*ФИЗИОЛОГИЯ ЧЕЛОВЕКА, 2023, том 49, № 4, с. 30–40*

*УДК 612.821*

# РОЛЬ СРЕДНЕГО МОЗГА В ВОСПРИЯТИИ ТОНАЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И РЕЧИ: АНАЛИЗ ИНДИВИДУАЛЬНЫХ НАБЛЮДЕНИЙ

**© 2023 г. Л. Б. Окнина[1, *], А. О. Канцерова[1], Д. И. Пицхелаури[2], В. В. Подлепич[2], Г. В. Портнова[1], И. А. Зибер[3], Я. О. Вологдина[1, 2], А. А. Слезкин[1, 4], А. М. Ланге[5], Е. Л. Машеров[2], Е. В. Стрельникова[1]**

*[1]ФГБУН Институт высшей нервной деятельности и нейрофизиологии РАН, Москва, Россия*

*[2]ФГАУ НМИЦ нейрохирургии имени академика Н.Н. Бурденко Минздрава РФ, Москва, Россия*

*[3]Национальный исследовательский институт "Высшая школа экономики", Москва, Россия*

*[4]ФГБОУ ВО Российский технологический университет (МИРЭА), Москва, Россия*

*[5]Сколковский институт науки и технологий, Москва, Россия*

*****E-mail: leliia@yandex.ru*

Поступила в редакцию 25.05.2022 г.
После доработки 10.11.2022 г.
Принята к публикации 31.01.2023 г.

Человеческая речь представляет собой сложную комбинацию звуков, слуховых событий. В настоящее время нет единого мнения о том, как происходит восприятие речи. Реагирует ли мозг на каждый звук в потоке речи отдельно или в звуковых рядах выделяются дискретные единицы, анализируемые мозгом как одно звуковое событие. В данном пилотном исследовании проанализированы ответы среднего мозга человека на простые звуки, комбинации простых тонов ("сложные" звуки) и лексические стимулы. Работа представляет собой описание единичных случаев, полученных в рамках интраоперационного мониторинга во время хирургического лечения опухоли глубинных срединно расположенных опухолей головного мозга или ствола мозга. В исследование включены данные регистрации потенциалов ближнего поля из среднего мозга у 6 пациентов (2 женщины, 4 мужчины). Были выделены *S*- и *E*-комплексы, возникающие при начале и окончании звучания, а также *S*-комплексы, возникающие при смене структуры звука. Полученные данные позволяют предположить, что выделенные комплексы являются маркерами первичного кодирования звуковой информации и генерируются структурами нейронной сети, обеспечивающей восприятие и анализ речи.

*Ключевые слова:* средний мозг, вызванные потенциалы, восприятие речи, восприятие звуков, связанные с событиями потенциалы.

**DOI:** 10.31857/S0131164623600052, **EDN:** XAVGVT

Человек способен различать звуки в диапазоне от 20 до 20000 Гц [1]. Человеческая речь представляет собой сложную комбинацию звуков, слуховых событий. Человек учится распознавать речь в течение нескольких лет, а потом совершенствует этот навык всю жизнь [2].

Для восприятия речи мозгу необходимо различать особенности звучания отдельных лексических единиц [3–6]. При этом восприятие лексической информации имеет четкую зависимость от способности воспринимать звуки окружающей среды [7]. Слуховая система способна воспринимать два звука как отдельные, если они отстоят друг от друга более, чем на 30 мс [8]. Кроме того, человек способен воспринимать отличие одного тона от другого, если изменения по частоте составляют больше 0.2% [9]. Это позволяет воспринимать речь, которая представляет собой постоянную смену частот и амплитуд звука.

В настоящее время нет единого мнения о том, как происходит восприятие речи. Реагирует ли мозг на каждый звук в потоке речи отдельно или в звуковых рядах выделяются дискретные единицы, анализируемые мозгом как одно звуковое событие. Существует предположение, что при восприятии "сложных", состоящих более чем из одного синусоидального тона, звуков может происходить достаточно сильное огрубление оценки и интерференция звуков [10]. При этом нейрофизиологические механизмы их восприятия сходны с таковыми для восприятия простых тонов [11].

В слуховой системе выделяют периферический и центральный отделы. Активность периферических структур при восприятии слуховой ин-

формации изучена достаточно полно, так же, как и активность коры полушарий [12–14]. Тогда как большая часть исследований роли подкорковых структур, входящих в слуховую систему, выполнена на животных. Однако прямой перенос данных, полученных в исследованиях на животных, на человека невозможен. Кроме того, необходимо учитывать, что многие исследования, которые сейчас признаны классическими и составляют базовую основу нейрофизиологии слуховой системы, проводились в середине прошлого века, и имеют ограниченный, по сравнению с настоящим временем, технические возможности [1, 10, 13, 15–17].

В клинических исследованиях на человеке широко используют регистрируемые с поверхности головы акустические стволовые вызванные потенциалы (АСВП), которые отражают проведение слуховой информации по структурам слуховой системы к коре головного мозга. В последнее десятилетие благодаря быстрому развитию высокотехнологичных медицинских методик в отдельных случаях стало возможным получить информацию об активности структур, локализованных в глубине мозга, регистрируя биопотенциалы при помощи электродов, имплантированных непосредственно в исследуемую структуру. В таких случаях выполняется регистрация потенциалов ближнего поля (*local field potential*, *LFP*) [18]. Анализируя *LFP* исследователями предпринимаются попытки проследить, как происходит восприятие звуков и человеческой речи в структурах слуховой системы и непосредственно в слуховой коре [19–21]. При этом остается открытым вопрос, будет ли мозг реагировать на каждое изменение звука отдельно, собирая все изменения в лексические единицы только в коре, или же существует дискретизация речи на более ранних этапах.

В нашей предыдущей работе было показано, что средний мозг реагирует на начало и окончание простого звука, появлением *S*-комплексов и *E*-комплексов, соответственно. Названия комплексов даны от *start* (*S*-комплекс) и *end* (*E*-комплекс). *S*-комплексы, по-видимому, сходны с описанными в литературе *on*-потенциалами, тогда как *E*-комплексы предположительно отличаются от описанных *off*-потенциалов как по морфологии ответа, так и по функциональному значению [22]. Однако полученные ранее данные не отвечают на вопрос, как будут восприниматься естественные звуки и отдельные речевые стимулы. Исходя из этого, в настоящем исследовании была сформулирована гипотеза, что уже на уровне среднего мозга происходит разделение поступающей слуховой информации на единицы, которые в коре будут анализироваться как единая структура. Для подтверждения гипотезы была поставлена задача – проанализировать ответы среднего мозга человека на слуховые стимулы разной сложности: простые звуки, “сложные” звуки, лексические стимулы и искусственно сгенерированные стимулы, сходные по своему частотному составу с лексическими стимулами. Работа представляет собой описание единичных случаев, полученных в рамках интраоперационного мониторинга во время хирургического лечения опухоли глубинных срединно расположенных опухолей головного мозга или ствола мозга.

