

УДК 612.85

ОПРЕДЕЛЕНИЕ ПОЛА ДИКТОРА ПО ХАРАКТЕРИСТИКАМ ГОЛОСА НА ФОНЕ ШУМА МНОГОГОЛОСИЯ

© 2024 г. О. В. Лабутина, С. П. Пак, Е. А. Огородникова

*Институт физиологии им. И.П. Павлова Российской академии наук,
Санкт-Петербург, 199034, наб. Макарова, д. 6*

E-mail: ogorodnikovaea@infran.ru

Поступила в редакцию 05.02.2024 г.

После доработки 01.03.2024 г.

Принята к публикации 13.03.2024 г.

Психофизическими методами исследовали особенности опознания пола диктора по характеристикам голоса в условиях действия речеподобной помехи и стимуляции через головные телефоны. Использовали набор речевых сигналов и шум многоголосия из экспериментов в свободном звуковом поле – пространственной сцене (Andreeva et al., 2019). В набор входили 8 двусложных слов, произнесенных 4 дикторами: 2 мужских и 2 женских голоса со средней частотой основного тона 117, 139, 208 и 234 Гц. Шум многоголосия представлял результат микширования всех аудиофайлов (8 слов × 4 диктора). Соотношение сигнал/шум составляло 1:1, что субъективно соответствовало максимальному уровню зашумления в пространственной сцене (SNR = –14 дБ). В экспериментах участвовали 42 взрослых испытуемых (от 17 до 57 лет). Дополнительно выделяли 3 возрастных подгруппы: 18.6±1.5 лет ($n = 27$); 28±4.1 лет ($n = 7$); 46±5.4 лет ($n = 8$). Все испытуемые обладали нормальным слухом. Результаты исследования и их сравнения с данными указанной работы подтвердили значимость характеристик голоса для слухового анализа сложных пространственных (свободное звуковое поле) и непространственных (головные телефоны) сцен, а также продемонстрировали роль механизмов маскировки и бинаурального восприятия, в частности высоко-частотного механизма пространственного слуха. Кроме того, обнаружена зависимость перцептивной оценки гендерных характеристик голоса в шуме от возраста испытуемых и пола дикторов (мужской или женский голос). Результаты имеют практическое значение для организации слухоречевого тренинга, ранней диагностики нарушений помехоустойчивости речевого слуха, а также развития помехоустойчивых систем автоматической верификации дикторов и технологий слухопротезирования.

Ключевые слова: восприятие, голос, гендерная особенность, имитация сложной сцены, шум, многоголосие, пространственная акустическая сцена

DOI: 10.31857/S0235009224020041 **EDN:** DDOTRT

ВВЕДЕНИЕ

Изучение механизмов выделения целевого речевого сигнала в сложной акустической среде является одним из базовых направлений исследований в области физиологии слуха и восприятия речи, которое имеет теоретическую и практическую значимость в контексте развития технологий слухопротезирования и искусственного интеллекта (Balling et al., 2021; Lesica et al., 2021; Королева, 2022; Bharathi, Nalina, 2024).

Основной акцент в таких исследованиях сделан на процессах пространственной избирательности, энергетической и информационной маскировки, а также перцептивной группировки речевых потоков, восприятия речи целевого диктора на фоне шумовых помех и в условиях конкуренции речевых сигналов (Cherry, 1953; Bregman,

1990; Darwin et al., 2003; Bronkhost, 2015; Andreeva, 2018; Балякова и др., 2023).

Известно, что в свободном звуковом поле наиболее важными факторами, определяющими успешность выделения речевого сигнала в сложной сцене, выступают близость спектрально-временных характеристик источников речи и шумовой помехи, их пространственное расположение и эффекты бинаурального освобождения от маскировки (Shamma et al., 2011; Gutschalk, Dykstra, 2014; Bronkhost, 2015; Andreeva, 2018).

В условиях стимуляции через головные телефоны, когда бинауральные механизмы освобождения от маскировки не срабатывают, на первый план выходят гендерные и индивидуальные характеристики голоса диктора (основной тон, тембр, особенности произношения), фонемная связность и контекст речевого высказывания,

слуховое селективное внимание (Darvin et al., 2003; Shamma et al., 2011; Moore, 2012; Gutschalk, Dykstra, 2014; Popper, Fay, 2015). Однако есть ряд трудностей в сопоставлении данных, полученных в этих условиях и в свободном поле, поэтому при исследовании влияния фактора голоса представлялась целесообразной попытка сблизить схему и параметры формирования используемых тестовых стимулов и фоновой помехи.

