

ОБЗОРНЫЕ СТАТЬИ

СЛУХОВАЯ АДАПТАЦИЯ К ХАРАКТЕРИСТИКАМ РЕЧЕВОГО СИГНАЛА

© 2022 г. И. Г. Андреева^{1,*}, Е. А. Огородникова^{1,2}

¹ Институт эволюционной физиологии и биохимии им. И.М. Сеченова Российской академии наук, Санкт-Петербург, Россия

² Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург, Россия

*e-mail: ig-andreeva@mail.ru

Поступила в редакцию 12.06.2022 г.

После доработки 29.07.2022 г.

Принята к публикации 02.08.2022 г.

В обзоре представлены данные исследований слуховой адаптации к ключевым параметрам речи — временным (темпо-ритмическим) и спектральным характеристикам голосового речевого сигнала, включая частоту основного тона голоса, формантные частоты, особенности тембра. Рассмотрены проявления адаптации к лингвистическим характеристикам голоса — полу и возрасту диктора, его эмоциональному состоянию. Показано, как механизмы адаптации участвуют в разделении конкурирующих речевых потоков и в механизмах слухового внимания. Обсуждается роль слуховой адаптации в становлении речевого слуха и формировании голосовых прототипов в процессе онтогенеза. Приведены основные модели процесса восприятия речи и данные по изучению нейрофизиологических механизмов фонематического анализа, на основе которых они формировались. Представлены результаты экспериментальных и модельных исследований, которые свидетельствуют о том, что адаптационные процессы играют важную роль в усилении контраста сигнала с фоном и приводят к улучшению идентификации сигнала. Рассмотренные в обзоре особенности слуховой адаптации к речевому сигналу имеют практическое значение для организации реабилитационных мероприятий по восстановлению и развитию речевого слуха пациентов после слухопротезирования, а также для разработки технических систем распознавания речи.

Ключевые слова: слуховое восприятие речи, слуховая адаптация, речевой сигнал, экстралингвистические характеристики речи, эффект вечеринки, слуховое внимание

DOI: 10.31857/S0044452922050035

ВВЕДЕНИЕ

Повседневное общение людей происходит на фоне шумов различной временной и спектральной структуры. Поэтому даже доклинические нарушения слуховой функции могут вызывать затруднения в распознавании речевого сигнала. Значительная потеря слуха приводит, как правило, к невозможности эффективной коммуникации в большинстве общественных мест. Современный уровень слухопротезирования не способен полностью решить задачу выделения целевого сигнала в шуме при общении пациента — пользователя слухового аппарата или кохлеарного импланта. Разработка фундаментальных вопросов о слуховых механизмах распознавания речи и ориентации в пространстве на фоне нестационарной помехи относится к области физиологии слуха, условно обозначаемой как проблема вечеринки (“the cocktail-party problem”). Она является одной из наиболее актуальных проблем как в фундаментальном плане, так и в прикладных разработках по разви-

тию технологий слухопротезирования, распознавания речи диктора в реальной зашумленной среде [1, 2].

Пространственные аспекты проблемы вечеринки подробно рассмотрены нами в обзоре [3], где обсуждаются бинауральные механизмы освобождения от маскировки и роль пространственной избирательности слухового внимания в выделении речевого сигнала на фоне окружающего шума. Изучение бинауральных механизмов в решении задачи пространственной настройки на голос диктора и в пространственном освобождении от маскировки имело существенный практический выход. Оно позволило обосновать необходимость двухстороннего протезирования слуха и разработку слухопротезирующих систем нового поколения, ориентированных на реализацию механизмов бинаурального взаимодействия [4].

Помимо локализации источника речевого сигнала его выделению из шума способствуют спектральные характеристики голоса диктора, такие

как основная частота голоса, его тембр, звукопронизношение и другие пара- и экстралингвистические характеристики. Индивидуальные или ситуационные особенности речи диктора в отношении ее темпа и ритмической организации также могут являться признаками, на основе которых происходит выделение и распознавание его речи в шуме. Распознавание речевого сигнала предусматривает его сегментацию и, таким образом, включает аспекты временного слухового анализа в условиях зашумления нестационарным сигналом. На фоне разнообразных шумов особое значение имеют процессы слуховой адаптации, которые приводят к усилению контраста сигнала и фона и, тем самым, способствуют улучшению условий для выделения и распознавания целевых признаков речи. Процессы адаптации широко распространены в слуховой системе и проявляются на разных уровнях обработки акустического сигнала — от волокон слухового нерва до структур коры головного мозга, где они связаны с центральной обработкой характеристик окружающей среды и речевых последовательностей, а также процессами слухового внимания и разделения конкурирующих звуковых потоков [5, 6]. Проявления адаптации в отношении индивидуальных спектрально-временных характеристик речи диктора при анализе сложной сцены будут рассмотрены в данном обзоре.

МАСКИРОВАНИЕ РЕЧЕВОГО СИГНАЛА И СЛУХОВАЯ АДАПТАЦИЯ

При разделении звуковых потоков в ситуации акустической сцены с множеством источников звука определяющее значение имеет способность слушателя выделять индивидуальные характеристики речи диктора [1]. В такой ситуации речь диктора выступает как полезный или целевой сигнал от интересующего нас в данный момент источника звука. Остальные составляющие акустической сцены — как помехи, конкурирующие с ним, и маскиры. Их действие обычно описывают в терминах энергетической и информационной маскировки. Энергетическая маскировка характеризуется спектральным и временным взаимодействием целевого сигнала и маскера [7]. Этот вид маскировки объясняют как результат перекрывания паттернов возбуждения, возникающих в ответ на целевой сигнал и маскер на периферии слуховой системы [8]. В отличие от энергетической информация маскировка влияет на центральные процессы в слуховой системе, связанные с вниманием и с анализом информации [9, 10]. Заметим, что в ситуации сложной акустической сцены, т.е. на фоне других акустических сигналов, величина информационной маскировки лингвистической информации (смысла речи) и экстралингвистических составляющих целевого речевого сообщения (пол, возраст, эмо-

циональное состояние диктора) могут быть различными.

При этом первой задачей в условиях сложной сцены оказывается обнаружение целевого сигнала. Обнаружению нового объекта в окружающей среде способствует адаптация слуховой системы к текущей информации. Считают, что адаптация к биологически значимым сигналам, в том числе речевым, в значительной степени обусловлена процессами в соответствующих областях коры больших полушарий головного мозга. В результате адаптации ограниченный диапазон реакции нейронов в процессе кодирования слуховых сигналов сдвигается в направлении расширения динамического диапазона этих реакций [11]. При изменении статистических свойств сигналов слуховая адаптация приводит к улучшению представления их информативных признаков, что позволяет рассматривать адаптационные процессы как эффективную стратегию кодирования [12–14]. Адаптивное кодирование определенных параметров стимулов было показано в нескольких подкорковых областях и областях коры [15, 16]. В частности, было выявлено, что слуховые кортикальные реакции у животных избирательно кодируют особенности вокализации в условиях шумов, имеющих разные спектральные характеристики [15–19].

Свойство слуховой системы обнаруживать новый источник информации (нового диктора) и разделять акустические потоки имеет высокую потенциальную ценность для развития технических систем распознавания речи, голосовых интерфейсов (voice user interface) и устройств голосового управления (voice command device) [20, 21]. Понимание и применение этого физиологического механизма способствуют решению технических вопросов распознавания речи на новом уровне.

СЛУХОВАЯ АДАПТАЦИЯ К ПАРАМЕТРАМ ЗВУКОВОГО СИГНАЛА

Слуховую адаптацию можно наблюдать при длительном предъявлении постоянного или повторяющегося стимула с одинаковыми характеристиками, которое вызывает временное изменение в восприятии последующих стимулов (слуховое последствие). Обычно отмечают негативный эффект слухового последствия, при котором наблюдается смещение перцептивной оценки в противоположном направлении относительно характеристик адаптирующего стимула. В основе механизма адаптации лежит изменение работы нейронных популяций, участвующих в анализе ключевых признаков последовательности сигналов. Одним из первых наблюдений слуховой адаптации является так называемый “тон Цвиккера”. Его появление связано с тем, что после прослушивания широкополосного шума, содержащего спектральную вырезку шириной около одной

трети октавы, у слушателя на несколько секунд возникает ощущение звучания тона, соответствующего ее характеристикам [22]. Потенциальные нейрональные корреляты для этой слуховой иллюзии были выявлены на уровне слуховой коры [23, 24]. По-видимому, ее проявление обусловлено временным усилением чувствительности, связанным с освобождением от торможения в центральных слуховых нейронах с частотой настроечной кривой в пределах спектральной вырезки. Эти нейроны были наименее стимулированы во время звучания адаптирующего шума. Тон Цвиккера демонстрирует особенность адаптации, которая отражает настройку анализатора на текущую структуру информационного потока. В результате воздействия сигналов настройка меняется, и адаптация непосредственно оказывает влияние на процесс сегрегации и группировки акустических потоков.