## МЕТОДИКА

В настоящее исследование были включены данные регистрации потенциалов ближнего поля из среднего мозга у 6 пациентов (2 женщины, 4 мужчины), проходивших хирургическое лечение в ФГАУ МНИЦ нейрохирургии им. акад. Н.Н. Бурденко МЗ РФ (г. Москва). У всех пациентов была полная сохранность слуховой системы и когнитивных функций до и после операции. Характеристики пациентов представлены в табл. 1.

*Регистрация потенциалов*. До операции пациентам проводили краниометрическую оценку точек постановки электродов. Перед началом операции на поверхности головы устанавливали электроды по схеме 10–20%. Из схемы исключали электроды, расположенные в непосредственной близости от области работы нейрохирурга.

Глубинный электрод устанавливался нейрохирургом в водопровод мозга во время проведения интраоперационного электрофизиологического мониторинга с целью картирования ядер и проводящих путей внутри ствола. Электрод состоял из силиконовой трубки диаметром 2.7 мм, которая на конце содержит три электрод-контакта, расположенные через 3.5 мм друг от друга. Ширина электрода-контакта – 3 мм. Электрод в водопроводе устанавливали таким образом, чтобы электрод-контакты располагались в проекции верхних и нижних бугорков четверохолмия. Референтом служил третий электрод-контакт, который в зависимости от хирургического доступа располагали в просвете третьего или четвертого желудочка. Более подробно методика установки глубинных электродов и регистрации локальных потенциалов описана в работе [22].

В качестве референтного для электродов на поверхности головы использовали объединенный ушной электрод. Глубинные потенциалы регистрировали биполярно с собственным референтом. Заземляющий электрод был общим и располагался в проекции плечевого сустава. Отдельным каналом регистрировали осцилограмму звуковых стимулов, отражающую ток, подаваемый в наушники при звуковой стимуляции.

Регистрацию проводили прибором ИОМ-4 “Нейрософт” (Россия). Частота дискретизации

**Таблица 1.** Характеристика пациентов, принявших участие в исследовании, и использованные парадигмы

| Характеристика пациентов | | | | | | Парадигмы | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| пациент | пол | опухоль | хирургический доступ | пропофол мг/кг/ч во время стимуляции | тип ЭЭГ | простые тоны | “сложные” звуки | свое имя | шум | гласные, слоги |
| 1 | ж | Левый ЗБ | ТВ | 7 | *BS* | + | + | + | + | |
| 2 | ж | ПМ | МС | 7 | *NBS-BS* | + | + | + | + | |
| 3 | м | ПО и ПМ | МС | 7.8 | *BS* | + | + | + | + | |
| 4 | м | III желудочек | ТВ | 10.4 | *BS* | + | + | + | + | |
| 5 | м | IV желудочек | МС | 6.09 | *BS* | + | | + | | + |
| 6 | м | IV желудочек | МС | 5.81 | *BS* | + | | + | | + |

*Примечание:* ЗБ – зрительный бугор, ПМ – продолговатый мозг, ПО – пинеальная область; ТВ – транскортико-трансвентрикулярный доступ, МС – срединный субокципитальный доступ; *BS* – наличие вспышек типа “вспышка–подавление” (*burst-suppression*), *NBS* – отсутствие вспышек типа “вспышка–подавление” (*no burst-suppression*).

составляла 10000 Гц. Использовали полосу пропускания 0.01–4000 Гц, режекторный фильтр 50 Гц.

Регистрацию проводили на фоне общего эндотрахеального наркоза. В качестве анестетика использовали пропофол, доза которого устанавливалась анестезиологом на основании клинических данных (табл. 1). Во время предъявления всех слуховых стимулов на ЭЭГ регистрировали комплексы “вспышка–подавление”.

*Звуковые последовательности*. Звуковые последовательности предъявляли с использованием программы “*Presentation” Neurobehavioral Systems, Inc*. (США). Предъявляемые звуки имели частоту дискретизации 44.1 кГц и разрядность 16 бит. Все звуки предъявляли бинаурально через накладные наушники. В каждой последовательности стимулы предъявляли в псевдослучайном порядке.

*Простые звуки* входили в *oddball* последовательность, которая состояла из тонов 600 (20%) и 800 Гц (80%), длительностью 80 мс, из которых восходящая и нисходящая фаза составляли 10 мс. Всего подавали 100 звуков. Межстимульный интервал варьировал от 1100 до 1170 мс.

*Комбинации простых тонов, или “сложные” звуки*. Каждый “сложный” звук включал 7 простых тонов. Все стимулы начинались и заканчивались тоном 1500 Гц. Начальный участок имел восходящий фронт длительностью 10 мс, финальный участок имел нисходящий фронт длительностью 10 мс. Между двумя тонами частотой 1500 Гц происходило чередование тонов с постоянной частотой 600, 800, 1000, 2000 и 4000 Гц. Переход от одной частоты к другой осуществляли в нулевой точке синусоиды, чтобы избежать возникновения “щелчка”. Были использованы 4 “сложных” звука, отличающихся друг от друга последовательностями фрагментов разной частоты: первый тип – 600, 800, 1000, 2000, 4000 Гц, второй тип – 4000, 2000, 1000, 800, 600 Гц, третий тип – 1000, 600, 2000, 800, 4000 Гц, четвертый тип – 2000, 800, 1000, 4000, 600 Гц. Амплитуда фрагментов разных частот, входящих в состав “сложного” звука, была идентичной. Длительность фрагментов каждой частоты составляла 50 мс. Длительность “сложного” звука – 350 мс.

Каждый “сложный” звук предъявляли 60 раз. Общее число стимулов в последовательности – 240. Межстимульный интервал составил 1500 мс с варьированием 10%.

*Лексические стимулы*. Всего использовали две лексические последовательности.

Последовательность “гласные и слоги” включала шесть основных русских гласных – *а*, *о*, *у*, *э*, *и*, *ы* и слоги *па*, *та*, *ка*, *ба*, *да*, *га*, *са*, *ша*. Гласные и слоги были записаны в изолированном произнесении диктора-мужчины в тихом помещении на диктофон *ZoomH2n* и преобразованы в монофонический формат. В естественном произнесении гласные имели разную длительность и были искусственно приведены к 390 мс. В слогах был использован только один гласный звук *а* [а], чтобы избежать влияния гласного на качество согласного. Согласные различались по глухости/звонкости (*па-ба*, *ка-га*), месту образования (*па-та-ка*, *ба-га-да*, *са-ша*), способу образования (*та-са*). Звонкие и глухие звуки, звуки разных мест и способов образования имеют разную собственную длительность, поэтому слоги по длительности не унифицировали. Длительность слогов варьировала от 262 до 417 мс. Каждый стимул в последовательности предъявляли 85 раз.

