перемешивания состояний описанной выше были расчитаны распределения. Например, когерентности ЭЭГ в дельта диапазоне для пары отведений Cz/O2 (Рис 4.4 слева) наблюдаемое среднее (красная линия) лежит за пределами распределения, a p-value < 0.001, то есть полученное значение для этой пары не является случайным. Таким образом канал считается значимым. справа

|  |  |
| --- | --- |
|  |  |

Рисунок 4.4 Распределение разницы состояний открытые и закрытые глаза при проведении процедуры перемешивания. Красной линией изображена вероятность, соответствующая разности вычисленной по реальным данным. Черными - значимости на общеиспользуемых уровнях 2.5%, 97.5% и 95%

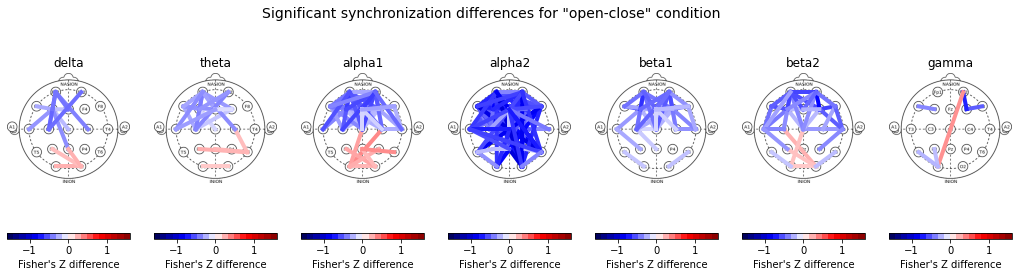
**Карты распределений значимых различий для подгрупп разных размеров.**

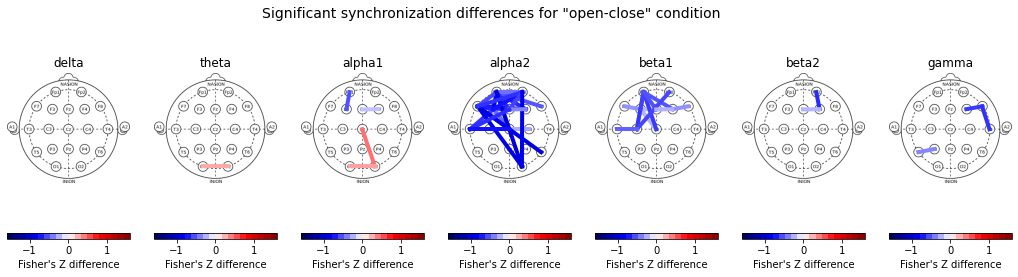
Также, для сравнения с предыдущими результатами, были получены примеры карт распределений для подгрупп с разными размерами выборки. В случае с подвыборкой большого размера - результат получился очень похожий, однако на малых подвыборках результат оказался нестабильным