В этом контексте рассматривалась экспериментальная модель, реализованная в исследовании (Andreeva et al., 2019). С ее помощью было показано, что пространственное разнесение (на два метра и более) близких по перцептивным характеристикам источников речи и речеподобной помехи (шум многоголосия) повышает вероятность обнаружения речевой цели в свободном звуковом поле. Дополнительный анализ данных показал, что в этих условиях также наблюдается влияние голосовых характеристик дикторов (частота основного тона, тембр) на результаты выделения речевого сигнала из речеподобного шума (Огородникова и др., 2020).

Целью нашей работы стало исследование слуховой оценки частоты основного тона (ЧОТ) и гендерной принадлежности голоса диктора в условиях, имитирующих сходную коммуникативную сцену при стимуляции через головные телефоны, т. е. без ее пространственной составляющей.

МЕТОДИЧЕСКИЕ УСЛОВИЯ ИССЛЕДОВАНИЯ

Для исследования использовали речевые стимулы и речеподобный шум (шум многоголосия), которые были включены в эксперименты в свободном звуковом поле в работах (Andreeva et al., 2019; Smirnova et al., 2019) при моделировании пространственной коммуникативной сцены.

Речевые стимулы представляли собой восемь двусложных слов: “пОчва”, “стрОкА”, “рубЕж”, “плАта”, “вЫпуск”, “крЕдит”, “рУчка”, “набОр” с близкими частотами употребления (Ляшевская, Шаров, 2009) и основными гласными звуками русского языка в ударных позициях (выделены прописными буквами). Слова произносили четверо дикторов (двое мужчин – М1, М2 и две женщины – Ж1, Ж2) – носителей русского языка без речевых патологий. Средние значения частоты основного тона их голосов составили: 117 (М1), 139 (М2), 208 (Ж1) и 234 (Ж2) Гц.

Речеподобный шум был получен в результате микширования 32 аудиофайлов (8 слов × 4 диктора). Его предъявление создавало у испытуемых

слуховые ощущения, подобные тем, которые возникают у слушателей, находящихся в многолюдной среде – шуме многоголосия.

Рабочее соотношение сигнал/шум составляло 1:1, что было перцептивно близко к варианту стимуляции при SNR = -14 дБ в пространственной сцене. Речевой сигнал располагался в центре отрезка шума, длительность которого составляла 2 с. Стимулы предъявляли диотически через головные телефоны Sennheiser HD 280 PRO с близкими амплитудно-частотными характеристиками левого и правого каналов (0.6–1.0 дБ в диапазоне 40 Гц – 15 кГц). Порядок предъявления стимулов был случайным.

Эксперименты проводили на комфортном уровне стимуляции (60–65 дБ) с помощью специальной компьютерной программы (Королева и др., 2013). Задача испытуемого состояла в определении гендерной принадлежности голоса диктора (мужской или женский голос) на фоне шума многоголосия. Оценивали количество правильных распознаваний, ошибок и время реакции.

В экспериментах приняли участие 42 взрослых испытуемых (10 мужчин и 32 женщины) в возрасте от 17 до 57 лет (средний возраст – 25.4 ± 6.5 года). Дополнительно выделяли 3 возрастных подгруппы: 1 – 18.6 ± 1.5 года ($n = 27$), 2 – 28 ± 4.1 года ($n = 7$), 3 – 46 ± 5.4 года ($n = 8$). Все испытуемые обладали нормальным слухом по данным аудиометрии (пороги слышимости в речевом диапазоне частот) и теста на обнаружение паузы в звуковом сигнале (Musiek, Chermak, 2014). Процедуры измерений с участием испытуемых, проведенные в рамках исследования, соответствовали требованиям этического комитета Института физиологии им. И.П. Павлова РАН и Хельсинкской декларации 1964 г. с ее последующими изменениями.