Последовательность “имя и шум” подготавливалась индивидуально для каждого пациента и включала в себя собственное имя пациента, чужое имя, анаграмму имени и альтернативный частотный шум (розовый шум). Анаграмма имени –

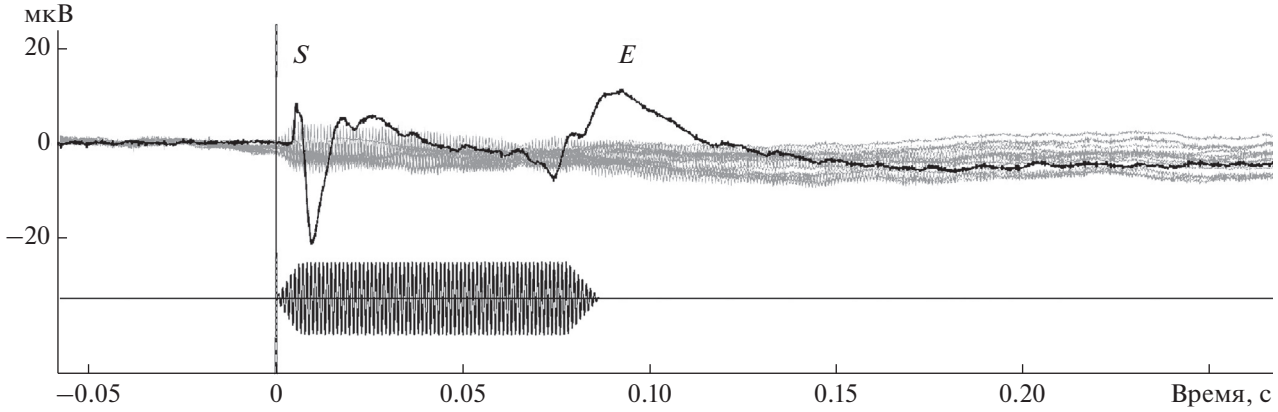


**Рис. 1.** Ответ среднего мозга пациента (индивидуальные данные) в ответ на простой тон (800 Гц).
Верхняя кривая – вызванные потенциалы: черная линия – на глубинном электроде, серые линии – на электродах на поверхности головы. Данные представлены без фильтрации. Нижняя линия – осцилограмма тона, отражающая звучание тона.

специально подготовленный звук, складывающийся из перестановки звуков имени, который удовлетворял следующим условиям: 1) звуковой стимул воспринимался как псевдослово (нельзя было узнать имя); 2) переставляли четко дифференцируемые фонемы; 3) использовали одинаковое количество перестановок. Альтернативный шум генерировали на основе средней частоты и громкости звука собственного имени при помощи модифицированного *toolbox Generate pink noise* (*MATLAB*, *MathWorks*). В результате полученный звуковой файл имел ту же длительность, среднюю громкость и звонкость, что и звук собственного имени. Длительность стимулов в последовательности варьировала от 0.5 до 1 с (в зависимости от длины имени пациента). Межстимульный интервал составлял 1500 мс с вариацией 10%. Каждый стимул предъявляли 35 раз.

*Анализ данных.* Зарегистрированные потенциалы обрабатывали в программе "*Brainstorm*" [23]. К анализу принимали безартефактные участки записи. Вызванные потенциалы в ответ на каждый стимул вычисляли отдельно.

Учитывая малое количество наблюдений, проводили визуальный анализ полученных данных.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На электродах, расположенных на поверхности головы, не были выделены ВП в ответ на звуковые стимулы ВП ни в одной из предъявляемых звуковых последовательностей.

На глубинных электродах регистрировались ВП в ответ на все предъявляемые стимулы. Однако пики, выявляемые в ВП в ответ на разные стимулы, отличались.

*В ВП в ответ на простые тоны* у всех пациентов отчетливо выделились два комплекса пиков. Первый комплекс следовал сразу за началом звучания – *S*-комплекс. Второй – *E*-комплекс – выделялся после окончания звука. Данные комплексы пиков выделялись в ответ на все простые тоны. Данные пики были описаны ранее [22]. На рис. 1 показан ответ среднего мозга с выделенными комплексами *S* и *E* в ответ на начало и конец звука, соответственно, и осцилограмма простого тона. Подобный характер ответов был характерен для всех пациентов.

*В ВП в ответ на "сложный" звук* также выделялись комплексы *S* и *E*, которые регистрировались в начале и в конце стимула, соответственно (рис. 2). Помимо этого, при каждой смене частоты отмечался комплекс, сходный с *S*-комплексом начала стимула. Необходимо отметить, что этот комплекс проявлялся вне зависимости от того, между какими частотами происходила смена. При этом *E*-комплекса при смене частот не регистрировалось. Подобный характер пиков в ВП был характерен для всех 4 пациентов.

*В ВП в ответ на гласные звуки* также регистрировались комплексы *S* и *E*. ВП в ответ на слог мог содержать дополнительные комплексы, сходные с *S* в середине звучания звука. На рис. 3 (II) показан ВП пациента в ответ на слог *да* [да]. В середине звука отмечается комплекс сходный с комплексом *S*, предположительно соответствующий переходу взрывного звука *д* [д] в гласный *а* [а]. У второго пациента, которому предъявляли данную последовательность, ВП имели аналогичное распределение пиков.

Необходимо отметить, что на рис. 1 и 2, помимо ВП на глубинном электроде, приведена активность, регистрируемая на поверхности головы. Учитывая неизменный характер ЭЭГ на протяжении всей записи, и отсутствие ВП в ответ на все