Исследования сегрегации акустических потоков показывают, что она определяется как обработкой сенсорных сигналов (обработка снизу вверх), так и когнитивным состоянием слушателя, например, фокусом внимания (обработка сверху вниз) [6, 25, 26]. Согласно представлению о слуховой адаптации последнее в значительной мере зависит от непосредственно предшествующей (текущей) информации. В ЭЭГ-исследованиях [27] было обнаружено, что эффективность разделения потоков при условии фокусировании внимания вне этих потоков обратно пропорциональна сходству между их корковыми представительствами. Этот факт свидетельствует о том, что адаптация к текущей слуховой информации является универсальным процессом, который лишь отчасти, но не полностью определяется состоянием внимания слушателя.

Адаптация слуховой системы к нестационарному речевому или речеподобному сигналу происходит по ряду его параметров, в том числе, уровню звука, его спектральным составляющим, ритмической структуре, положению его источника, спектрально-временному контексту речевой последовательности [28]. В эксперименте, как правило, исследуют адаптацию к определенному признаку звукового стимула, например, к амплитудной модуляции [29–31], длительности [32], ритму [33], спектральным характеристикам [34, 35]. Наряду с адаптацией к простым или “одномерным” признакам известны эффекты для сложных признаков, таких как положение источника звука в пространстве [36–38] или его движение в определенном направлении и с определенной скоростью [39–41]. Такие признаки являются сложными или комбинированными, так как содержат монауральные и бинауральные признаки локализации.

Слуховое последствие было выявлено для многих характеристик слухового восприятия речевого сигнала: фонематическая категория [30, 42–

44], пол диктора [45], тембр голоса [46], вокальная эмоция [47]. Адаптация к биологически значимым сигналам, которые содержат большой комплекс признаков, оказывается более выраженной по сравнению с адаптацией к одному из них [48]. Таким образом, слуховую адаптацию можно рассматривать как общий принцип обработки акустических сигналов, включая голоса дикторов, адаптивное кодирование которых играет особую роль в слуховом анализе речи.

АДАПТАЦИЯ К РИТМУ. ТЕМПО-РИТМИЧЕСКИЕ ХАРАКТЕРИСТИКИ РЕЧИ

Акустический речевой поток по своим физическим характеристикам очень неоднороден, так как образован не одним излучателем, а сложной системой источников звука разной природы при голосообразовании. В нем присутствуют квазипериодические элементы типа гласных и вокализованных согласных, шумовые сегменты разной длительности, паузы. В соответствии с известными закономерностями первичного слухового анализа сцен последовательность из подобных сегментов должна разделяться на отдельные потоки. Это подтверждается результатами экспериментов, в которых слушатели воспринимали последовательности гласных звуков, дополнительно разделенных встроенными паузами [49]. Слушатели разделяли предъявляемую последовательность на потоки, если гласные принадлежали к разным фонемным классам и имели разную форму спектральной огибающей. Это означает, что гласные звуки в такой ситуации воспринимались как сигналы из разных источников, несмотря на то, что их произносил один и тот же диктор. Предполагается, что единство (слитность) речевых потоков, произнесенных одним диктором (голосом), обеспечивается плавностью спектральных переходов между смежными речевыми звуками – фонемами, и интонационной (просодической) когерентностью [50].

Речевые сигналы являются квазиритмическими, причем интервалы между усилением звуков речи (выделение ударением) довольно изменчивы. Коэффициенты вариации составляют более 30% от среднего интервала между ударениями [51, 52]. Вместе с тем речевые паттерны могут вызывать субъективное впечатление изохронии [53]. Более высокая регулярность обнаруживается в метрической речи, такой как поэзия [54–56], и при групповой речепродукции – молитвах и хоровом пении [57]. Адаптация к ритмам была исследована в диапазоне 100–700 мс между ударами, что примерно соответствует ритмическому диапазону речи. После адаптации к быстрому слуховому ритму, умеренно быстрый тестовый ритм (400 мс между ударами) казался медленным и наоборот [33]. Таким образом, были получены данные о типичном нега-

тивном последствии при восприятии ритма. Величина этого эффекта уменьшалась по мере того, как расхождение ритмического рисунка или паттерна между адаптирующими звуковыми последовательностями и тестовыми последовательностями становилось больше. Результаты последней упомянутой работы подтверждают предположение, что сенсорная временная информация кодируется нейрональными единицами, настроенными на определенные временные интервалы с широкими перекрывающимися кривыми настройки. Последствие ритма не зависело от того, на какое из ушей подавали адаптирующий и тестовый стимулы, а также от того, было это одно или оба уха. Данные факты означают, что информация о ритме звукового сигнала от обоих ушей обрабатывается в слуховых центрах с общим механизмом анализа времени. Информация о пространственно-временных характеристиках кодируемых звуковых сигналов достигает таламуса и мозжечка на самых ранних стадиях слуховой обработки [58, 59]. Мозжечок проецирует важную информацию, закодированную в слуховом сигнале (начало, конец сигнала и резкие изменения уровня), через таламус непосредственно в лобную кору. Полученные в работе [33] экспериментальные данные подтверждают, что описываемые временные свойства формируют основу временного паттерна, обнаруживаемого в обработке простого и сложного ритма. Эта ритмическая основа опирается на ту же нейронную фронто-стриарную архитектуру, что и любая другая временная обработка. Таким образом, адаптация слуха к относительно простым — временным или, иначе, темпо-ритмическим характеристикам речи, демонстрирует сложную функциональную организацию процесса адаптации.

АДАПТАЦИЯ К НЕЛИНГВИСТИЧЕСКИМ ХАРАКТЕРИСТИКАМ ГОЛОСА — ПОЛУ ДИКТОРА И ЕГО ЭМОЦИОНАЛЬНОМУ СОСТОЯНИЮ

Темпо-ритмическая характеристика речи отчасти отражает эмоциональное состояние и может свидетельствовать о нарушении центральных механизмов формирования речи и эмоций [60]. Наибольшее значение для оценки нелингвистических составляющих речи диктора и, в особенности, его голоса имеют спектральные характеристики. К последним относятся частота основного тона голоса диктора, ее модуляция и тембр. Показано, что адаптация к нелингвистической информации голоса также вызывает слуховое последствие.

Так, в работе [45] были получены первые экспериментальные доказательства слуховой адаптации к качеству голоса, а точнее, к половой принадлежности диктора. Стимулы представляли собой высококачественные аудиозаписи голосов пяти женщин и пяти мужчин — носителей немецкого языка

(в возрасте 20–27 лет). Дикторы произносили четыре триграммы, содержащие гласный — согласный — гласный звуки (ГСГ), — /aba/, /aga/, /ibi/ и /igi/. Аудиозаписи составили пять пар женских и мужских голосов, которые были модифицированы с помощью особого алгоритма обработки голосовых характеристик — auditory morphing [61]. В итоге для каждой из пяти пар были созданы семь стимулов, в которых пропорции женского/мужского голоса соответствовали: 80/20%, 70/30%, 60/40%, 50/50%, 40/60%, 30/70% и 20/80%. Было показано, что предварительная адаптация к мужским голосам по контрасту определяет восприятие последующих голосов как более женских и наоборот. При использовании в качестве адаптирующих сигналов синусоидальных тонов с частотой, соответствующей частоте основного тона мужского или женского голоса, подобного последствия не наблюдалось. Эффект отсутствовал и в ситуациях кроссмодальной адаптации, когда предъявлялись фотографии мужских и женских лиц или текстовые варианты имен соответствующего рода. Предполагалось, что это связано с использованием статических (без артикуляции) изображений, нивелирующих проявления аудио-визуального взаимодействия (эффект Мак-Гурка). В целом результаты исключали объяснения слухового последствия голоса через адаптацию только к высоте звука и демонстрировали влияние процессов адаптации на голосовые репрезентации высокого уровня, аналогично контрастным эффектам при распознавании лиц. Области коры больших полушарий мозга, в которых реализуется нейрофизиологический механизм адаптации к гендерной принадлежности диктора, по предположениям Belin и Zatorre [62], а также Lattner и соавт. [63], находится в слуховой ассоциативной коре перед извилиной Гешля и/или в верхней височной борозде правого полушария. Таким образом, адаптация может рассматриваться как общее свойство высокоуровневого сенсорного кодирования при восприятии нелингвистической социально значимой информации от изображений лиц до голосовых характеристик дикторов. При этом адаптационные эффекты к характеристикам голоса зарегистрированы не только в отношении определения пола [45, 64], но и при оценке вокального возраста диктора [65].