Для анализа результатов использовали непараметрические критерии Вилкоксона и Манна–Уитни для зависимых и независимых выборок, соответственно.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Согласно усредненным по всей группе испытуемых данным, пол диктора на фоне шума многоголосия правильно распознавался в $69.8 \pm 1.4\%$ случаев, среднее время реакции составляло 3.1 ± 0.1 с. При этом проявились различия в оценке гендерных характеристик голоса. Так, доля успешных испытуемых (более 70% правильных ответов) при выделении женского голоса была существенно меньше, чем при выделении мужского – 45 и 60% соответственно. Усредненный показатель ошибок опознания ($M \pm m$) для женского голоса составил $32 \pm 2.5\%$, для мужского – $27 \pm 1.5\%$.

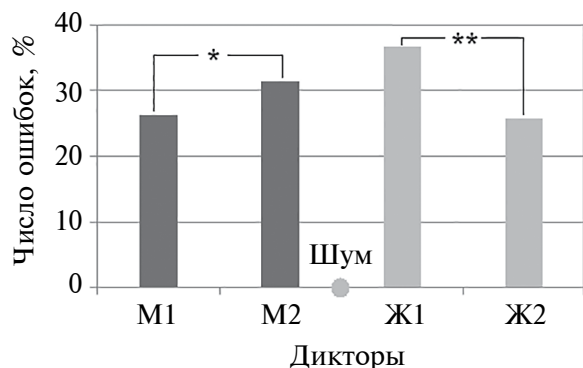


Рис. 1. Распределение ошибок при определении гендерной принадлежности голоса в шуме многоголосия среди дикторов. По горизонтали: условный ряд по показателям ЧОТ для голосов дикторов (М1, М2, Ж1, Ж2) и шума многоголосия. По вертикали: число ошибок в определении пола диктора, %. *, ** – соответственно уровни достоверности различий $p < 0.05$ и $p < 0.01$ по критерию Вилкоксона.

В целом доля ошибок опознания женского голоса от общего числа ошибочных распознаваний в группе составила 55%. Однако эти различия не достигали уровня статистической значимости. Достоверных различий в перцептивной оценке мужских и женских голосов не было выявлено и в отношении времени реакции, а также у испытуемых разного пола.

В то же время значимые различия в восприятии гендерной принадлежности голоса проявились при сопоставлении данных по ряду дикторов (рис. 1).

Можно видеть, что максимальное число ошибок наблюдается для голосов дикторов М2 и Ж1, основной тон которых наиболее близок к диапазону ЧОТ шумовой помехи. Эти данные позволяют говорить о влиянии фактора ЧОТ (условно, высоты голоса), а также о действии механизмов частотной маскировки.

Достоверные различия в восприятии гендерных характеристик голосов дикторов обнаружены

и при сравнении результатов в выделенных возрастных подгруппах (табл. 1).

Результаты сравнения выявили определенную зависимость эффективности слуховой оценки голоса в шуме многоголосия от возраста испытуемых. Наиболее успешными оказались испытуемые средней возрастной группы (группа 2). Их результат по опознанию пола диктора по голосу достоверно отличался и от группы 1 с испытуемыми более младшего возраста ($p < 0.05$) и от более старшей возрастной группы 3 ($p < 0.01$). Такая зависимость, по-видимому, обусловлена периодом оптимального функционирования слуховой системы и когнитивных процессов (внимания), который выделяют в жизненном цикле человека: первый период зрелого возраста – от 20 до 35 лет (Сапогова, 2001; Хухлаева, 2006; Лопотко и др., 2008; Davis et al., 2016).

Свой вклад в эти различия могли внести и возрастные особенности помехоустойчивости речевого слуха, связанные с ухудшением восприятия речи на фоне речевой помехи. Эти особенности могут проявляться достаточно рано (в возрасте около 40 лет) при сохранении других аудиологических показателей в пределах нормы (Fostick et al., 2013).

Другим направлением анализа результатов стало их сопоставление с данными, полученными в условиях пространственной сцены, где использовали тот же набор речевых стимулов и шум многоголосия (Andreeva et al., 2019). Сравнение показало, что, несмотря на различие в задаче испытуемого – обнаружение речевого сигнала vs опознание пола диктора, наблюдалась сходная зависимость правильных ответов от характеристик голоса диктора (рис. 2).