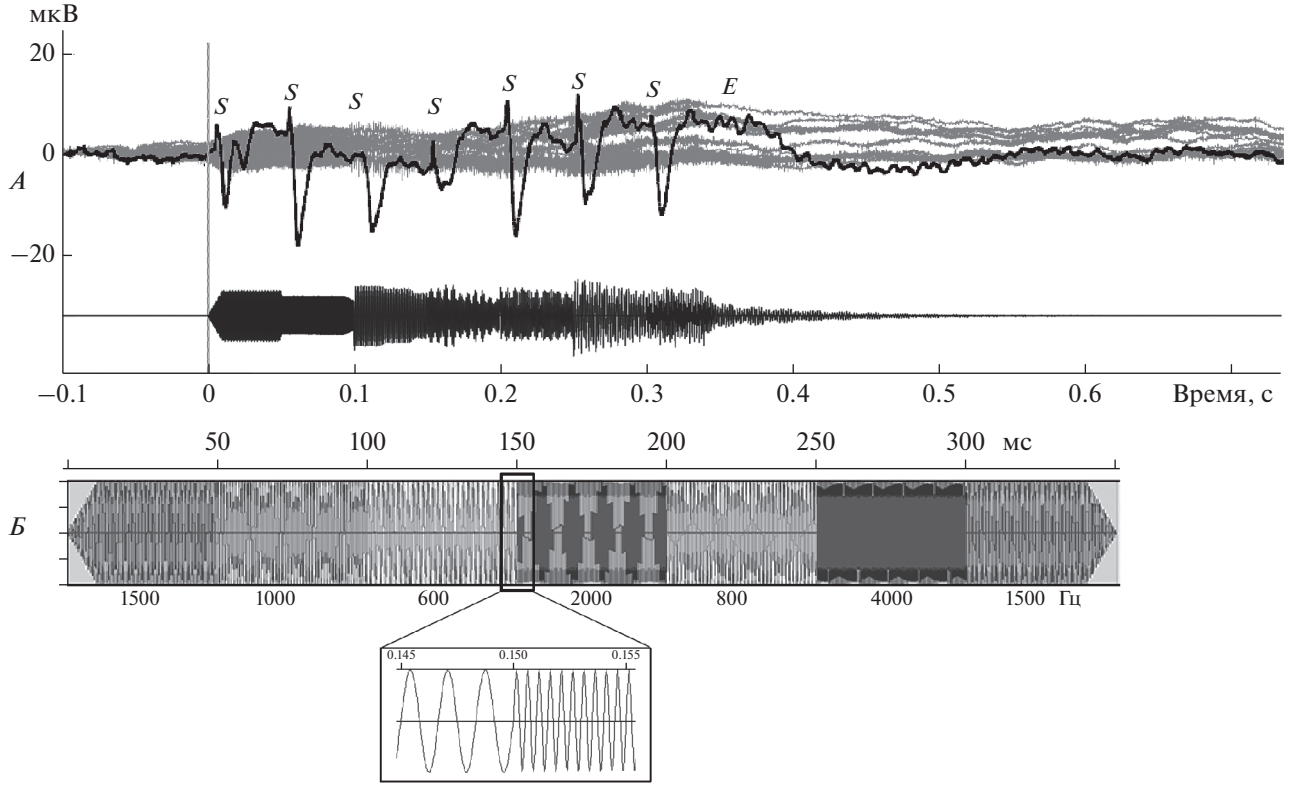


**Рис. 2.** Ответ среднего мозга пациента (индивидуальные данные) в ответ на “сложный” составной тон.
*А* – верхняя кривая – вызванные потенциалы (ВП) пациента в ответ на “сложный” тон. Черная линия – ВП на глубинном электроде, ВП серые линии – на электродах на поверхности головы. Нижняя кривая – осцилограмма звука. Отличия амплитуды участков с разными частотами являются не истинным различием, а артефактом визуализации при цифровой записи, возникающие вследствие зависимости соотношения частоты сигнала и частоты оцифровки. Данные представлены без фильтрации. *Б* – структура “сложного” звука. Для визуализации использован звуковой редактор “*Audacity*”. Выноска показывает переход одной частоты в другую в нулевой точке синусоиды при неизменной амплитуде колебаний.

стимулы, используемые в исследовании, в дальнейшем она не приводится.

*ВП в ответ на другие гласные и слоги* имели сходный характер – наличие комплексов *S* и *E*, соответствующие началу и концу стимула и возникновение комплекса *S* при переходе от одного звука к другому.

*В ВП в ответ на имена*, искусственно сгенерированное псевдоимя и шум также выделяются комплексы *S* и *E*, соответствующие началу и окончанию стимула. При этом в ответ на имена и искусственное псевдоимя в центре слова также могли присутствовать комплексы, сходные с *S*, которые соответствовали времени резкой смены частот в стимуле. В ВП в ответ на шум в середине звучания стимула ни у одного из пациентов не было выявлено комплексов, сходных ни с *S* ни с *E*.

На рис. 4 представлены индивидуальные ответы одного и того же пациента на собственное имя, чужое имя, псевдоимя и розовый шум.

На рис. 5 показаны индивидуальные ответы на собственное имя. *S*-комплекс отчетливо выявляется в начале стимула и при смене слогов или возникновении нового звука при произнесении имени. В конце у всех пациентов отмечается *E*-комплекс.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В настоящем исследовании проверялась гипотеза, что у человека первичная оценка лексических стимулов происходит уже на уровне среднего мозга. В работе были зарегистрированы и проанализированы потенциалы ближнего поля, зарегистрированные с помощью глубинных электродов непосредственно из среднего мозга.

Для повышения точности выделения ответов среднего мозга в работе, помимо активности мозга, отдельным каналом регистрировалась осцилограмма звукового сигнала, что позволило контролировать точность постановки метки стимула и снизить вероятность технических ошибок при усреднении ВП. Необходимо отметить, зарегистрированная осцилограмма не позволяет оценить параметры подаваемого звука в силу соотно-