N=30

N=60





N=100

N=177

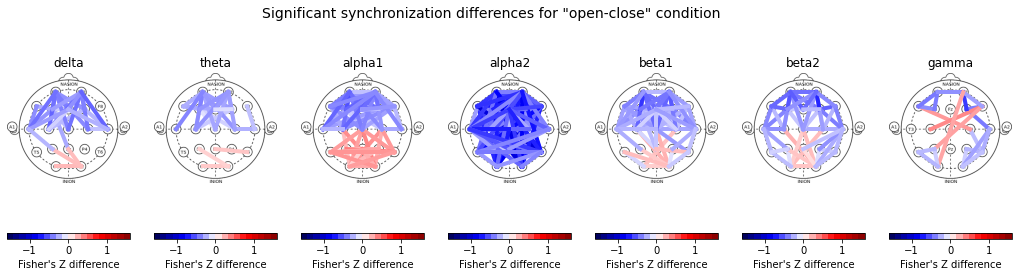
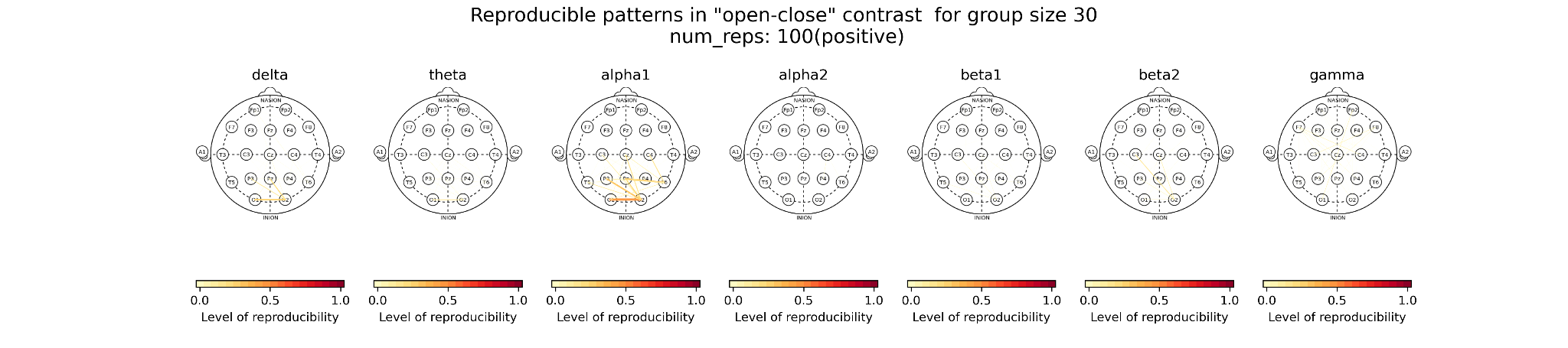


Рис 4.5 Карты различий парных синхронизаций состояний с открытыми и закрытыми глазами для разных размеров подвыборок

**Оценка частоты воспроизводимости эффекта**

В результате многократных повтора экспериментов на различных подвыборках одного и того же размера и вычисления различия между состояниями было обнаружено, что карты различий получаются каждый раз разные. Для того, чтобы количественно оценить разницы был проведена оценка частоты воспроизводимости для каждого канала. Процедура состояла из нескольких шагов: случайный выбор подвыборки размера N, оценка индивидуальных различий, фиксация пар каналов, которые значимы на заранее выбранном уровне (0.05 и 0.001), повтор процедуры 100 раз. В результате для каждой пары каналов для каждой итерации получаем бинарное значение (есть значимое отличие или нет), после этого можно посчитать долю значимых синхронизации для каждой пары каналов. Данную величину можно интерпретировать как воспроизводимость эффекта. Результат будет зависеть от заранее выбранного уровня значимости. На Рис 4.6 представлены результаты для разных размеров подвыборок и разных уровней значимостей. Для удобства восприятия положительные и отрицательные эффекты изображены на разных графиках.