Можно также выделить условный ряд возрастания перцептивной успешности: непространственная сцена – сцена без разделения источников речи и шума в пространстве – сцена с пространственным разделением источников

Таблица 1. Средние показатели правильных ответов и ошибок ($M \pm m$) при определении пола диктора на фоне шума многоголосия в возрастных группах сравнения

Группа	Возраст, лет	Правильные ответы, %	Ошибки для женского голоса, %	Ошибки для мужского голоса, %	Критерий Вилкоксона
1 ($n = 27$)	17–21 (18.6 ± 1.5)	69.1 ± 1.8	31.2 ± 4.3	29.1 ± 4.2	–
2 ($n = 7$)	24–35 (28 ± 4.1)	76.4 ± 2.1	30.4 ± 4.5	16.1 ± 4.0	$p < 0.05$
3 ($n = 8$)	40–57 (46 ± 5.4)	66.3 ± 3.3	39.8 ± 5.6	25.7 ± 3.8	$p < 0.05$

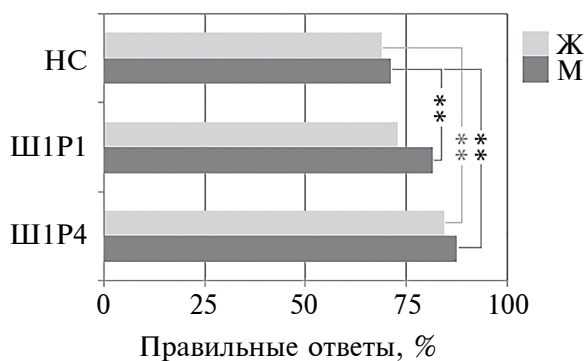


Рис. 2. Показатели правильных ответов при восприятии одного набора речевых стимулов и шума многоголосия в разных условиях. Непространственная сцена (НС): определение пола диктора по голосу при стимуляции через головные телефоны. Пространственные сцены: обнаружение речевого сигнала в свободном звуковом поле при локализации источников речи и шума на расстоянии 1 м от слушателя (Ш1Р1), при их разнесении по удаленности от слушателя – источник шума на 1 м, источник речи на 4 м (Ш1Р4). ** – достоверность различий $p < 0.01$ (критерий Манна–Уитни).

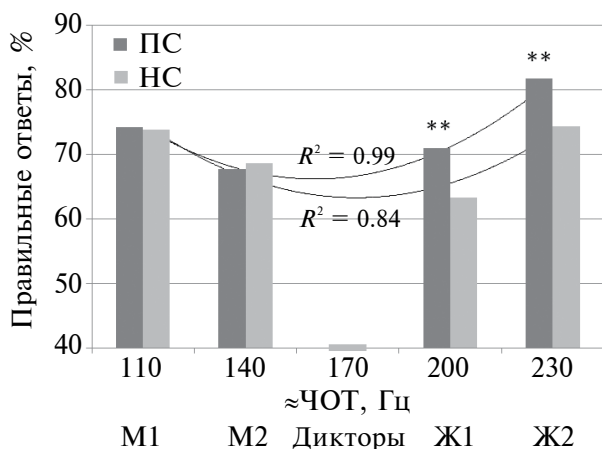


Рис. 3. Распределения правильных ответов (%) по ряду дикторов с разными характеристиками голоса (ЧОТ). НС (непространственная сцена) – данные исследования по определению пола диктора по голосу в шуме многоголосия. ПС (пространственная сцена) – результаты обнаружения речевого сигнала в пространственной сцене при разнесении источников речи и шума (Ш1Р4) и максимальном зашумлении ($SNR = -14$ дБ). ** – уровень значимости различий при восприятии женских голосов ($p < 0.01$, критерий Манна–Уитни).

на 3 м. Ряд хорошо аппроксимируется линейной зависимостью для речевых сигналов, произнесенных как женским голосом ($R^2 = 0.93$), так и мужским ($R^2 = 0.97$). Он также свидетельствует, что помимо фактора ЧОТ, в пространственной сцене проявляется действие базового механизма бинаурального освобождения от маскировки, особенно

выраженное при разнесении источников речи и шума по удаленности ($p < 0.01$).

Аналогичные проявления наблюдали и в случае сравнения данных по ряду дикторов (рис. 3).