Другим важным видом экстралингвистической информации, содержащейся в голосе, является эмоциональное состояние диктора. Восприятие эмоциональных состояний имеет решающее значение для адекватного социального взаимодействия. Известно, что эмоции отражаются в лице, но также и в голосе [66], жесте [67], движении тела [67–69]. В работе [70] впервые была показана слуховая адаптация при восприятии вокального аффекта. В этом исследовании адаптация к гневной вокализации (одиночные/а/-гласные) приводила к восприятию эмоционально неоднозначных голо-

сов (трансформированных в “гнев—испуг” континуум) как более испуганных и наоборот. В другом эксперименте этой же работы аналогичное последствие обнаружено для естественных и синтезированных адаптирующих голосов. Полученные эффекты интерпретировались авторами как указывающие на то, что эмоциональное последствие не является исключительно следствием низкоуровневой адаптации, а скорее может зависеть от более высокого уровня восприятия аффективной категории адаптирующего стимула. Негативное последствие адаптации к счастливым или гневным голосам приводило к тому, что тестовые голоса, трансформированные в континуум “счастье—гнев”, воспринимались как более счастливые после предшествующей адаптации к сердитым (гневным) голосам и наоборот [47]. Эти результаты демонстрируют контрастные механизмы при анализе аффективной информации и расширяют данные работы [70], где при адаптации к эмоциональной составляющей были выявлены гендерно-специфические различия. Эффект гендерной специфичности при адаптации к вокальным эмоциям особенно четко проявился для адаптирующих женских голосов. Для женских тестовых голосов адаптационные эффекты были более выражены на эмоционально-неоднозначных морфологических (стимульных) уровнях, в то время как для мужских тестовых голосов адаптационные эффекты были одинаковыми для всего континуума эмоций. Аналогичные гендерно-специфические эффекты были выявлены и при изучении последствия в условиях восприятия возраста диктора по голосу [71].

В одной из последних работ на эту тему [72] была предпринята попытка оценить относительную роль основной частоты (F0) и тембра в адаптации к вокальным эмоциям гнева и страха. В экспериментах применяли адаптацию к голосам, содержащим информацию об эмоциях либо в F0, либо в тембре. Остальные параметры оставались постоянными и не имели однозначно выраженной эмоциональной составляющей, что соответствовало пропорции 50%/50% согласно отмеченному ранее алгоритму (auditory morphing). Голоса с полностью выраженной эмоцией (100%/0%) и неоднозначные голоса использовали в качестве эталонных условий. Все адаптирующие стимулы были сформированы голосами дикторов либо того же пола, что и пол диктора в тестовых стимулах (Эксперимент 1), либо противоположного пола (Эксперимент 2). В Эксперименте 1 обнаружили устойчивое последствие при всех условиях адаптации. Однако последствие при адаптации к тембру было выражено намного больше, чем при адаптации к частоте основного тона голоса, и лишь незначительно меньше, чем в случае адаптации к исходному (не трансформированному) голосу. В Эксперименте 2 величина последствия была значительно меньше, а различия в величине последствия между адаптирующими

сигналами с разными типами обработки оказались незначимыми. Данные показывают, что в адаптации вокальных эмоций тембр играет большую роль, чем частота основного тона голоса, а также, что адаптация к вокальным эмоциям нарушается из-за устранения гендерного соответствия между адаптирующими сигналами и целевыми (тестовыми) стимулами. Эти результаты дополняют растущее количество свидетельств, указывающих на важную роль тембра в слуховой адаптации к речевому сигналу.

ПРОТОТИПНОЕ КОДИРОВАНИЕ И АДАПТАЦИЯ К ГОЛОСУ

При исследовании адаптации к голосам конкретных дикторов были получены экспериментальные данные, подтверждающие прототипное кодирование голосовой идентичности [73]. Оно состоит в следующем: индивидуальный голос с характерным для него тембром кодируется в многомерном перцептивном пространстве голосовых характеристик, и его идентичность определяется положением в этом пространстве [71, 74]. Два измерения перцептивного пространства, объясняющие большую часть дисперсии в суждениях о несхожестве, хорошо аппроксимируются параметрами частоты основного тона голоса (F0) и формантными частотами [75]. Все пространство формируется относительно среднего голоса, который может выступать в качестве эталона (прототипа) для кодирования других голосовых идентичностей. Таким образом, прототип определяется как голос, относительно которого оцениваются все остальные. Он считается центром массы многомерного голосового пространства и может быть аппроксимирован экспериментально как трансформация на основе усреднения 16 мужских голосов [74, 75]. Важно отметить, что результаты моделирования пространства получены в упрощенном контексте восприятия голосовой идентичности из кратких гласных. Хотя этот контекст аналогичен упрощенному контексту восприятия лица из статических изображений лиц и дает сопоставимые выводы, их нельзя обобщить на более сложные, естественные ситуации. Как правило, другие речевые характеристики, выявляющиеся в течение более длительного времени, также способствуют распознаванию диктора, например, конкретные фонетические шаблоны, диалект, интонация, скорость речи и т.д.

Формирование индивидуальных особенностей речи в процессе раннего онтогенеза и дальнейшей жизнедеятельности также можно отнести к эффектам длительной речевой адаптации с прототипической настройкой в восприятии голосов [76–80]. К этим процессам относятся закрепление фонемных категорий и интонационных конструкций родной речи, влияние особенностей окружающей речевой среды, формирование психоэмоциональных реак-

ций на определенные характеристики голоса, особенно в отношении голоса матери. Известно, что младенцы изначально способны различать фонетические единицы по всему континууму звуков речи. По мере развития и адаптации к акустическим характеристикам родного языка, у них происходит закрепление его специфических контрастов, в то время как чувствительность к неродной фонетической информации постепенно снижается [76, 77]. Наиболее “сензитивным” периодом для фонетического обучения и формирования основы категориального восприятия речевых звуков является возраст от 6 до 12 месяцев. К 6 месяцам младенцы учатся выделять гласные фонемы родного языка, к 11–12 месяцам — осваивают контрасты согласных звуков [77, 81]. Предполагается, что в этот период развития закладываются нейронные связи и формируется система “native language neural commitment (NLNC)”, облегчающая дальнейшее языковое развитие [77, 79, 81, 82]. По завершении ее формирования освоение нового языка будет проходить без эффектов “облегчения”, т.е. аналогично обучению иностранному языку у взрослых.

В сензитивный период эти процессы пластичны и находятся в зависимости от сенсорного опыта. Это подтверждают исследования, связанные с возможностью “обогащения” акустической среды вокруг младенцев. Показано, что включение иностранной речи, в частности второго языка, ведет к расширению формируемого диапазона фонемных контрастов (категорий) и проявлению эффектов адаптации. Регулярное прослушивание музыкальных сигналов (ритм, высота) улучшает освоение просодических паттернов речи — структуры слогов, словесного ударения, интонационных конструкций, а также фонемного различения для ряда языков с высотной оппозицией речевых единиц, например, китайского мандаринского языка [78, 83]. Обучение звуковым категориям языка достигается, по мнению ряда исследователей, только посредством пассивного слушания (статистическое обучение) и может начинаться уже во внутриутробном периоде [84, 85]. Аргументом в пользу этого предположения выступают, в частности, данные о том, что уже в первый день после рождения младенцы способны выделять голос матери и реагировать на него поворотом головы [80]. В раннем постнатальном периоде фиксируются также и другие реакции — увеличение скорости сосания при появлении новых акустических сигналов и, наоборот, ее снижение до фоновых значений по мере “привыкания” ребенка к их звучанию [81, 84]. Наряду с элементами перцептивной предрасположенности и пассивного обучения, в процессе освоения речевой деятельности включаются и активные процессы — врожденные программы лепета, компоненты имитации воспринимаемых речевых сигналов и их просодических характеристик. Эти процессы базируются на развитии у ребенка меха-

низмов слухо-проприоцептивного и слухо-зрительного взаимодействия. При этом голос матери и адекватная коммуникация в диалоге “мать—дитя” оказываются важнейшими факторами не только языковой и социальной, но и биологической адаптации ребенка [86, 87]. Эффективное воздействие голоса матери на психофизиологический статус младенцев подтверждается результатами его терапевтического использования в клинической практике неонатальных центров [88–92]. Таким образом, голос матери играет особую роль и в освоении образцов родной речи, и в формировании прототипических голосовых характеристик (модели протодиктора).

Сходные процессы адаптации и обучения, связанные с освоением звукового строя языка, интонационных особенностей речи и формированием способности к распознаванию голосов дикторов (кодированию голосовой идентичности) на основе направленной тренировки, наблюдаются и при реабилитации ранооглохших пациентов после операции кохлеарной имплантации [93–99], а также в период адаптации к новым условиям “слушания” у постлингвальных пациентов с кохлеарными имплантами и у начинающих пользователей слуховых аппаратов, имеющих выраженную тугоухость [96, 100–103]. В этих условиях формируются или корректируются исходные протообразы как речевых единиц, так и голосовых характеристик [104]. При этом процесс адаптации затрагивает не только возможности гендерного различения дикторов, но и настройку собственного голоса (развитие слухового контроля) и узнавание голосов близких людей, перцептивные образцы которых находятся в долговременной памяти [96, 100, 105–107].

К перечисленным выше проявлениям долговременной слуховой адаптации близки результаты экспериментов по изучению эффектов адаптации к антиголосам (голосам, отличающимся от прототипа в противоположном направлении относительно конкретных). Эти эффекты можно считать подтверждением прототипного представления идентичности голоса [45, 61, 73]. Ограничение экспериментов по долговременной адаптации состоит в том, что в них использовали небольшие наборы голосов или простые (гласные) стимулы. Для рассмотрения прототипного кодирования идентичности голоса как единого принципа обработки речевой информации желательно получить подтверждение при применении больших наборов знакомых дикторов и произвольных образцов речи.