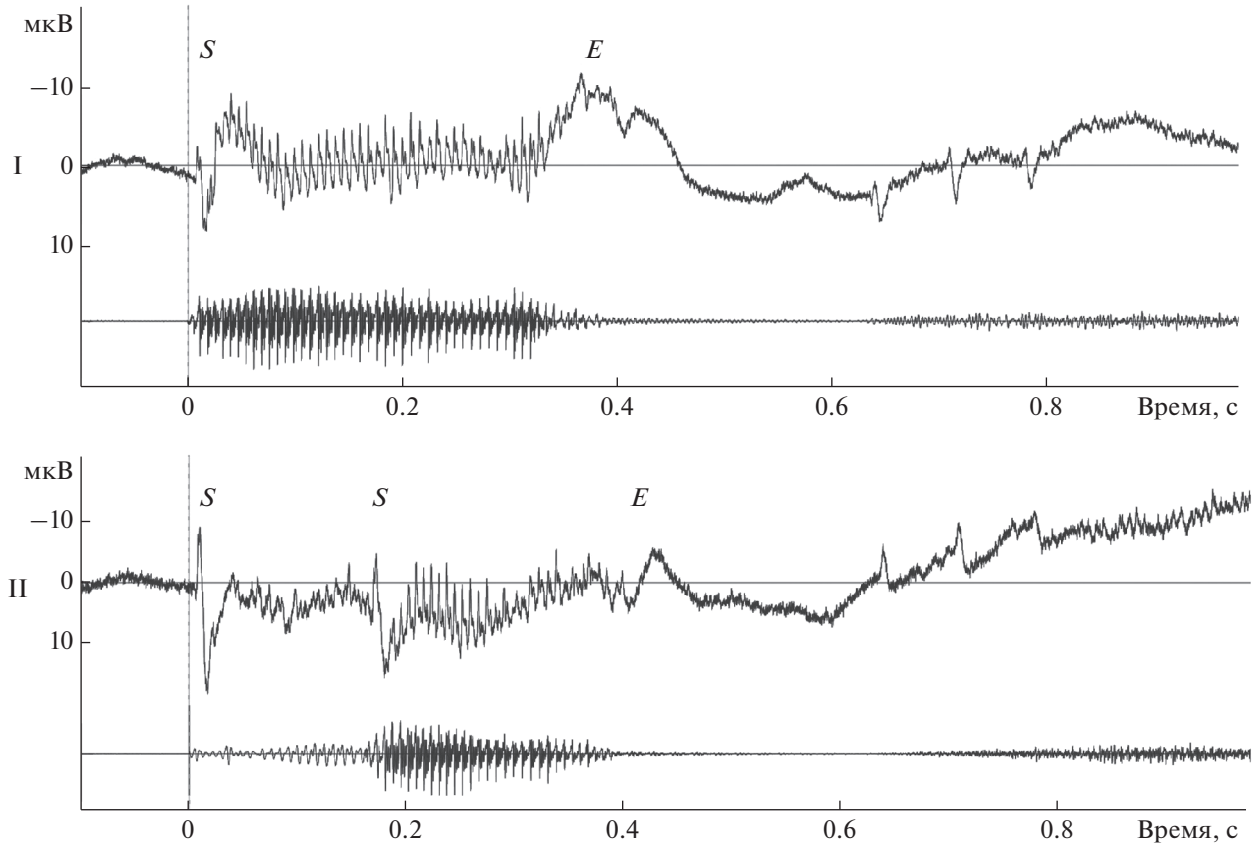


**Рис. 3.** Вызванные потенциалы пациента I – в ответ на гласный *а* [а], II – в ответ на слог *да* [да].
Верхняя кривая – ответ среднего мозга, зарегистрированный на глубинном электроде, нижняя кривая – осцилограмма стимула. Данные представлены без фильтрации.

шения частоты подаваемого звукового стимула и чистоты дискретизации.

Учитывая, что восприятие речи невозможно без способности воспринимать остальные звуки, в работе, помимо лексических стимулов и сходных по частотному составу искусственно сгенерированных стимулов, были оценены ВП среднего мозга в ответ на простые тоны и “сложные” звуки, включающие несколько простых тонов. При этом части “сложного” звука отличались между собой только по частоте, тогда как амплитуда была неизменной. Это позволило исключить влияние амплитуды на параметры регистрируемых ответов среднего мозга, описанных в литературе [21]. Кроме того, переход одной частоты в другую был через нулевую точку синусоиды, что позволило избежать “щелчка”. Однако в лексических стимулах полностью исключать влияние амплитуды на ответы среднего мозга невозможно, поскольку стимулы были подготовлены в естественном произнесении.

Ранее нами уже были описаны реакции среднего мозга на начало и конец простого тона [22]. В настоящем исследовании при смене частоты “сложного” звука были выявлены компоненты, сходные с *S*-компонентом, регистрируемом в ответ на начало звука. При этом *E*-компонент не регистрируется. Можно полагать, что *S*-комплекс в ответ на начало тона сходен по своему значению с описанными ранее *on*-потенциалами, тогда как в середине тона отражает изменение частоты. То есть можно полагать, что выделяемый комплекс разделяет структуры, которые анализируются в вышележащих структурах слуховой системы как отдельная единица.

Наименьшая единица языка, имеющая собственный смысл – морфема. Минимальная единица языка, позволяющая различать схожие по звучанию, но имеющие разный смысл слова – фонема. Носители языка хорошо распознают фонемы, т.е. смыслоразличительные единицы языка [6, 24–26]. Кроме того, в восприятии окружающей действительности большую роль играют субъективный опыт человека и контекст предъявляемых речевых стимулов. Он формирует ожидание входящей информации, которая подвергается интеграции с ранее полученным опытом. И только на основании подобной интеграции происходит оценка входного сигнала [27]. При подготовке стимульных последовательностей мы учитыва-