Таким образом, полученные данные хорошо согласуются с общими представлениями о том, что основой слуховой оценки пола по голосу выступают тембр (спектральный профиль естественных голосов дикторов разного пола) и диапазон основной частоты – высоты голоса (Pernet, Belin, 2012). При этом условная категориальность восприятия пола (мужской или женский голос), формируемая в процессе сенсорного опыта, может отражаться не только в индивидуальных и возрастных различиях, но и в специфических реакциях слуховой коры головного мозга (Mooge, 2012; Weston et al., 2014). Поэтому важным результатом работы стало сравнение особенностей восприятия гендерных характеристик голоса в разных экспериментальных условиях, что позволило продемонстрировать эффекты, связанные со слуховой оценкой голосов дикторов и механизмами бинаурального слуха, способствующими речевой коммуникации на фоне сильной помехи – шуме многоголосия.

ЗАКЛЮЧЕНИЕ

В рамках исследования подтверждена значимость характеристик голоса (тембр, основной тон) для слухового анализа сложных сцен: пространственных (свободное звуковое поле) и непространственных (стимуляция через головные телефоны). Кроме того, в условиях непространственной сцены получены новые данные о проявлении зависимости эффективности слуховой оценки гендерной принадлежности голоса диктора в шуме многоголосия от возраста испытуемых. Показано, что топ-возрастом в отношении помехоустойчивости восприятия голоса может выступать первый период зрелости человека (от 21 года до 35 лет).

В возрастных подгруппах старше 18–20 лет выявлены также достоверные различия в определении пола говорящего по голосу на фоне шума для дикторов-мужчин и дикторов-женщин (мужской или женский голос). Предполагается, что лучшее выделение мужского голоса при стимуляции через телефоны может определяться энергетической составляющей маскировки в области низких частот, а также начальными проявлениями снижения помехоустойчивости восприятия речи при воздействии речеподобной помехи у более возрастных испытуемых. При этом выигрыш в обнаружении женской речи в шуме в условиях свободного звукового поля может быть

обусловлен включением высокочастотного механизма пространственного слуха.

Полученные результаты создают основу для продолжения исследования и сравнения особенностей восприятия голосовых характеристик в других условиях, в частности в пространственной сцене с нелокализованным источником помехи – интерализованным шумом многоголосия.

Данные имеют практическое значение для организации слухоречевого тренинга, ранней диагностики нарушений помехоустойчивости речевого слуха и центральных слуховых расстройств, а также развития помехоустойчивых систем автоматической верификации дикторов, технологий слухопротезирования и реабилитации пациентов с кохлеарными имплантами.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Работа поддержана средствами федерального бюджета в рамках государственного задания Института физиологии им. И.П. Павлова РАН (№ 1021062411645-5-3.1.8).

КОНФЛИКТ ИНТЕРЕСОВ

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

ВКЛАД АВТОРОВ

Сбор данных литературы, проведение серий экспериментов (О.В. Лабутина), подготовка методики, обработка и анализ результатов (С.П. Пак, Е. А. Огородникова, О. В. Лабутина), подготовка рукописи и иллюстраций (О.В. Лабутина, Е. А. Огородникова, С. П. Пак).

БЛАГОДАРНОСТИ

Авторы выражают благодарность заведующему и сотрудникам лаборатории сравнительной сенсорной физиологии Института эволюционной физиологии и биохимии им. И.М. Сеченова РАН за помощь в проведении исследования.

СПИСОК ЛИТЕРАТУРЫ

Балякова А.А., Лабутина О.В., Медведев И.С., Пак С.П., Огородникова Е.А. Особенности распознавания речевых сигналов в условиях голосовой конкуренции в норме и при нарушениях слухоречевой функции. *Сенсорные системы*. 2023. Т. 37. № 4. С. 342–347. DOI: 10.31857/S0235009223040029

Королева И.В. *Основы аудиологии и слухопротезирования*. СПб: КАРО, 2022. 448 с.

Королева И.В., Огородникова Е.А., Пак С.П., Левин С.В., Балякова А.А., Шапорова А.В. Методические подходы к оценке динамики развития процессов слухоречевого восприятия у детей с кохлеарными имплантами. *Российская оториноларингология*. 2013. № 3. С. 75–85.

Лопотко А.И., Бердникова И.П., Бобошко М.Ю., Журавлева Т.А., Журавский С.Г., Квасова Т.В., Ломоватская Л.Г., Мальцева Н.В., Молчанов А.П., Рындина А.М., Савенко И.В., Слесаренко Н.П., Солдатова Г.Ш. *Практическое руководство по сурдологии*. СПб: Диалог, 2008. 273 с.