ПРИМЕНЕНИЕ АДАПТАЦИИ К ИЗУЧЕНИЮ НЕЙРОФИЗИОЛОГИЧЕСКИХ МЕХАНИЗМОВ ФОНЕМАТИЧЕСКОГО АНАЛИЗА

Начальный анализ речевого сигнала основан на сегментации потока речи и выделении его звуковых единиц (фонем). Для доказательства действия

нейрофизиологических механизмов — детекторов признаков, которые опосредуют восприятие речи и отдельных речевых звуков, рассматривали результаты избирательной адаптации и различения стимулов вблизи фонемных границ [108]. В частности, чтобы выяснить, могут ли эффекты границы фонемы и ее сдвига после адаптации быть отнесены к изменениям предвзятости ответа (*response bias*) или к явлениям слуховой адаптации, была использована модель теории обнаружения. Эта модель была применена при анализе фонемной идентификации разных наборов стимулов (/ap—ab—/, /bd/, /bp/) до и после адаптации. Результаты показали, что изменения в идентификации фонемы после адаптации могут быть вызваны сдвигами в критериях оценки, а не изменениями на сенсорном уровне.

Влияние избирательной адаптации на восприятие контрастов языковых признаков исследовалось также в работе [109]. Тестовый стимул состоял из серии синтетических речевых звуков — 13 слогов согласный-гласный, которые варьировали по частоте основного тона и по направлению переходов второй и третьей формант. Изменения в формантных переходах были достаточными для восприятия разных согласных в слогах [b æ], [d æ] и [g æ]. Установлено, что значительный адаптационный эффект был получен как в задаче фонемной идентификации, так и в задаче различения согласных. Использование в качестве адаптирующего сигнала слога [bi], у которого переходы второй и третьей формант отличались от тестовых стимулов, позволило предположить, что последствие наблюдается на уровне фонетического анализа признака, а не отдельных акустических составляющих.

Сходные выводы были сделаны и при исследовании эффекта избирательной адаптации на материале русских взрывных согласных [30]. В работе сравнивали результаты адаптации к неречевым и речевым стимулам с близкими акустическими характеристиками — спектральный состав, изменения амплитудной огибающей (двухтоновый комплекс и фонемный ряд [a] — [ta] — [da]). Для неречевых звуков адаптацию выполняли путем многократного предъявления стимула с максимальным значением длительности между началом сигнала и моментом приращения интенсивности, для речевых — соответственно слогом [da]. При восприятии неречевых стимулов изменений в границе обнаружения амплитудной неравномерности после адаптации выявлено не было. При восприятии речевых стимулов граница между гласным и глухим согласным [ta] оставалась неизменной, а для слогов с глухим и звонким взрывным согласным значимо смещалась по принципу контраста — расширение категории [ta] и сокращение категории [da].

Результаты другой группы экспериментов по избирательной адаптации с использованием взрывных согласных были интерпретированы с точки зрения изменения состояния детекторов слуховых и фонетических признаков, а также контраста откликов [110]. В этих исследованиях использовали процедуры избирательной адаптации и парных сравнений между последовательно предъявляемыми стимулами — слогами с сочетанием разных категорий согласных. Один из стимулов представлял звукосочетание [spa], за которым следовал интервал тишины (75 мс) и слог [ba] (время начала голоса 10 мс). При предъявлении он воспринимался как слог [spa] с глухим согласным [p] (без голоса), несмотря на то, что спектральная структура ее конечного элемента сочетания “взрывной—гласный” внутри него была идентична слогу, оцениваемому как [ba] в тестовой серии [ba]-стимулу [pha]. Как адаптирующие стимулы, конечные слоги [spa] и [ba] имели одинаковые эффекты. В процедуре парного сравнения [spa] приводил к оценке неоднозначных тестовых стимулов как “b”, слог [ba] — как “p”. Результаты этих экспериментов показали, что ни контраст ответа, ни обнаружение фонетических признаков не участвовали в избирательных адаптационных эффектах, обнаруженных для ряда “гласный — взрывной согласный”. На основании полученных данных можно утверждать, что избирательные адаптационные эффекты возникают на раннем, слуховом, уровне обработки, который реагирует на спектральное перекрытие между адаптирующими и тестовыми стимулами.

Выводы работы [110] были позднее подтверждены в исследовании по избирательной адаптации к отдельным фонемам, в котором был использован синтезированный согласный-гласный континуум в диапазоне от [ba] до [da] [44]. В экспериментах участвовали дети в возрасте от пяти до шести лет и взрослые. В одном из условий избирательной адаптации внимание слушателей было сосредоточено на адаптационном стимуле [ba], который являлся конечной точкой синтезированного континуума. В другом случае внимание участников эксперимента было сосредоточено не на противоположной конечной точке континуума — [da], а на контралатерально и одновременно прослушиваемых слогах “she” и “see”. Результаты этих экспериментов, по сравнению с двумя более типичными условиями адаптации, показали, что внимание не усиливает избирательные эффекты адаптации. В группе детей данные демонстрировали меньшую величину эффекта при условии сосредоточения внимания на адаптирующем стимуле. В отличие от взрослых дети незначительно изменяли свои ответы после воздействия адаптирующих стимулов [ba]. Этот результат совпадал с данными работы [111]. Вместе с тем дети значительно изменили оценки после воздействия адаптирующего стимула [da]. Таким образом, результаты исследования по различению

стимулов “согласный-гласный” поддержали представление о важности акустической обработки для реализации избирательной адаптации. Совокупность этих данных подтверждает гипотезу о различиях в сенсорной обработке у нормально развивающихся детей по сравнению со взрослыми и демонстрирует, что способности к различению фонем связаны с навыками восприятия речи, т.е. удержания внимания на речевом стимуле.

Для изучения адаптации речевых континуумов в переходах слогов [ba]–[wa] и [cha]–[sha] были проведены три эксперимента с использованием неречевых стимулов (музыка и шум) [112]. Адаптирующие стимулы вызывали значительные сдвиги границ фонем на речевых континуумах только тогда, когда они совпадали по периодичности: музыкальные стимулы влияли на восприятие [ba]–[wa], а шумовые стимулы – на [cha]–[sha]. Причем такие эффекты возникали даже тогда, когда адаптирующие стимулы и тестовые континуумы не совпадали в других простых акустических признаках (время нарастания или длительность согласных). Спектральное перекрытие адаптирующих и тестовых стимулов также оказалось несущественным для результата адаптации. Эти данные подтверждают существование слуховой обработки избирательной к сложным акустическим сигналам, а также нейронно-детекторов, которые реагируют на более сложные признаки речи, характеризующие фонемы. Заметим, что в данном случае наблюдалась асимметричная адаптация, которая свидетельствовала против оппонентного процесса функционирования таких групп нейронов, т.е. против формирования контраста.

На основе данных, полученных в рассмотренных выше работах по избирательной адаптации к различным фонемным конструкциям, была предложена двухуровневая акустическая модель процесса восприятия речи. Ее первый уровень отражает процессы слухового анализа речевого сигнала, на основе которого производится преобразование следующего уровня и переход к более абстрактным признакам и их дальнейшей фонетической и лексической интерпретации. Однако пока не существует единой согласованной модели восприятия речевых сигналов. Разнообразие предложенных разными авторами моделей объясняется сложностью процесса и вовлечением в него не только процедур слуховой обработки, но и сенсорно-моторных программ речепродукции (артикуляция, просодика), а также когнитивных процессов – язык, память, мышление. Модели восприятия речи представлены моторной теорией [113], теориями “следов” [114], категориального восприятия [115], “двойного потока” [116] и другими.

Помимо выделения различных аспектов речевой функции, модельные подходы можно разделить на “пассивные”, т.е. ориентированные в ос-

новном на процессы восприятия, и “активные” – предполагающие взаимосвязь восприятия речи и речепродукции. Другой принцип разделения выбран по отношению к включению в работу моделей как восходящего, так и нисходящего путей обработки. Здесь можно выделить теории “следов” и “двойного потока”, допускающих реализацию стратегий влияния и “снизу–вверх” и “сверху–вниз” [117]. При этом вопросы моделирования процессов нормализации речевого сигнала и его адаптации к дикторской вариативности остаются пока не решенными [118].

НЕЙРОФИЗИОЛОГИЧЕСКИЕ МЕХАНИЗМЫ СЛУХОВОЙ АДАПТАЦИИ К ХАРАКТЕРИСТИКАМ РЕЧИ

В основе кодирования ключевых признаков речи и избирательного реагирования на определенный информационный поток (речь конкретного диктора) лежат нейрофизиологические механизмы, реализующиеся на разных уровнях слуховой системы – от стволовых структур до слуховых и мультисенсорных областей коры больших полушарий. Таким образом, в анализе акустической сцены задействована обширная нейронная сеть, которая реализует функции сегрегации речевых потоков. Инвариантное по отношению к шуму представление речи было показано в слуховой коре человека [6, 119, 120]. Вместе с тем пространственно-временное разрешение неинвазивных методов нейровизуализации ограничено, поэтому данные методы не позволяют четко выявить кодирование речевых характеристик у людей в шуме. Другой особенностью исследований нейронного представления речевого сигнала или вокализаций животных в шуме является использование постоянных фоновых шумов. Эти экспериментальные ограничения привели к тому, что наблюдали только результат адаптации и свойства нейронного представления речи после постоянного шумового воздействия. Остается неясным, каким образом, в какие временные периоды и в каких областях коры больших полушарий адаптация формируется, когда новый фоновый шум внезапно появляется в акустической сцене или, наоборот, исчезает из нее. По этой причине многие важные вопросы, касающиеся динамических свойств адаптации к речи на фоне шума в слуховой коре человека, остаются без ответа. Вместе с тем ответы на них имеют решающее значение для создания полной динамической модели обработки речи в слуховой коре человека.