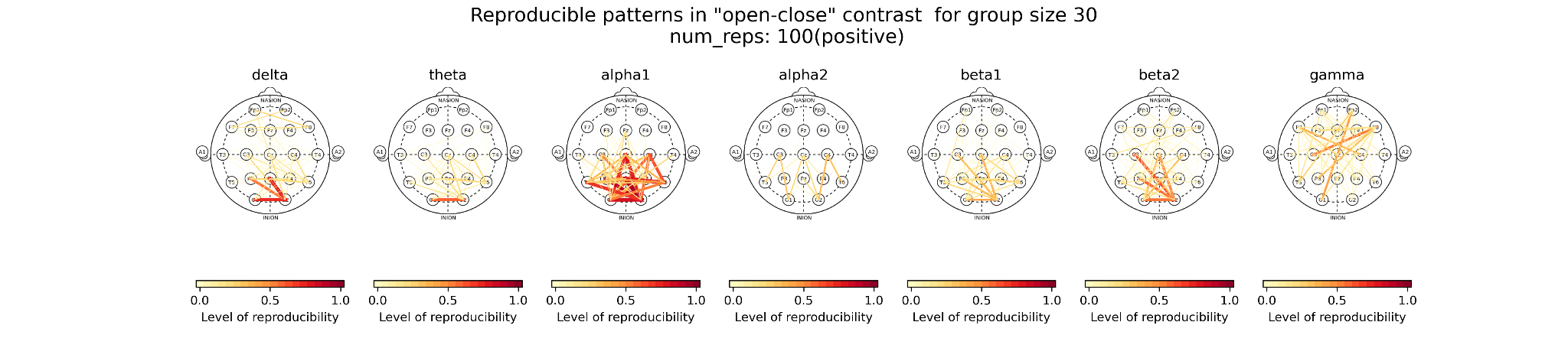
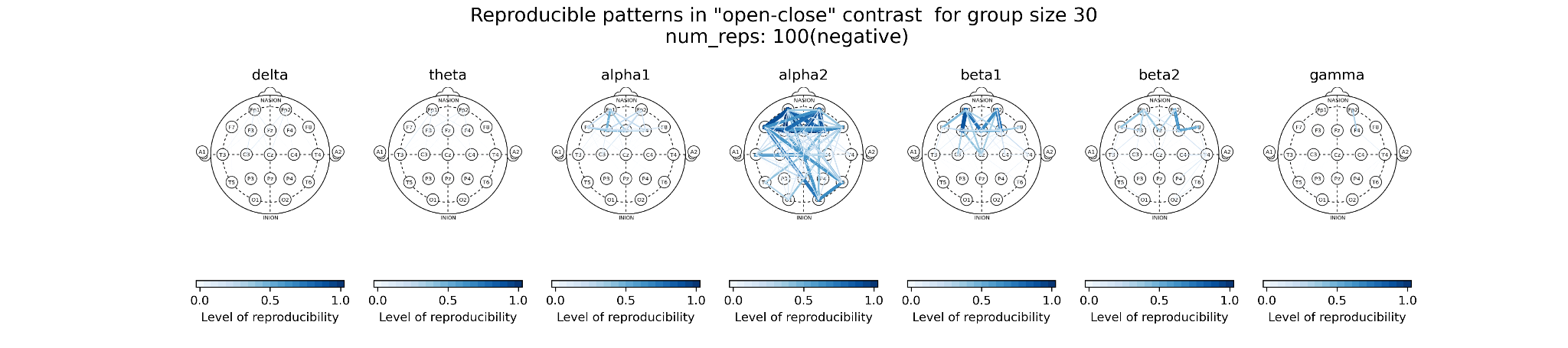


Рис 4.6 (а) Частота воспроизведения эффекта при заданном пороге и размере подгруппы