Ляшевская О.Н., Шаров С.А. *Частотный словарь современного русского языка (на материалах Национального корпуса русского языка)*. М.: Азбуковник, 2009. 1090 с.

Огородникова Е.А., Лабутина О.В., Андреева И.Г., Гвоздева А.П., Баулин Ю.А. Фактор просодики в восприятии коммуникативной сцены с пространственным разделением источников речи и речеподобной помехи. *Тезисы докладов Международной конференции “Лингвистический форум 2020: Язык и искусственный интеллект”* / Под ред. А.А. Кибрика, В.Ю. Гусева, Д.А. Залманова. М.: Институт языкознания РАН, 2020. С. 127–128.

Сапогова Е.Е. *Психология развития человека*. М.: Аспект пресс, 2001. 460 с.

Хухлаева О.В. *Психология развития. Молодость, зрелость, старость*. М.: Академия, 2006. 208 с.

Andreeva I.G. Spatial selectivity of hearing in speech recognition in speech-shaped noise environment. *Hum. Physiol.* 2018. V. 44(2). P. 226–236. <https://doi.org/10.1134/S0362119718020020>

Andreeva I.G., Dymnikowa M., Gvozdeva A.P., Ogorodnikova E.A., Pak S.P. Spatial separation benefit for speech detection in multi-talker babble-noise with different egocentric distances. *Acta Acustica united with Acustica*. 2019. V. 105. № 3. P. 484–491. <https://doi.org/10.3813/AAA.919330>

Balling L.W., Mølgaard L.L., Townend O., Nielsen J.B.B. The collaboration between hearing aid users and artificial intelligence to optimize sound. *Seminars in Hearing*. 2021. № 42(3). P. 282–294. <https://doi.org/10.1055/s-0041-1735135>

Bharathi R., Nalina H.D. Survey of Recent Advances in Hearing Aid Technologies and Trends. *International Research Journal on Advanced Engineering Hub*. 2024. V. 2. I. 2. P. 303–308. <https://doi.org/10.47392/IRJAEH.2024.0046>

Bregman A.S. *Auditory scene analysis: the perceptual organization of sound*. Cambridge: MIT Press, 1990.

- Bronkhorst A.W. The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception & Psychophysics*. 2015. V. 77(5). P. 1465–1487. <https://doi.org/10.3758/s13414-015-0882-9>.
- Cherry E.C. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 1953. V. 25. № 5. P. 975.
- Darvin C.J., Brungart D.S., Simpson B.D. Effects of fundamental frequency and vocal-tract length changes on attention to one or two simultaneous talkers. *J. Acoust. Soc. Am.* 2003. V. 114. P. 2913–2922.
- Davis A., McMahon C.M., Pichora-Fuller K.M., Russ S., Lin F., Olusanya B.O., Chadha S., Tremblay K.L. Aging and Hearing Health: The Life-course Approach. *Gerontologist*. 2016. № 56 (Suppl 2). P. 256–267. <https://doi.org/10.1093/geront/gnw033>.
- Fostick L., Ben-Artzi E., Babkoff H. Aging and speech perception: beyond hearing threshold and cognitive ability. *J. Basic Clin Physiol Pharmacol*. 2013. № 24(3). P. 175–183. <https://doi.org/10.1515/jbcpp-2013-0048>.
- Gutschalk A., Dykstra A.R. Functional imaging of auditory scene analysis. *Hear. Res.* 2014. V. 307. P. 98.
- Lesica N.A., Mehta N., Manjaly J.G., Deng L., Wilson B.S., Zeng F.-G. Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat. Mach. Intell.* 2021. № 3. P. 840–849. <https://doi.org/10.1038/s42256-021-00394-z>
- Moore B.C.J. *An Introduction to the Psychology of Hearing*. Leiden. Brill., 2012. 442 p.
- Musiek F.E., Chermak G.D. *Handbook of central auditory processing disorder*. San Diego. Plural Publishing, 2014. V. 1. Auditory neuroscience and diagnosis. 768 p.
- Pernet C.R., Belin P. The Role of Pitch and Timbre in Voice Gender Categorization. *Front. Psychol.* 2012. Sec. Perception Science. V. 3. <https://doi.org/10.3389/fpsyg.2012.00023>
- Popper A.N., Fay R.R. (Eds). Perspectives on auditory research. *Springer handbook of auditory research*. 2014. 680 p.
- Shamma S.A., Elhilali M., Micheyl C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 2011. V. 34. P. 114.
- Smirnova V.A., Labutina O.V., Gvozdeva A.P. Chapter 9: Speech detection in spatially distributed speech-like noise. In: *Neural Networks and Neurotechnologies* (eds: Yu. Shelepin, E. Ogorodnikova, N. Solovyev, E. Yakimova). St. Petersburg, VVM, 2019. P. 52–60.
- Weston P., Hunter M.D., Sokhi D.S., Wilkinson I. Discrimination of voice gender in the human auditory cortex. *NeuroImage*. 2014. V. 105. P. 208–214. <https://doi.org/10.1016/j.neuroimage.2014.10.056>