В работе [121] были объединены методы инвазивной электрофизиологии и психоакустические эксперименты, что позволило исследовать динамические механизмы обработки речи на фоне шума в слуховой коре человека. Нейрональная запись производилась с помощью глубинных и поверхностных электродов высокого разрешения, им-

плантированных в слуховую кору нейрохирургических больных. На основе данных электрокортикографии, полученных при участии шести человек, анализировали нейрональную активность в гамма-диапазоне (75–150 Гц). Усредненная нейрональная реакция показала кратковременный переходный пик со средней длительностью 670 мс, который возникал сразу после изменения фона. Эта кратковременная реакция проявлялась во всех четырех условиях эксперимента (речь на фоне трех типов шума и контроль в тишине), даже при переходе от речи в шуме к речи в тишине. Наблюдали кратковременное и значительное снижение соотношения сигнал – шум (С/Ш) для представления речевого сигнала в слуховой коре в начальном процессе адаптации. По завершении адаптации к шуму соотношение С/Ш увеличивалось. Таким образом, когда новый фоновый шум появлялся на акустической сцене во время восприятия речи, слуховые нейроны мгновенно реагировали на его особенность, но затем быстро адаптировались, что приводило к подавлению нейронного кодирования шума и усилению нейронального кодирования речи, а также улучшению восприятия ее фонетических особенностей и разборчивости. Таким образом, методический подход, использованный в работе [121], позволил выявить специфические особенности нейронального кодирования речи и фонового шума во время развития адаптационного процесса. Было обнаружено, что величина эффектов адаптации к различным фоновым шумам различна в разных нейрональных областях. Данный результат можно рассматривать как вполне предсказуемый, учитывая особенности спектрально-временных характеристик настройки этих областей.

В дополнение к спектральному перекрытию выделение слухового объекта из фонового шума, имеющего временную структуру, требует интеграции сигнала во времени [122, 123]. Реконструктивный анализ нейрональной активности, выполненный в работе [121], выявил кодирование спектрально-временных особенностей стимула в реакциях нейронных популяций. Используя высокую гамма-активность, авторы работы исследовали относительные расстояния между нейронными представлениями фонем во время и после адаптации. Было обнаружено, что относительное расстояние фонем в реакциях нейронных популяций значительно уменьшается в начальный период адаптации, но затем существенно увеличивается. Различимость разных реконструированных фонетических признаков также снижается в начальный период адаптации к новому фоновому состоянию и далее увеличивается.

В психоакустических экспериментах распознавание фонем сходным образом ухудшалось в начальный период адаптации к новому типу шума [121]. При этом величина эффекта адаптации к фоновым шумам была больше при прослушивании

речи в шуме по сравнению с прослушиванием только шумовой помехи. Это означает, что наличие речевого сигнала было необходимым условием для наблюдаемого подавления реакций нейронных популяций на шум. Наблюдалось также снижение точности в восприятии фонем при переходе от прослушивания шумного фона к тишине. Эти проявления хорошо согласуются с психофизическими исследованиями прямой маскировки, где обнаружение целевого сигнала может быть нарушено предшествующим звуком [124], особенно когда акустические характеристики шума и цели перекрываются [125]. При этом повышается роль адаптационных процессов в усилении контраста сигнала и фона, а также в улучшении условий для выделения и распознавания признаков речи целевого диктора. Таким образом, адаптационные процессы играют важную роль в восприятии речевого сигнала при различных условиях, что приводит к улучшению его разборчивости и определения дикторской принадлежности.

ЗАКЛЮЧЕНИЕ

Слуховая адаптация – это полезный механизм, позволяющий гибко и быстро изменять чувствительность слуха к внешним акустическим раздражителям в соответствии с меняющейся окружающей средой. Универсальность процессов адаптации по отношению к различным динамическим характеристикам акустической сцены указывает на эффективность такого кратковременного подстраивания к параметрам информации о текущих событиях, что одновременно обеспечивает увеличение чувствительности к новым акустическим событиям [126]. Процессы адаптации развиваются уже в волокнах слухового нерва и проявляются по отношению к частоте при сохранении временной информации (начало, продолжительность звуков), при этом волокна с фазовой синхронизацией демонстрируют более слабую адаптацию [127]. Баланс адаптации к текущей слуховой информации и ее синхронизация слева и справа обеспечивают возможность обработки в ядрах ствола мозга таких акустических характеристик, как местоположение звука. На более высоких уровнях анализа проявляется адаптация к более сложным типам сигналов, которая может способствовать дальнейшей обработке информационных потоков. Она позволяет нейрональным популяциям выделять комплексные характеристики слуховой среды, сложные последовательности сигналов, категориальные отношения и участвовать в таких процессах, как слуховое внимание и разделение акустических потоков. В результате выделение целевого речевого сигнала (целевого диктора) из нестационарного шума может быть обусловлено адаптацией к любому характерному для речи данного диктора признаку. В этом качестве может выступать какая-либо индивидуальная

особенность голоса или темпо-ритмическая организация речевого потока. Явление слухового последствия дает уникальный психофизический инструмент для подтверждения существования специфических детекторов различных признаков и соответствующих нейрональных популяций в слуховых центрах. Этот инструмент важен также для изучения того, как формируются нейросети, связанные с обработкой и выделением голосовых и речевых спектрально-временных характеристик в процессе онтогенеза. Особенности слуховой адаптации к спектрально-временным характеристикам голоса и речи следует учитывать и при организации реабилитационных мероприятий по восстановлению и развитию речевого слуха пациентов после слухопротезирования и, в первую очередь, после кохлеарной имплантации. Кроме того, важно подчеркнуть близкую нейрофизиологическую основу, которую демонстрируют механизмы зрительного восприятия лица собеседника и слуховой идентификации голоса диктора [128, 129].

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено за счет гранта Российского научного фонда (проект № 22-25-00068).

КОНФЛИКТ ИНТЕРЕСОВ

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

ВКЛАД АВТОРОВ

И.Г.А. — разработка концепции работы, анализ данных по проблематике исследования и написание рукописи; Е.А.О. — анализ данных литературы, написание раздела по онтогенетическому развитию речевой функции и редактирование рукописи.

СПИСОК ЛИТЕРАТУРЫ

1. Bronkhorst AW (2015) The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten Percept Psychophys* 77 (5): 1465–1487. <https://doi.org/10.3758/s13414-015-0882-9>
2. Qian Ym, Weng C, Chang Xk, Wang S, Yu D (2018) Past review, current progress, and challenges ahead on the cocktail party problem. *Front Inf Technol Electronic Eng* 19 (1): 40–63. <https://doi.org/10.1631/FITEE.1700814>
3. Andreeva IG (2018) Spatial Selectivity of Hearing in Speech Recognition in Speech-shaped Noise Environment. *Human Physiol* 44 (2): 226–236. <https://doi.org/10.1134/S0362119718020020>
4. Misurelli SM, Litovsky RY (2012) Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *J Acoust Soc Am* 132 (1): 380–391. <https://doi.org/10.1177/1084713808325880>
5. Pérez-González D, Malmierca MS (2014) Adaptation in the auditory system: an overview. *Front Integr Neurosci* 8: 19. <https://doi.org/10.3389/fnint.2014.00019>
6. Auerbach BD, Gritton HJ (2022) Hearing in Complex Environments: Auditory Gain Control, Attention, and Hearing Loss. *Front Neurosci* 16: 799787. <https://doi.org/10.3389/fnins.2022.799787>
7. Licklider JCR (1948) The influence of interaural phase relations upon the masking of speech by white noise. *J Acoust Soc Am* 20: 150. <https://doi.org/10.1121/1.1906358>
8. Durlach NI, Mason CR, Kidd JrG, Arbogast TL, Colburn HS, Shinn-Cunningham BG (2003) Note on informational masking (L). *J Acoust Soc Am* 113 (6): 2984–2987. <https://doi.org/10.1121/1.1570435>
9. Brungart DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am* 109 (3): 1101–1109. <https://doi.org/10.1121/1.1345696>
10. Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) The role of perceived spatial separation in the unmasking of speech. *J Acoust Soc Am* 106 (6): 3578–3588. <https://doi.org/10.1121/1.428211>
11. Chung S, Li X, Nelson SB (2002) Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses *in vivo*. *Neuron* 34 (3): 437–446. [https://doi.org/10.1016/S0896-6273\(02\)00659-1](https://doi.org/10.1016/S0896-6273(02)00659-1)
12. Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8: 1684–1689. <https://doi.org/10.1038/nn1541>
13. Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. *Curr Opin Neurobiol* 17: 423–429. <https://doi.org/10.1016/j.conb.2007.07.001>
14. Robinson BL, McAlpine D (2009) Gain control mechanisms in the auditory pathway. *Curr Opin Neurobiol* 19: 402–407. <https://doi.org/10.1016/j.conb.2009.07.006>
15. Rabinowitz NC, Willmore BDB, King AJ, Schnupp JWH (2013) Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol* 11: e1001710. <https://doi.org/10.1371/journal.pbio.1001710>
16. Mesgarani N, David SV, Fritz JB, Shamma SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. *Proc Natl Acad Sci U S A* 111: 6792–6797. <https://doi.org/10.1073/pnas.131801711>
17. Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci* 10: 1601–1607. <https://doi.org/10.1038/nn2009>
18. Moore RC, Lee T, Theunissen FE (2013) Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. *PLoS Comput Biol* 9: e1002942. <https://doi.org/10.1371/journal.pcbi.1002942>