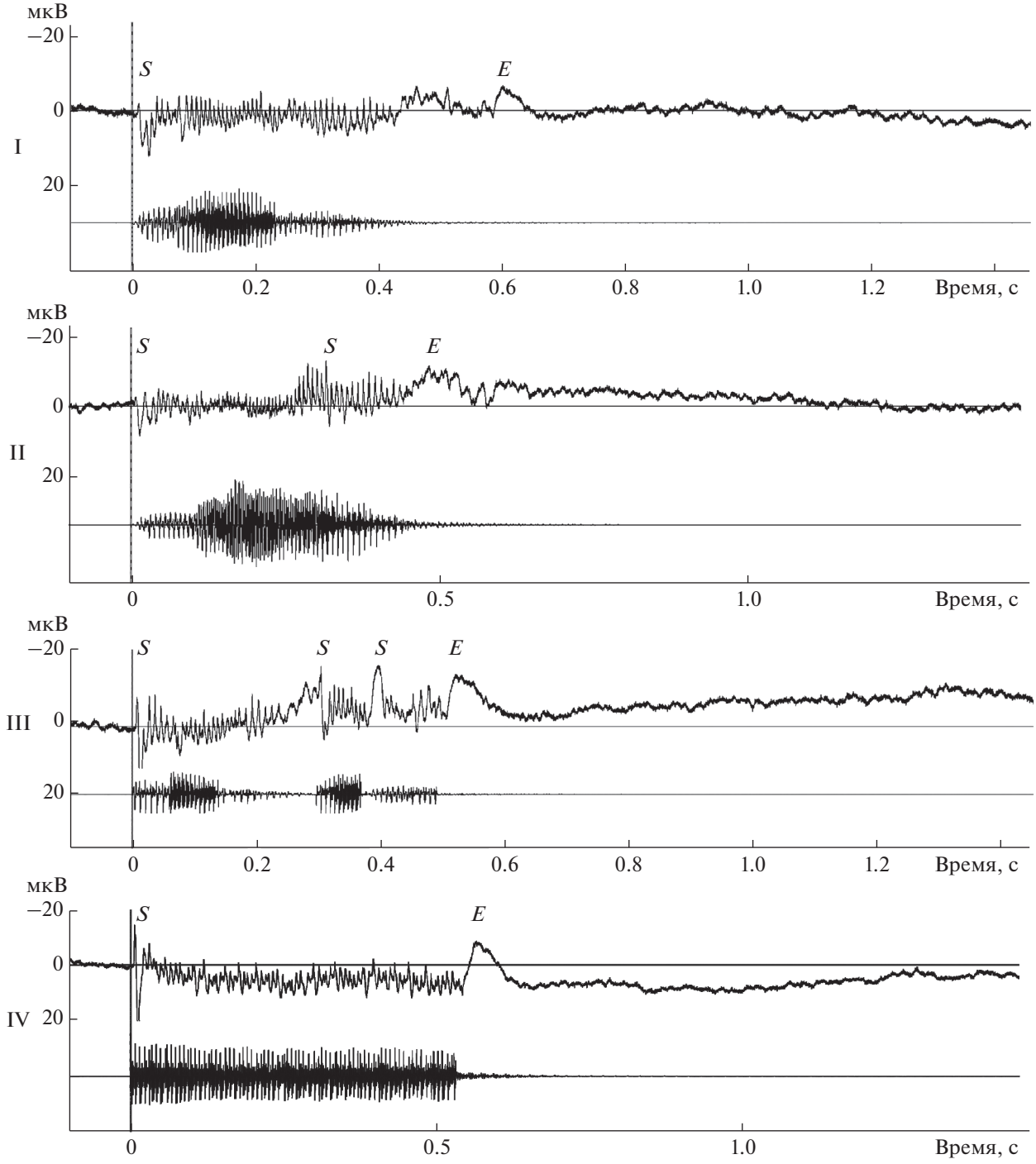


**Рис. 4.** Вызванные потенциалы в ответ на разные типы стимулов у одного и того же пациента.
I – на свое имя [Галя], II – чужое имя [Юля], III – искусственно сгенерированное псевдоимя, IV – розовый шум. Верхняя кривая – ответ среднего мозга, зарегистрированный на глубинном электроде, нижняя кривая – зарегистрированная осцилограмма стимула. Данные представлены без фильтрации.

ли субъективный опыт человека: лексическая последовательность была индивидуальной для каждого пациента и включала собственное имя человека, которое, предположительно, должно распознаваться легче других речевых стимулов и может восприниматься не как набор звуков, а как единый смысловой образ.

Признавая многоступенчатость восприятия речи, в настоящее время нет единого мнения о том, какая из областей мозга важна для восприя-

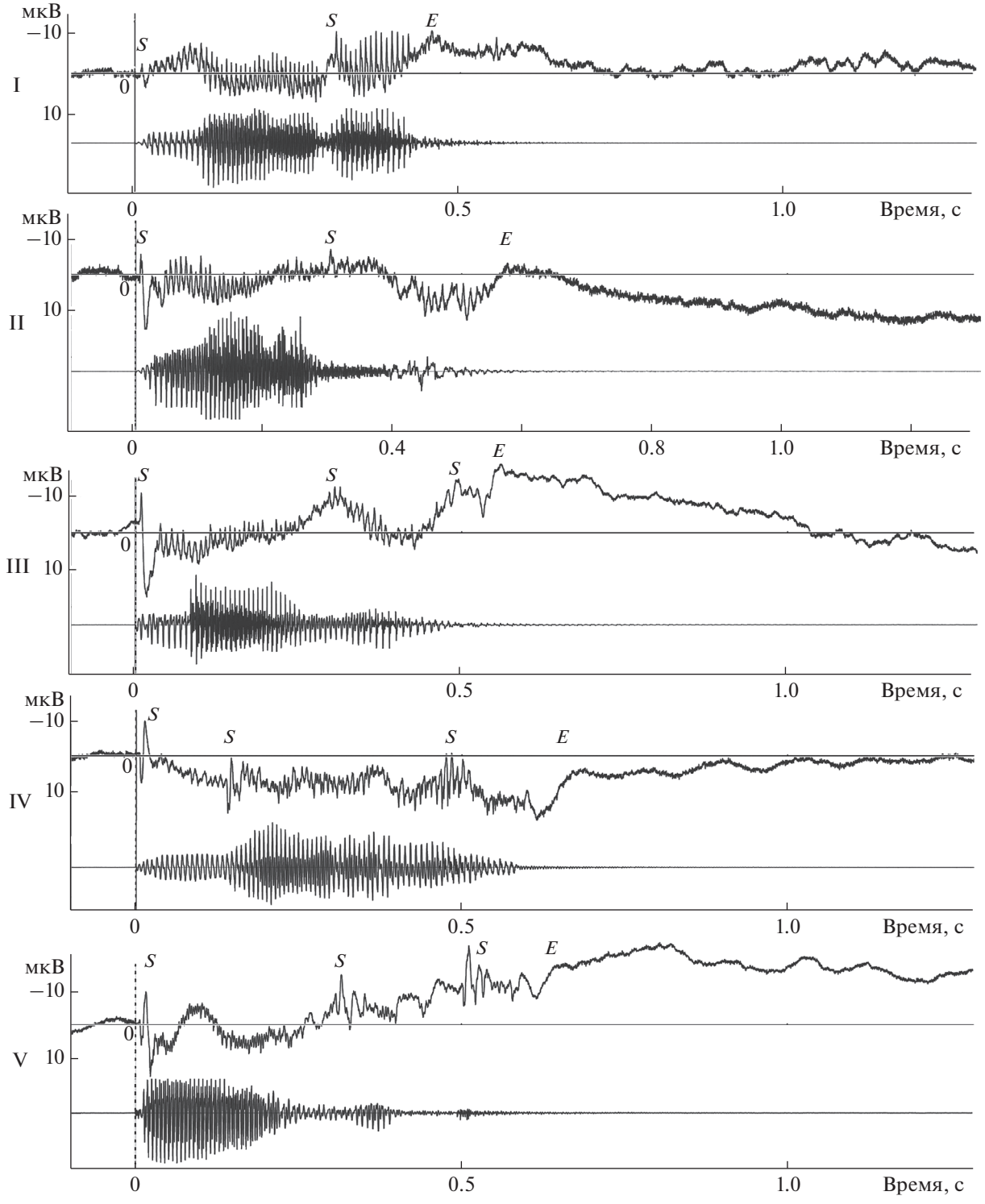


**Рис. 5.** Вызванные потенциалы в ответ на свое имя у 5 пациентов (индивидуальные данные).
I – имя [Вера], II – имя [Леша], III – имя [Шурик], IV – имя [Роман], V – имя [Вадим]. Верхняя кривая – ответ среднего мозга, зарегистрированный на глубинном электроде, нижняя кривая – зарегистрированная осцилограмма стимула. Данные представлены без фильтрации.