alpha = 0.001(вверху), alpha = 0.05 (внизу), положительный эффект



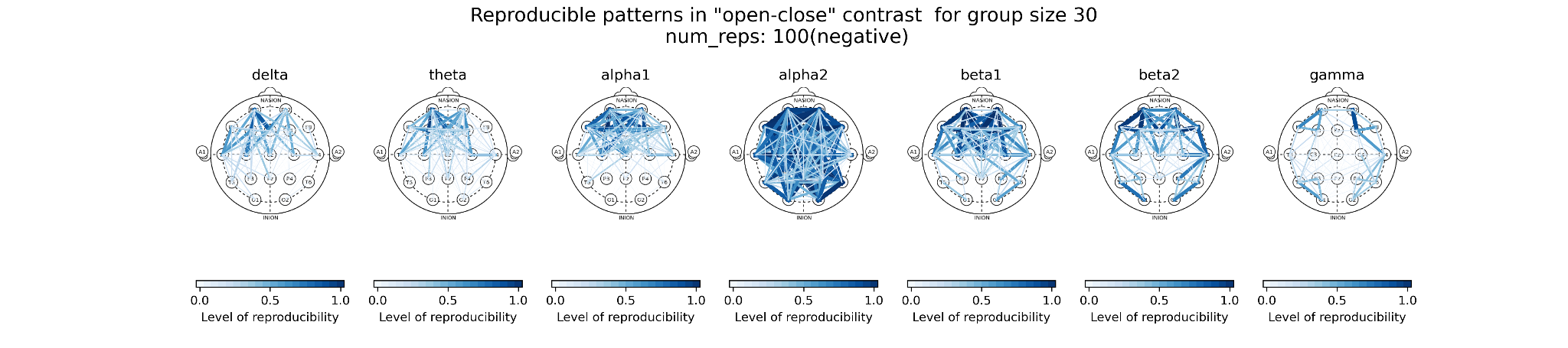


Рис 4.6 (б) Частота воспроизведения эффекта при заданном пороге и размере подгруппы 30, alpha = 0.001(вверху), alpha = 0.05 (внизу), отрицательный эффект

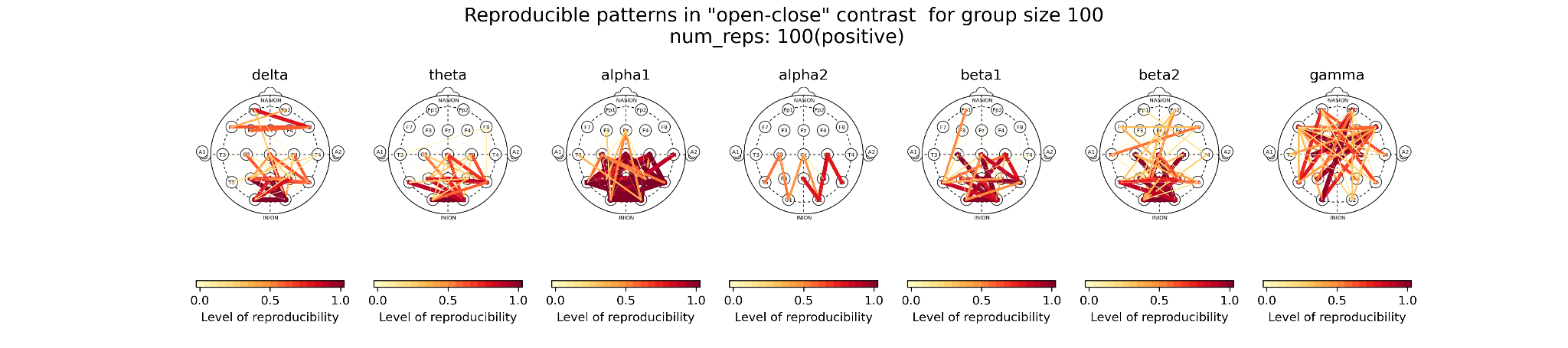
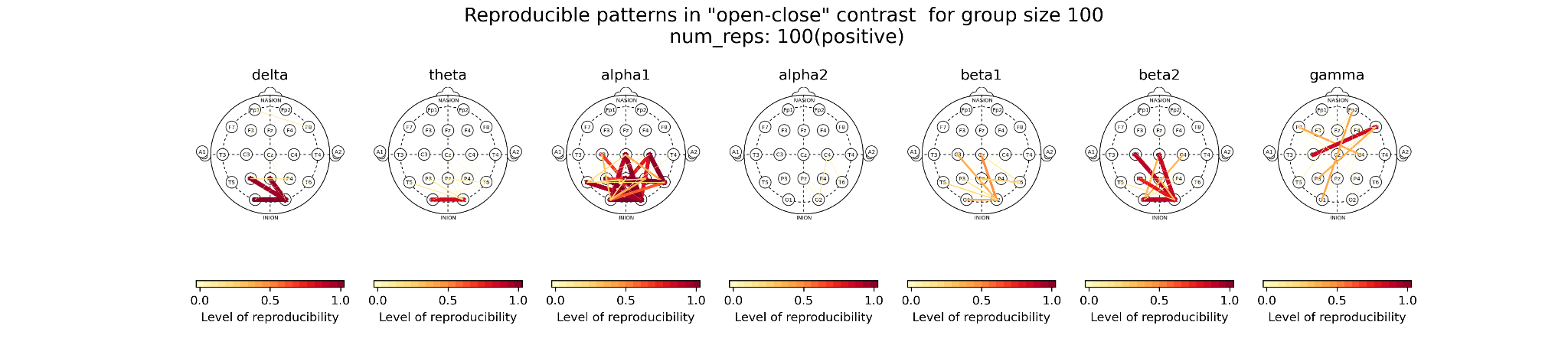