19. *Schneider DM, Woolley SMN* (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79: 141–152.
<https://doi.org/10.1016/j.neuron.2013.04.038>
20. *Kröger JL, Lutz OHM, Raschke P* (2020) Privacy Implications of Voice and Speech Analysis - Information Disclosure by Inference. In: Friedewald M, Önen M, Lievens E, Krenn S, Fricker S (eds) *Privacy and Identity Management. Data for Better Living: AI and Privacy. IFIP Advances in Information and Communication Technology*. Springer Cham 576: 242–258.
https://doi.org/10.1007/978-3-030-42504-3_16
21. *Jin H, Wang S* (2018) Voice-based determination of physical and emotional characteristics of users U.S. Patent No. 10,096,319. Washington, DC: U.S. Patent and Trademark Office.
<https://www.us.hsb.com/customer-service/voice>
22. *Zwicker E* (1964) ‘Negative afterimage’ in hearing. *J Acoust Soc Am* 36: 2413–2415.
<https://doi.org/10.1121/1.1919373>
23. *Hoke ES, Hoke M, Ross B* (1996) Neurophysiological correlate of the auditory after-image ‘Zwicker tone’. *Audiol Neuro-Otol* 1: 161–174.
<https://doi.org/10.1159/000259196>
24. *Noren AJ, Eggermont JJ* (2003) Neural correlates of an auditory afterimage in primary auditory cortex. *J Assoc Res Otolaryngol* 4: 312–328.
25. *Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Courtenay WE* (2007) The role of auditory cortex in the formation of auditory streams. *Hear Res* 229: 116–131.
<https://doi.org/10.1016/j.heares.2007.01.007>
26. *Snyder JS, Alain C* (2007) Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* 133: 780–799.
<https://doi.org/10.1037/0033-2909.133.5.780>
27. *Hausfeld L, Riecke L, Valente G, Formisano E* (2018) Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *Neuroimage* 181: 617–626.
<https://doi.org/10.1016/j.neuroimage.2018.07.052>
28. *Stilp CE* (2020) Acoustic context effects in speech perception. *Wiley interdisciplinary reviews. Cognit Sci* 11 (1): 1–8.
<https://doi.org/10.1002/wcs.1517>
29. *Rosenblith WA, Miller GA, Egan JP, Hirsh IJ, Thomas GJ* (1947) An auditory afterimage? *Science* 106: 333–335.
<https://doi.org/10.1126/science.106.2754.333>
30. *Огородникова ЕА* (1978) Эффект “селективной адаптации” при восприятии элементарных неречевых стимулов. *Физиол журн СССР* 64 (12): 1803–1808. [*Ogorodnikova EA* (1978) The effect of selective adaptation on perception of elementary non-speech stimuli. *Sechenov Physiol J USSR*. 64 (12): 1803–1807. (In Russ)].
31. *Gutschalk A, Micheyl C, Oxenham AJ* (2008) The pulse-train auditory aftereffect and the perception of rapid amplitude modulations. *J Acoust Soc Am* 123 (2): 935–945.
<https://doi.org/10.1121/1.2828057>
32. *Shima S, Murai Y, Hashimoto Y, Yotsumoto Y* (2016) Duration Adaptation Occurs Across the Sub- and Supra-Second Systems. *Front Psychol* 7: 114.
<https://doi.org/10.3389/fpsyg.2016.00114>
33. *Becker MW, Ian P, Rasmussen IP* (2007) The rhythm aftereffect: Support for time sensitive neurons with broad overlapping tuning curves. *Brain and Cognition* 64: 274–281.
<https://doi.org/10.1016/j.bandc.2007.03.009>
34. *Masutomi K, Kashino M* (2013) Frequency-change aftereffect produced by adaptation to real and illusory unidirectional frequency sweeps. *J Acoust Soc Am* 134 (1): EL14–EL18.
35. *Wang N, Oxenham AJ* (2014) Spectral motion contrast as a speech context effect. *J Acoust Soc Am* 136 (3): 1237–1245.
<https://doi.org/10.1121/1.4892771>
36. *Frissen I, Vroomen J, de Gelder B, Bertelson P* (2003) The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychol (Amst)* 113 (3): 315–327.
[https://doi.org/10.1016/S0001-6918\(03\)00043-X](https://doi.org/10.1016/S0001-6918(03)00043-X)
37. *Phillips DP, Hall SE* (2005) Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hear Res* 202 (1–2): 188–199.
<https://doi.org/10.1016/j.heares.2004.11.001>
38. *Malinina ES* (2014) Perception of approaching and withdrawing sound sources following exposure to broadband noise. The effect of spatial domain. *Zh Evol Biokhim Fiziol* 50 (1): 59–68.
39. *Ehrenstein WH* (1978) Direction-specific acoustical aftereffects. *J Acoust Soc Am* 64 (Suppl. 1): S35.
<https://doi.org/10.1121/1.2004165>
40. *Grantham DW, Wightman FL* (1979) Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *J Acoust Soc Am* 65: 1509–1517.
<https://doi.org/10.1121/1.382915>
41. *Andreeva IG, Malinina ES* (2010) Auditory Motion Aftereffects of Approaching and Withdrawing Sound Sources. *Human Physiol* 36 (3): 290–294.
<https://doi.org/10.1134/S0362119710030060>
42. *Eimas PD, Corbit JD* (1973) Selective adaptation of linguistic feature detectors. *Cogn Psychol* 4: 99–109.
[https://doi.org/10.1016/0010-0285\(73\)90006-6](https://doi.org/10.1016/0010-0285(73)90006-6)
43. *Landahl KL, Blumstein SE* (1982) Acoustic invariance and the perception of place of articulation: a selective adaptation study. *J Acoust Soc Am* 71 (5): 1234–1241.
<https://doi.org/10.1121/1.387772>
44. *Sussman JE* (1993) Auditory processing in children’s speech perception: Results of selective adaptation and discrimination tasks. *J Speech Hear Res* 36 (2): 380–395.
<https://doi.org/10.1044/jshr.3602.380>
45. *Schweinberger SR, Casper C, Hauthal N, Kaufmann JM, Kawahara H, Kloth N, Robertson DMC, Simpson AP, Zäske R* (2008) Auditory Adaptation in Voice Perception. *Curr Biol* 18: 684–688.
<https://doi.org/10.1016/j.cub.2008.04.015>
46. *Latinus M, Belin P* (2011) Human voice perception. *Curr Biol* 21: R143–R145.
<https://doi.org/10.1016/j.cub.2010.12.033>