тия речи в целом, а какая – для восприятия ее отдельных лексических структур. Наиболее распространены две модели восприятия речи. В первую модель включены акустико-фонетические особенности, фонемы и, собственно, слова [28]. В последнее десятилетие модель дополнена двух-

сторонними связями между близлежащими уровнями. Это позволяет предшествующему лексическому или фонологическому опыту влиять на протекающую фонологическую или акустико-фонетическую оценку. Это делает модель удовлетворяющей требованиям акустико-фонологического взаимодействия в процессе анализа речевого сигнала. Вторая модель, которую можно использовать для моделирования работы слухоречевой системы в данном контексте, несколько лучше объясняет возможность интеграции сенсорного и обратного ("сверху-вниз") контроля обработки информации. Модель описывает такую интеграцию как "предсказательное кодирование". Эта модель имеет достаточно экспериментальных подтверждений. Модель предполагает, что контроль "сверху-вниз" осуществляет предсказание, сравнивая шаблон с входящим сигналом, и только незнакомая активность (ошибка предсказания) проходит дальнейшую обработку [27]. Однако обе модели не учитывают возможности дискретизации речи на целостные структуры, превышающие отдельные звуки, на более низких по отношению к коре иерархических уровнях.

Верхние иерархические уровни оценки речи локализуются в областях коры больших полушарий головного мозга. Это, прежде всего, верхняя височная извилина (*gyrus temporalis superior*) при первичном восприятии слуховых стимулов и нижняя лобная извилина (*gyrus frontalis inferior*), которую связывают с оценкой абстрактных лингвистических единиц и принятием решения [29, 30]. Анализ данных, полученных у животных (приматов), а также у человека методами оценки функциональных связей выявил, что между этими областями существуют реципрокные взаимоотношения, на основе которых осуществляется оценка входящего речевого сигнала [31, 32].

В лобной области происходит достаточно ранняя оценка стимулов до начала реализации контроля "сверху-вниз". Предполагается, что в нижней лобной извилине на временно́м интервале 90–130 мс от подачи стимула происходит оценка в соответствии с эффектом предшествующего опыта. Несмотря на искажения подаваемой речи и смысловую неоднозначность фонетической последовательности в первые 100 мс от начала стимула, уже в данном временно́м диапазоне происходит оценка речевого сигнала, хотя она носит достаточно "грубый" характер. Тем не менее, в этот момент в речевом сигнале уже имеется достаточно информации, чтобы произошло сопоставление входного сигнала с предшествующим опытом [27, 28].

Можно полагать, что подобная оценка становится возможной в том случае, когда к коре информация поступает уже разделенной на отдельные структурные единицы. Выявленные в случае сложного тона и искусственно сгенерированного стимула аналоги *S*-комплекса, сходные с таковыми в начале звука, могут служить маркерами сегментации сложных слуховых стимулов, включая речь. В таком случае в коре анализируется не отдельно каждый частотный элемент структуры речи, а набор частот, объединенных единой структурой, разделенной *S*-комплексами. Анализ учитывает наличие пауз, которые ведут к возникновению *E*-комплекса, свидетельствующему о завершении сегмента.

Можно полагать, что предшествующий опыт человека интегрируется с входящим речевым сигналом через реализацию механизмов контроля как "снизу-вверх", так и "сверху-вниз". Кодирование "снизу-вверх" начинается на уровне и с непосредственным участием среднего мозга, который осуществляет первичное выделение речевых субъединиц, соответствующих фонетической структуре. И уже эти субъединицы анализируются в коре больших полушарий с учетом персонального опыта восприятия речи и подчиняются общим принципам предсказательного кодирования для дальнейшего контроля "сверху-вниз".

Полученные данные позволяют предположить, что выделенные *S*- и *E*-комплексы, возникающие при начале и окончании звучания, а также *S*-комплексы, возникающие при смене структуры звука, являются маркерами первичного кодирования звуковой информации.

*Ограничения исследований.* Исследование выполнено в рамках интраоперационного мониторинга и носит характер описания и обобщения единичных наблюдений. Это ограничило возможности предъявления пациентам большого количества звуковых последовательностей. Важно отметить, что описанные феномены фиксировались и отчетливо проявлялись у всех пациентов, принятых в исследование. Вместе с тем, из-за небольшого количества пациентов, проведение полноценного статистического анализа невозможно.

## ЗАКЛЮЧЕНИЕ

Средний мозг реагирует на начало стимула, его окончание и резкую смену частотных характеристик стимула. Можно полагать, что, таким образом, происходит разделение входящей слуховой информации на отдельные дискретные единицы, которые в коре анализируются как целая единица.

***Этические нормы.*** Все исследования были проведены в соответствии с принципами биомедицинской этики, сформулированными в Хельсинкской декларации 1964 г. и ее последующих обновлениях. Исследования проводились в рамках плановых научных тем и комплексного клинического обследования при оказании медицин-

ской помощи в ФГАУ НМИЦ нейрохирургии им. академика Н.Н. Бурденко Минздрава России (Москва). Исследование было одобрено местным этическим комитетом.

***Информированное согласие.*** Каждый участник исследования представил добровольное письменное информированное согласие, подписанное им после разъяснения ему потенциальных рисков и преимуществ, а также характера предстоящего исследования.

***Финансирование работы.*** Исследование выполнено на средства Государственного задания Министерства науки и высшего образования РФ для ИВНД и НФ РАН (Москва); в рамках Программы фундаментальных исследований НИУ ВШЭ (Москва).