Рис 4.6 (с) Частота воспроизведения эффекта при заданном пороге и размере подгруппы 100, alpha = 0.001(вверху), alpha = 0.05 (внизу), положительный эффект

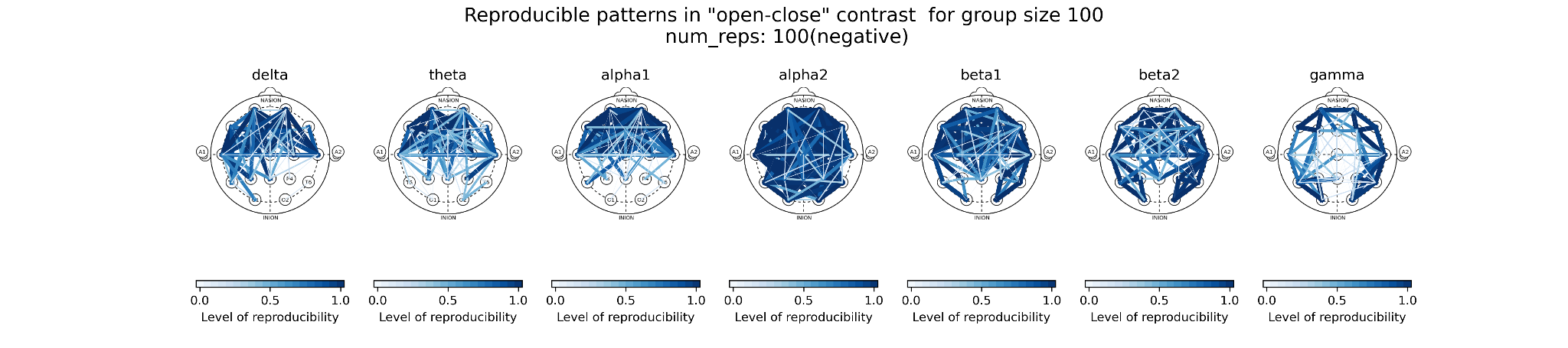
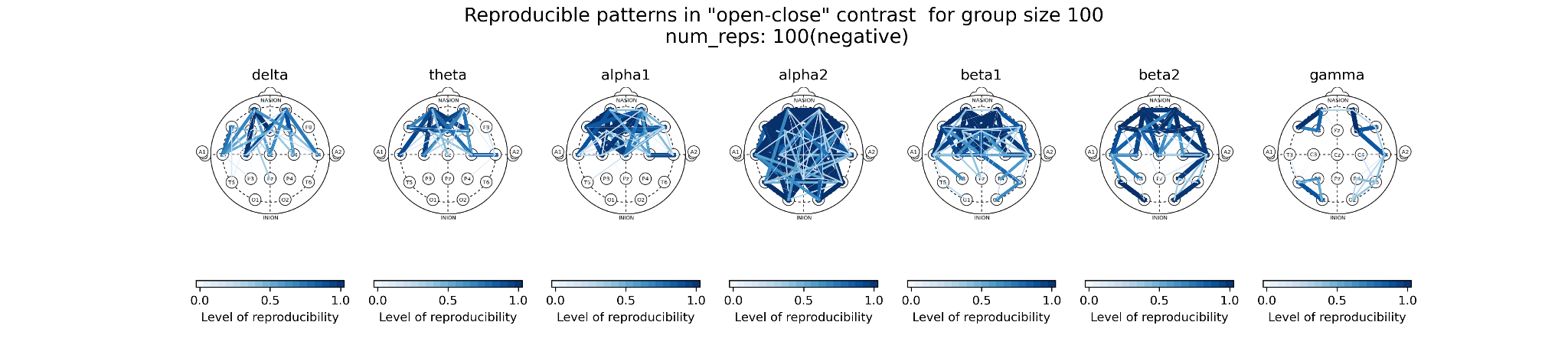