47. Skuk VG, Schweinberger SR (2013) Adaptation Aftereffects in Vocal Emotion Perception Elicited by Expressive Faces and Voices. *PLoS One* 8 (11): e81691. <https://doi.org/10.1371/journal.pone.0081691>
48. Bestelmeyer PEG, Mühl C (2021) Individual differences in voice adaptability are specifically linked to voice perception skill. *Cognition* 210 (1): 104582. <https://doi.org/10.1016/j.cognition.2021.104582>
49. Lublinskaya VV, Ross J (1991) Perception of the temporal structure in speech-like sound sequence. *Proc XIIth Int Congress Phon Sci*: 318–321.
50. Darwin CJ, Bethell-Fox CE (1977) Pitch Continuity and Speech Source Attribution. *J Exp Psychol: Human Perception and Performance* 3 (4): 665–672. <https://doi.org/10.1037/0096-1523.3.4.665>
51. Dauer RM (1983) Stress-timing and syllable-timing reanalyzed. *J Phonet* 11: 51–62. [https://doi.org/10.1016/S0095-4470\(19\)30776-4](https://doi.org/10.1016/S0095-4470(19)30776-4)
52. Lidji P, Palmer C, Peretz I, Morningstar M (2011) Listeners feel the beat: entrainment to English and French speech rhythms. *Psychon Bull Rev* 18: 1035–1041. <https://doi.org/10.3758/s13423-011-0163-0>
53. Lehiste I (1977) Isochrony reconsidered. *J Phonet* 5: 253–263. [https://doi.org/10.1016/S0095-4470\(19\)31139-8](https://doi.org/10.1016/S0095-4470(19)31139-8)
54. Lerdahl F (2001) The sounds of poetry viewed as music. *Ann NY Acad Sci* 930: 337–354. <https://doi.org/10.1111/j.1749-6632.2001.tb05743.x>
55. Obermeier C, Menninghaus W, von Koppenfels M, Raettig T, Schmidt-Kassow M, Otterbein S, Kotz SA (2013) Aesthetic and emotional effects of meter and rhyme in poetry. *Front Psychol* 4: 10. <https://doi.org/10.3389/fpsyg.2013.00010>
56. Obermeier C, Kotz SA, Jessen S, Raettig T, Von Koppenfels M, Menninghaus W (2016) Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cogn Affect Behav Neurosci* 16: 362–373. <https://doi.org/10.3758/s13415-015-0396-x>
57. Cummins F (2009) Rhythmas entrainment: the case of synchronous speech. *J Phonet* 37: 16–28. <https://doi.org/10.1016/j.wocn.2008.08.003>
58. Grahn JA (2009) The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. *Ann NY Acad Sci* 1169: 35–45. <https://doi.org/10.1111/j.1749-6632.2009.04553.x>
59. Nozaradan S, Schwartze M, Obermeier C, Kotz SA (2017) Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex* 95: 156–168. <https://doi.org/10.1016/j.cortex.2017.08.015>
60. Bóna J (2016) Characteristics of pausing in normal, fast and cluttered speech. *Clin Linguist Phon* 30 (11): 888–898. <https://doi.org/10.1080/02699206.2016.1188421>
61. Kawahara H, Matsui H (2003) Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. *Proc 2003 IEEE Int Confer Acoustics, Speech, and Signal Processing (Piscataway, NJ: IEEE)*: 256–259. <https://doi.org/10.1109/ICASSP.2003.1198766>
62. Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14: 2105–2109. <https://doi.org/10.1097/01.wnr.0000091689.94870.85>
63. Lattner S, Meyer ME, Friederici AD (2005) Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp* 24: 11–20. <https://doi.org/10.1002/hbm.20065>
64. Zäske R, Schweinberger SR, Kaufmann JM, Kawahara H (2009) In the ear of the beholder: neural correlates of adaptation to voice gender. *Eur J Neurosci* 30: 527–534. <https://doi.org/10.1111/j.1460-9568.2009.06839.x>
65. Zäske R, Schweinberger SR (2011) You are only as old as you sound: Auditory aftereffects in vocal age perception. *Hear Res* 282: 283–288. <https://doi.org/10.1016/j.heares.2011.06.008>
66. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70: 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
67. Montepare J, Koff E, Zaitchik D, Albert M (1999) The use of body movements and gestures as cues to emotions in younger and older adults. *J Nonverbal Behav* 23: 133–152. <https://doi.org/10.1021/1435526134>
68. Dael N, Mortillaro M, Scherer KR (2012) Emotion expression in body action and posture. *Emotion* 12: 1085–1101. <https://doi.org/10.1037/a0025737>
69. Grandjean D (2020) Brain networks of emotional prosody processing. *Emot Rev* 13(1): 34–43. <https://doi.org/10.1177/1754073919898522>
70. Bestelmeyer PE, Rouger J, DeBruine LM, Belin P (2010) Auditory adaptation in vocal affect perception. *Cognition* 117: 217–223. <https://doi.org/10.1016/j.cognition.2010.08.008>
71. Zäske R, Schweinberger SR, Kawahara H (2010) Voice aftereffects of adaptation to speaker identity. *Hear Res* 268: 38–45. <https://doi.org/10.1016/j.heares.2010.04.011>
72. Nussbaum C, von Eiff CI, Skuk VG, Schweinberger SR (2022) Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency. *Cognition* 219 (7): 104967. <https://doi.org/10.1016/j.cognition.2021.104967>
73. Latinus M, Belin P (2011) Anti-voice adaptation suggests prototype-based coding of voice identity. *Front Psychol* 2: Article 175. <https://doi.org/10.3389/fpsyg.2011.00175>
74. Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z (2010) Neural mechanisms for voice recognition. *Neuroimage* 52 (4): 1528–1540. <https://doi.org/10.1016/j.neuroimage.2010.05.048>
75. Baumann O, Belin P (2010) Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychol Res* 74: 110–120. <https://doi.org/10.1007/s00426-008-0185-z>
76. Kuhl PK (2004) Early language acquisition: cracking the speech code. *Nat Rev Neurosci* 5: 831–843. <https://doi.org/10.1038/nrn1533>
77. Kuhl PK, Stevens E, Hayachi A, Deguchi T, Kiritani S, Iverson P (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12

- months. *Development Sci* 9: F13–F21.
<https://doi.org/10.1111/j.1467-7687.2006.00468.x>
78. Zhao TC, Kuhl P (2016) Effects of enriched auditory experience on infants' speech perception during the first year of life. *Prospects* 46: 235–247.
<https://doi.org/10.1007/s11125-017-9397-6>
79. Bates E, Thal D, Finlay BL, Clancy B (2002) Early language development and its neural correlates. In: F Boller, J Grafman (Series eds), SJ Segalowitz, I Rapin (Vol eds) *Handbook of Neuropsychol.* Amsterdam. 8 (Part II): 109–176.
80. DeCasper AJ, Fife WP (1980) Of human bonding: newborns prefer their mothers' voice. *Science* 208: 1174–1176.
<https://doi.org/10.1126/science.7375928>
81. Minagawa-Kawai Y, Mori K, Naoi N, Kojima S (2006) Neural Attunement Processes in Infants during the Acquisition of a Language-Specific Phonemic Contrast. *J Neurosci* 27 (2): 315–321.
<https://doi.org/10.1523/JNEUROSCI.1984-06.2007>
82. Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosoph Transact Royal Society Biol Sci* 363 (1493): 979–1000.
<https://doi.org/10.1098/rstb.2007.2154>
83. Conboy BT, Kuhl PK (2011) Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Development Sci* 14 (2): 242–248.
<https://doi.org/10.1111/j.1467-7687.2010.00973.x>
84. Crystal D (2005) *The Cambridge Encyclopedia of Language.* Cambridge. CUP.
85. Webb A, Heller H, Benson C, Lahar A (2015) Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proc Natl Acad Sci U S A* 112: 3152–3157.
<https://doi.org/10.1073/pnas.14149241>
86. DeCasper AJ, Spence MJ (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav Development* 9 (2): 133–150.
<https://doi.org/10.1126/science.737592>
87. Lam-Cassettari C, Peter V, Antoniou M (2021) Babies detect when the timing is right: Evidence from event-related potentials to a contingent mother-infant conversation. *Dev Cogn Neurosci* 48: 100923.
<https://doi.org/10.1016/j.dcn.2021.100923>
88. Doheny L, Hurwitz S, Insoft R, Ringer S, Lahav A (2012) Exposure to biological maternal sounds improves cardio-respiratory regulation in extremely preterm infants. *J Matern Fetal Neonatal Med* 25 (9): 1591–1594.
<https://doi.org/10.3109/14767058.2011.648237>
89. Randa K, Lahava A (2014) Maternal sounds elicit lower heart rate in preterm newborns in the first month of life. *Early Hum Devel* 90 (10): 679–683.
<https://doi.org/10.1016/j.earlhumdev.2014.07.016>
90. Chirico G, Cabano R, Villa G, Bigogno A, Ardesi M, Dioni E (2017) Randomised study showed that recorded maternal voices reduced pain in preterm infants undergoing heel lance procedures in a neonatal intensive care unit. *Acta Paediatr* 106 (10): 1564–1568.
<https://doi.org/10.1111/apa.13944>
91. Best K, Bogossian F, New K (2018) Language Exposure of Preterm Infants in the Neonatal Unit: A Systematic Review. *Neonatology* 114: 261–276.
<https://doi.org/10.1159/000489600>
92. Efendi D, Caswini N, Rustina Y, Iskandar ADP (2018) Combination of Mother Therapeutic Touch (MTT) and Maternal Voice Stimulus (MVS) therapies stabilize sleep and physiological function in preterm infants receiving minor invasive procedures. *J Neonat Nursing* 24 (6): 318–324.
<https://doi.org/10.1016/j.jnn.2018.08.001>
93. Evans MK, Deliyski DD (2007) Acoustic voice analysis of prelingually deaf adults before and after cochlear implantation. *J Voice* 21: 669–682.
<https://doi.org/10.1016/j.jvoice.2006.07.005>
94. Ogorodnikova EA, Koroleva IV, Lublinskaja VV, Pak SP, Stoljarova EI, Baljakova AA (2009) Computer in rehabilitation of patients with cochlear implants. Proc 13-th Int Confer "Speech and Computer–SPECOM'2009". Spb. SPIIRAS: 483–486.
95. Королева ИВ, Огородникова ЕА, Пак СП, Левин СВ, Баякова АА, Шапорова АВ (2013) Методические подходы к оценке динамики развития процессов слухоречевого восприятия у детей с кохлеарными имплантами. *Рос оториноларингол* 3: 75–85 [Koroleva IV, Ogorodnikova EA, Pak SP, Levin SV, Balyakova AA, Shaporova AV (2013) Methodological approaches to assessment of the progress of auditory and speech perception in children with cochlear implants. *Russ Otorinolaringol* 3: 75–85. (In Russ)].
96. Beier LO, Pedroso F, Costa-Ferreira MID (2015) Auditory training benefits to the hearing aids users— a systematic review. *Rev CEFAC* 17 (4): 1327–1332.
<https://doi.org/10.1590/1982-0216201517422614>
97. Dettman SJ, Dowell R, Choo D, Arnott W, Abrahams Y, Davis A, Dornan D, Leigh J, Constantinescu G, Cowan R, Briggs RJ (2016) Long-term communication outcomes for children receiving cochlear implants younger than 12 months: a multicenter study. *Otol Neurotol* 37: e82–e95.
<https://doi.org/10.1097/MAO.0000000000000915>
98. Hall ML, Hall WC, Caselli NK (2019) Deaf children need language, not (Just) speech. *First Language* 39 (4): 367–395.
<https://doi.org/10.1177/0142723719834102>
99. Wie OB, Torkildsen JK, Schaubert S, Busch T, Litovsky R (2020) Long-Term Language Development in Children With Early Simultaneous Bilateral Cochlear Implants. *Ear and Hearing* 41 (5): 1294–1305.
<https://doi.org/10.1097/AUD.0000000000000851>
100. Higgins MB, McCleary EA, Carney AE, Schulte L (2003) Longitudinal changes in children's speech and voice physiology after cochlear implantation. *Ear Hear* 24 (1): 48–70.
<https://doi.org/10.1097/01.AUD.0000051846.71105.AF>
101. Miller JD, Watson CS, Dubno JR, Leek MR (2015) Evaluation of Speech-Perception Training for Hearing Aid Users: A Multisite Study in Progress. *Semin Hear* 36 (4):