***Конфликт интересов.*** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией данной статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. *Renals S., Hain T.* Speech Recognition / The Handbook of Computational Linguistics and Natural Language Processing // Eds. Clark A., Fox C., Lappin S. Blackwells, 2010. Chapter 12. P. 299.
2. *Attneave F., Olson R.K.* Pitch as a medium: a new approach to psychophysical scaling // Am. J. Psychol. 1971. V. 84. № 2. P. 147.
3. *Boruta L., Peperkamp S., Crabbé B., Dupoux E.* Testing the Robustness of Online Word Segmentation: Effects of Linguistic Diversity and Phonetic Variation / Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics. Portland. Oregon. USA, 2011. P. 1.
4. *Boruta L.* A note on the generation of allophonic rules [электронный ресурс]. RT-0401. 2011. inria-00559270v1.
5. *Kuhl P.K.* Early language acquisition: Cracking the speech code // Nat. Rev. Neurosci. 2004. V. 5. № 11. P. 831.
6. *Shea C., Curtin S.* Discovering the relationship between context and allophones in a second language: evidence for distribution-based learning // Stud. Second Lang. Acquis. 2010. V. 32. № 4. P. 581.
7. *Kazanina N., Phillips C., Idsardi W.* The influence of meaning on the perception of speech sounds // Proc. Natl. Acad. Sci. U.S.A. 2006. V. 103. № 30. P. 11381.
8. *Shahin K., Johnson K.* Acoustic and Auditory Phonetics // Language. 1999. V. 75. № 4. P. 870.
9. *Micheyl C., Xiao L., Oxenham A.J.* Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level // Hear. Res. 2012. V. 292. № 1–2. P. 1.
10. *Oxenham A.J.* How We Hear: The Perception and Neural Coding of Sound // Annu. Rev. Psychol. 2018. V. 69. № 1. P. 27.
11. *Lau B.K., Mehta A.H., Oxenham A.J.* Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies // J. Neurosci. 2017. V. 37. № 37. P. 9013.
12. *Радионова Е.А.* Опыты по физиологии слуха. Нейрофизиологические и психофизические исследования. СПб.: Ин-т физиологии им. И.П. Павлова, 2003. 256 с.
13. *Alavash M., Tune S., Obleser J.* Modular reconfiguration of an auditory control brain network supports adaptive listening behavior // Proc. Natl. Acad. Sci. U.S.A. 2019. V. 116. № 2. P. 660.
14. *Столярова Э.И.* Моделирование механизмов слуховой обработки речевых сигналов // Речевые технологии. 2010. № 2. С. 31.
15. *Hackett T.A., Barkat T.R., O'Brien B.M.J. et al.* Linking topography to tonotopy in the mouse auditory thalamocortical circuit // J. Neurosci. 2011. V. 31. № 8. P. 2983.
16. *Guo W., Clause A.R., Barth-Maron A., Polley D.B.* A Corticothalamic Circuit for Dynamic Switching between Feature Detection and Discrimination // Neuron. 2017. V. 95. № 1. P. 180.
17. *Moerel M., De Martino F., Formisano E.* An anatomical and functional topography of human auditory cortical areas // Front. Neurosci. 2014. V. 8. P. 225.
18. *Herreras O.* Local field potentials: Myths and misunderstandings // Front. Neural. Circuits. 2016. V. 10. P. 101.
19. *Nourski K.V., Steinschneider M., Rhone A.E. et al.* Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings // Brain Lang. 2015. V. 148. P. 37.
20. *Moses D.A., Mesgarani N., Leonard M.K., Chang E.F.* Neural speech recognition: Continuous phoneme decoding using spatiotemporal representations of human cortical activity // J. Neural. Eng. 2016. V. 13. № 5. P. 056004.
21. *Частович Л.А.* Физиология речи. Восприятие речи человеком. М.: "Книга по Требованию", 2012. 386 с.
22. *Канцерова А.О., Окнина Л.Б. Пицхелаури Д.И. и др.* Вызванные потенциалы среднего мозга, ассоциированные с началом и окончанием звучания простого тона // Физиология человека. 2022. Т. 48. № 3. С. 5.
*Kantserova A.O., Oknina L.B., Pitskhelauri D.I. et al.* Evoked potentials of the midbrain associated with the beginning and end of a sound of a simple tone // Human Physiology. 2022. V. 48. № 3. P. 229.
23. *Tadel F., Baillet S., Mosher J.C. et al.* Brainstorm: A user-friendly application for MEG/EEG analysis // Comput. Intell. Neurosci. 2011. V. 2011. P. 879716.
24. *Shen Y.* Some Allophones Can Be Important // Language Learning. 1959. V. 9. № 1–2. P. 7.
25. *Richter C.* Learning Allophones: What Input Is Necessary ? / Proceedings of the 42nd annual Boston University Conference on Language Development. United States. Boston (3–5 November 2017). Cascadilla Press, 2018. P. 659.
26. *Mitterer H., Reinisch E., McQueen J.M.* Allophones, not phonemes in spoken-word recognition // J. Mem. Lang. 2018. V. 98. P. 77.
27. *Davis M.H., Sohoglu E., Peelle J.E., Carlyon R.P.* Predictive top-down integration of prior knowledge during

speech perception // J. Neurosci. 2012. V. 32. № 25. P. 8443.

28. *McClelland J.L., Elman J.L.* The TRACE model of speech perception // Cogn. Psychol. 1986. V. 18. № 1. P. 1.

29. *Scott S.K., Johnsrude I.S.* The neuroanatomical and functional organization of speech perception // Trends Neurosci. 2003. V. 26. № 2. P. 100.

30. *Poeppel D., Hickok G.* The cortical organization of speech processing // Nat. Rev. Neurosci. 2007. V. 8. № 5. P. 393.

31. *Anwander A., Tittgemeyer M., Cramon D.Y. et al.* Connectivity-based parcellation of Broca's area // Cereb. Cortex. 2007. V. 17. № 4. P. 816.

32. *Frey S., Campbell J.S.W., Pike G.B., Petrides M.* Dissociating the human language pathways with high angular resolution diffusion fiber tractography // J. Neurosci. 2008. V. 28. № 45. P. 11435.

# The Role of Midbrain in Perception of Tone Sequences and Speech: an Analysis of Individual Studies

**L. B. Oknina[a, *], A. O. Kantserova[a], D. I. Pitshelauri[b], V. V. Podlepich[b], G. V. Portnova[a], I. A. Ziber[c], J. O. Vologdina[a, b], A. A. Slezkin[a, e], A. M. Lange[d], E. L. Masherow[b], E. V. Strelnikova[a]**

[a] *Institite of Higher Nervous Activity and Neurophysiology of the RAS, Moscow, Russia*
[b] *Burdenko National Medical Research Center of Neurosurgery, Moscow, Russia*
[c] *National Research University Higher School of Economics, Moscow, Russia*
[d] *Russian Technological University (MIREA), Moscow, Russia*
[e] *Skolkovo Institute of Science and Technology, Moscow, Russia*
**E-mail: leliia@yandex.ru*

Human speech is a complex combination of sounds, auditory events. To date, there is no consensus on how speech perception occurs. Does the brain react to each sound in the flow of speech separately, or are discrete units distinguished in the sound series, analyzed by the brain as one sound event. The pilot study analyzed the responses of the human midbrain to simple tones, combinations of simple tones ("complex" sounds), and lexical stimuli. The work is a description of individual cases obtained in the frame of intraoperative monitoring during surgical treatment of tumors of deep midline tumors of the brain or brain stem. The study included local-field potentials from the midbrain in 6 patients (2 women, 4 men). The S- and E-complexes that emerge at the beginning and end of the sound, as well as the S-complexes that emerge when the structure of the sound changes, were identified. The obtained data suggest that the selected complexes are markers of the primary coding of audio information and are generated by the structures of the neural network that provides speech perception and analysis.

*Keywords:* midbrain, evoked potentials, event-related potentials, speech perception, sound perception.