Рис 4.6 (д) Частота воспроизведения эффекта при заданном пороге и размере подгруппы 100, alpha = 0.001(вверху), alpha = 0.05 (внизу), отрицательный эффект

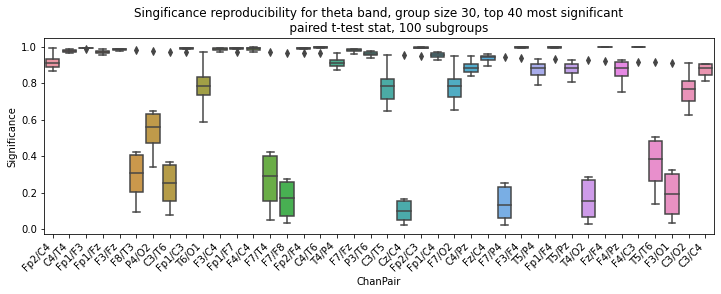
Таблицы со значимыми каналами:

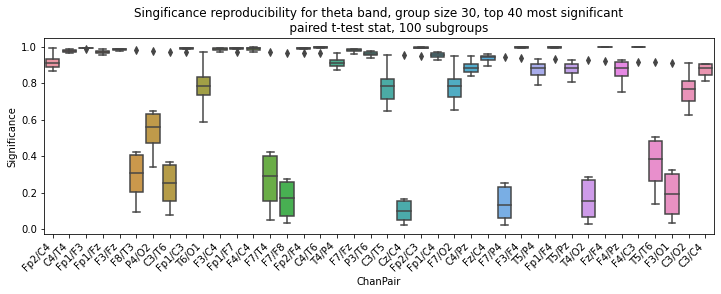
<https://drive.google.com/drive/folders/1YXL1rRJvJslXW8CoilG1xrP-A4MaxJQF?usp=sharing>

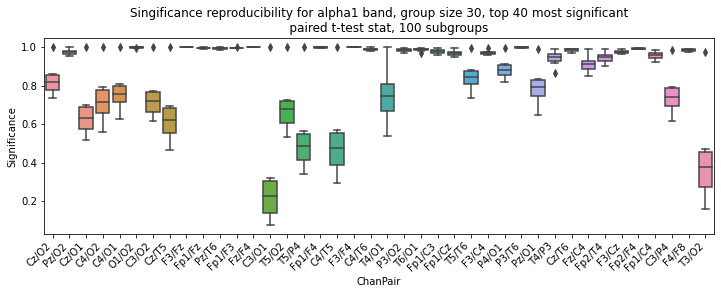
**Результаты оценки воспроизводимости уровней значимости**

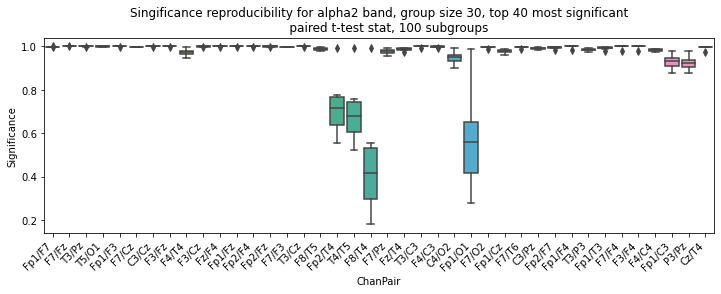
Для каждого ритма результаты были получены согласно процедуре описанной в части методы. Оказалось, что для самых устойчивых паттернов, особенно в альфа диапазоне, данный эффект действительно воспроизводится, однако для более слабых эффектов картина может разнится существенно. На Рис 4.7 показаны распределения значимости для различных размеров подвыборок и разных диапазонов. Каналы упорядочены в соответствии с убыванием значимости в первой зафиксированной подвыборке.