- 273–283.
<https://doi.org/10.1055/s-0035-1564453>
102. *Beyea JA, McMullen KP, Harris MS, Houston DM, Martin JM, Bolster VA, Adunka OF, Moberly AC* (2016) Cochlear Implants in Adults: Effects of Age and Duration of Deafness on Speech Recognition. *Otology & Neurotology* 37 (9): 1238–1245.
<https://doi.org/10.1097/MAO.0000000000001162>
 103. *Koroleva IV, Ogorodnikova EA* (2019) Modern achievements in cochlear and brainstem auditory implantation. In: *Shelepin Yu, Ogorodnikova E, Solovyev N, Yakimova E* (eds.) *Neural Networks and Neurotechnologies*. SPb. VVM. Chapt 30: 231–249.
 104. *Kovacic D, Balaban E* (2009) Voice gender perception by cochlear implantees. *J Acoust Soc Am* 126 (2): 762–775.
<https://doi.org/10.1121/1.3158855>
 105. *Horga D, Liker M* (2006) Voice and pronunciation of cochlear implant speakers. *Clini. Linguistics & Phonetics* 20 (2–3): 211–217.
<https://doi.org/10.1080/02699200400027015>
 106. *Vongpaisal T, Trehub S, Schellenberg EG, Lieshout P* (2010) Children With Cochlear Implants Recognize Their Mother's Voice. *Ear and Hearing* 31 (4): 555–566.
<https://doi.org/10.1097/AUD.0b013e3181daae5a>
 107. *Coelho AC, Brasolotto AG, Bevilacqua MC, Moret ALM, Bahmad JF* (2016) Hearing performance and voice acoustics of cochlear implanted children. *Braz J Otorhinolaryngol* 82: 70–75.
<https://doi.org/10.1016/j.bjorl.2015.11.002>
 108. *Elman JL* (1979) Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. *J Acoust Soc Am* 65: 190–207.
<https://doi.org/10.1121/1.382235>
 109. *Cooper WE* (1974) Adaptation of phonetic feature analyzers for place of articulation. *J Acoust Soc Am* 56: 617.
<https://doi.org/10.1121/1.1903300>
 110. *Sawusch JR, Jusczyk P* (1981) Adaptation and contrast in the perception of voicing *J Exp Psychol Hum Percept Perform* 7 (2): 408–421.
<https://doi.org/10.1037/0096-1523.7.2.408>
 111. *Sussman JE, Carney AE* (1989) Effects of Transition Length on the Perception of Stop Consonants by Children and Adults. *J Speech, Language, and Hearing Res* 32 (1): 151–160.
<https://doi.org/10.1044/jslr.3201.151>
 112. *Samuel AG, Newport EL* (1979) Adaptation of speech by nonspeech: evidence for complex acoustic cue detectors. *J Exp Psychol: Hum Percept Perform* 5 (3): 563–578.
<https://doi.org/10.1037/h0078136>
 113. *Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M* (1967) Perception of the Speech Code. *Psychol Rev* 74: 431–461. <https://doi.org/10.1037/h0020279>
 114. *McClelland J, Elman J* (1986) The TRACE Model of Speech Perception. *Cognit Psychol* 18: 1–86.
[https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
 115. *Goldstone L* (1994) Influences of categorization on perceptual discrimination. *J Exp Psychol* 123: 178–200.
<https://doi.org/10.1037/0096-3445.123.2.178>
 116. *Hickok G, Poeppel D* (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8 (5): 393–402.
<https://doi.org/10.1038/nrn2113>
 117. *Heald SLM, Nusbaum HC* (2014) Speech perception as an active cognitive process. *Front Systems Neurosci* 8: 35.
<https://doi.org/10.3389/fnsys.2014.00035>
 118. *Wilder RJ* (2018) Investigating Hybrid Models Of Speech Perception. *Publ Accessible Penn Dissertat*: 3202.
<https://repository.upenn.edu/edissertations/3202>
 119. *Ding N, Simon JZ* (2013) Adaptive temporal encoding leads to a background insensitive cortical representation of speech. *J Neurosci* 33: 5728–5735.
<https://doi.org/10.1523/JNEUROSCI.5297-12.2013>
 120. *Kell AJ, McDermott J* (2017) Robustness to real-world background noise increases between primary and non-primary human auditory cortex. *J Acoust Soc Am* 141: 3896.
<https://doi.org/10.1121/1.4988749>
 121. *Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N* (2019) Adaptation of the human auditory cortex to changing background noise. *J Nature Communicat* 10: 2509.
<https://doi.org/10.1038/s41467-019-10611-4>
 122. *Chait M, Poeppel D, Simon JZ* (2005) Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cereb Cortex* 16: 835–848.
<https://doi.org/10.1093/cercor/bhj027>
 123. *Teki S, Grube M, Kumar S, Griffiths TD* (2011) Distinct neural substrates of duration-based and beat-based auditory timing. *J Neurosci* 31: 3805–3812.
<https://doi.org/10.1523/JNEUROSCI.5561-10.2011>
 124. *Oxenham AJ* (2001) Forward masking: Adaptation or integration? *J Acoust Soc Am* 109: 732–741.
<https://doi.org/10.1121/1.1336501>
 125. *Jesteadt W, Bacon SP, Lehman JR* (1982) Forward masking as a function of frequency, masker level, and signal delay. *J Acoust Soc Am* 71: 950–962.
<https://doi.org/10.1121/1.387576>
 126. *Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A* (2014) Neuronal adaptation, novelty detection and regularity encoding in audition. *Front Syst Neurosci* 8.
<https://doi.org/10.3389/fnsys.2014.00111>
 127. *Sumner CJ, Palmer AR* (2012) Auditory nerve fibre responses in the ferret. *Eur J Neurosci* 36(4): 2428–2439.
<https://doi.org/10.1111/j.1460-9568.2012.08151.x>
 128. *Blank H, Anwender A, von Kriegstein K* (2011) Direct structural connections between voice-and face-recognition areas. *J Neurosci* 31 (36): 12906–12915.
<https://doi.org/10.1523/JNEUROSCI.2091-11.2011>
 129. *Von Kriegstein K, Giraud AL* (2006) Implicit multisensory associations influence voice recognition. *PLoS Biology* 4 (10): e326.
<https://doi.org/10.1371/journal.pbio.0040326>

Auditory Adaptation to Speech Signal Characteristics

I. G. Andreeva^{a, #}, and E. A. Ogorodnikova^{a, b}

^a *Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, St. Petersburg, Russia*

^b *Pavlov Institute of Physiology, Russian Academy of Sciences, St. Petersburg, Russia*

[#]*e-mail: ig-andreeva@mail.ru*

The review presents experimental data on auditory adaptation to specific speech parameters — temporal (tempo-rhythmic) and spectral characteristics of the speech signal, including the fundamental frequency, vowel formants, timbre features. The aftereffects of the non-linguistic characteristics of the voice — the gender and age of the speaker, his/her emotional state — are considered. It is shown how adaptation is involved in the separation of competing speech flows and in the mechanisms of auditory attention. The role of auditory adaptation in the development of speech processing and the occurrence of voice prototypes during ontogenesis are discussed. The main models of speech processing and experimental data on phonemic analysis are presented. The results of experimental and model studies indicate that adaptive processes play an important role in enhancing the signal-background contrast and improve signal identification. It is important to consider the auditory adaptation to specific speech parameters when developing rehabilitation procedures for patients with hearing aids and designing technical speech recognition systems.

Keywords: speech perception, auditory adaptation, speech signal, extralinguistic characteristics, the cocktail-party effect, auditory attention