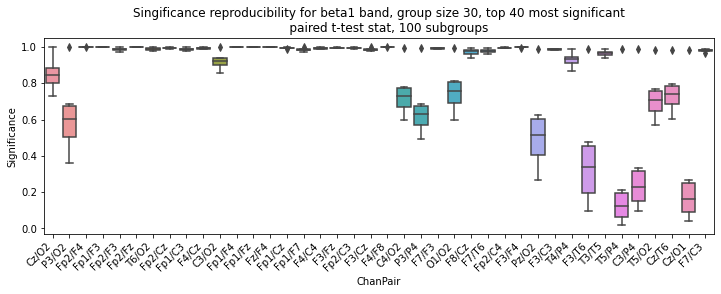


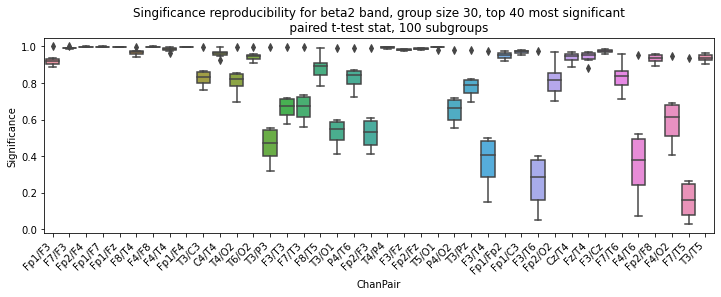


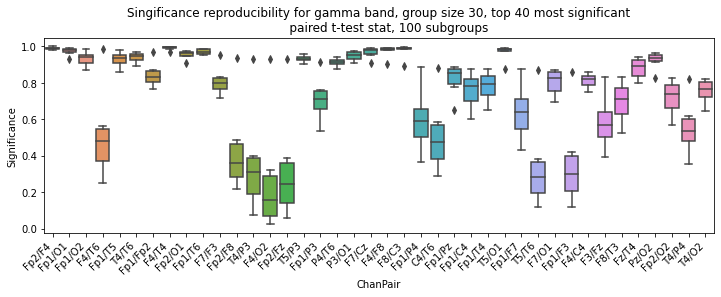




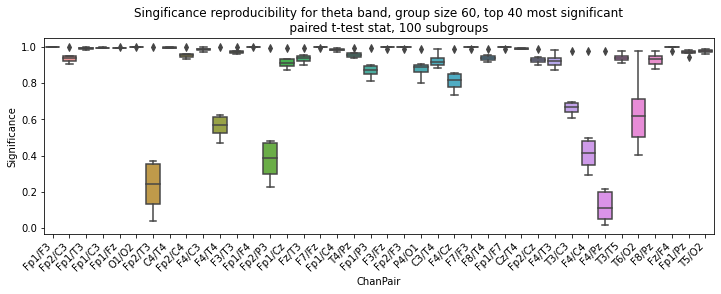


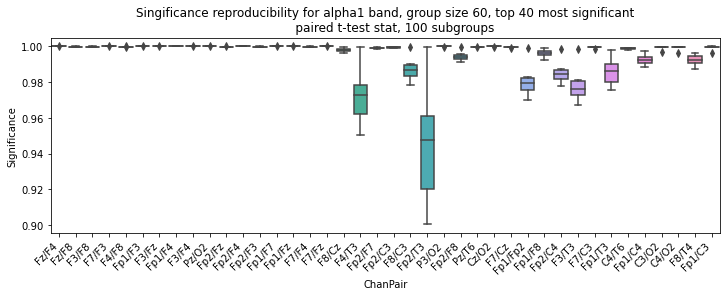


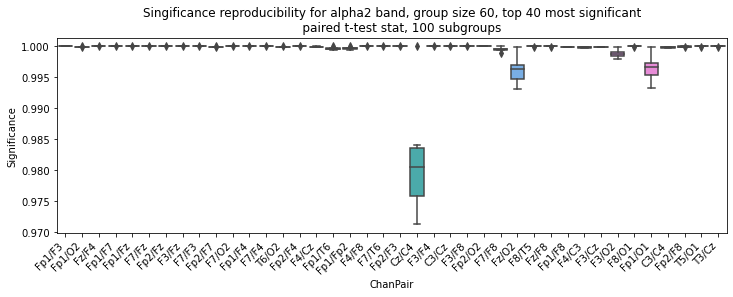


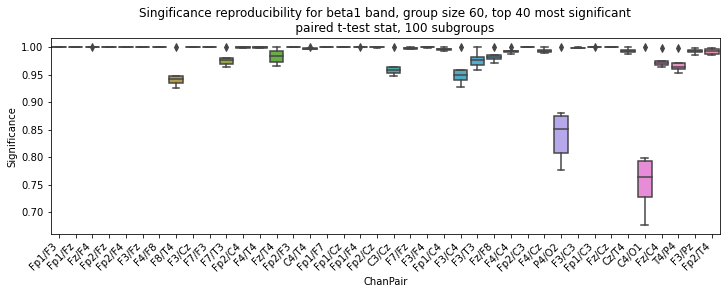


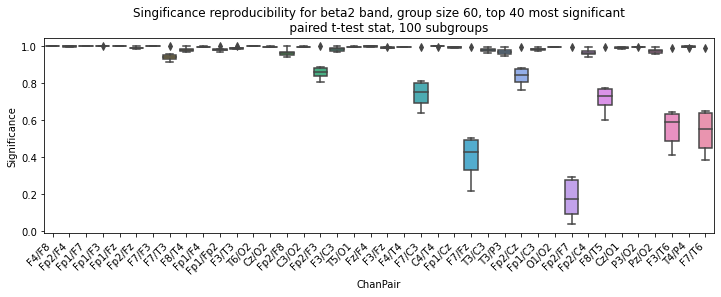












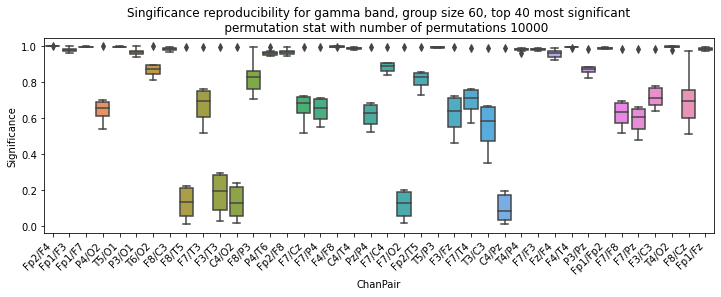


Рис 4.7 Распределения значимости для различных размеров подвыборок и разных диапазонов. Каналы упорядочены в соответствии с убыванием значимости в первой зафиксированной подвыборке

## Выводы

* при многократном повторении эксперимента с одним и тем же размером группы результаты могут отличаться очень существенно.
* при росте испытуемых наблюдается увеличение количества значимых “закономерностей”, которые не обнаруживаются на меньшем объеме выборки
* увеличение допустимого уровня значимости за счет коррекции на множественные сравнения приводит к очень существенному снижению шансов на обнаружение эффекта
* в рамках стандартных статистических процедур сложно получить уверенные выводы при типичных размерах выборки (до 60), необходимо обращаться к методам, которые учитывают кластерную структуру либо априорные данные (байесовские методы)

## Литература

Cohen, Mike X. 2014. *Analyzing Neural Time Series Data: Theory and Practice*. The MIT Press. https://doi.org/10.7551/mitpress/9609.001.0001.

Cremers, Henk R., Tor D. Wager, and Tal Yarkoni. 2017. ‘The Relation between Statistical Power and Inference in FMRI’. *PLOS ONE* 12 (11): e0184923. https://doi.org/10.1371/journal.pone.0184923.

Maris, Eric, Jan-Mathijs Schoffelen, and Pascal Fries. 2007. ‘Nonparametric Statistical Testing of Coherence Differences’. *Journal of Neuroscience Methods* 163 (1): 161–75. https://doi.org/10.1016/j.jneumeth.2007.02.011.