Identification of speaker gender by voice characteristics under background of multi-talker noise

O. V. Labutina, S. P. Pak, E. A. Ogorodnikova

*Pavlov Institute of Physiology, Russian Academy of Sciences,
199034, Makarov emb., 6, St. Petersburg, Russia
e-mail: ogorodnikovaea@infran.ru*

Psychophysical methods were used to study the features of identifying the gender of a speaker based on voice characteristics under conditions of speech-like interference and stimulation through headphones. We used a set of speech signals and multi-talker noise from experiments in a free sound field – a spatial scene (Andreeva et al., 2019). The set included 8 disyllabic words spoken by 4 speakers: 2 male and 2 female voices with average fundamental frequencies of 117, 139, 208 and 234 Hz. Multi-talker noise represented the result of mixing all audio files (8 words * 4 speakers). The signal-to-noise ratio was 1:1, which subjectively corresponded to the maximum noise level in the spatial scene ($SNR = -14$ dB). Adult subjects from 17 to 57 years old ($n = 42$) participated in the experiments. Additionally, 3 age subgroups were identified: 18.6 ± 1.5 years ($n = 27$); 28 ± 4.1 years ($n = 7$); 46 ± 5.4 years ($n = 8$). All subjects had normal hearing. The results of the study and their comparison with the data of mentioned work confirmed the importance of voice characteristics for the auditory analysis of complex spatial (free sound field) and non-spatial (headphones) scenes, and also demonstrated the role of mechanisms of the masking and binaural perception, in particular, the high-frequency mechanism of spatial hearing. A relation the perceptual assessment of the gender by voice in noise and the age of the subjects and the gender of the speakers (male/female voice) was also found. The results are of practical importance for the organization of hearing-speech training, early detection of speech hearing interference immunity impairment, as well as the development of noise-resistant systems for automatic speaker verification and hearing aid technologies.

Keywords: perception, voice, gender feature, imitation of a complex scene, noise, polyphony, spatial acoustic scene

REFERENCES

- Balyakova A.A., Labutina O.V., Medvedev I.S., Pak S.P., Ogorodnikova Ye.A. *Osobennosti raspoznavaniya rechevykh signalov v usloviyakh golosovoy konkurentzii v norme i pri narusheniyakh slukhorechevoy funktsii* [Features of speech signal recognition in conditions of vocal competition with normal hearing and with hearing or speech disorders]. *Sensornyye sistemy*. 2023. V. 37. № 4. P. 342–347. DOI: 10.31857/S0235009223040029.
- Koroleva I.V. *Osnovy audiologii i slukhoprotezirovaniya*. [Fundamentals of audiology and hearing aid]. St. Petersburg: KARO, 2022. 448 p. (in Russian).
- Koroleva I.V., Ogorodnikova E.A., Pak S.P., Levin S.V., Baliakova A.A., Shaporova A.V. *Metodicheskiye podkhody k otsenke dinamiki razvitiya protsessov slukhorechevogo vospriyatiya u detey s kokhlearnymi implantami*. [Methodological approaches to assessing the dynamics of the development of hearing and speech perception processes in children with cochlear implants] *Russian Otorhinolaryngology*. 2013. № 3. P. 75–85. (in Russian).
- Lopotko A.I., Berdnikova I.P., Boboshko M.Yu., Zhuravleva T.A., Zhuravskiy S.G., Kvasova T.V., Lomovatskaya L.G., Mal'tseva N.V., Molchanov A.P., Ryndina A.M., Savenko I.V., Slesarenko N.P., Soldatova G.Sh. *Prakticheskoye rukovodstvo po surdologii* [A practical guide to audiology]. St. Petersburg: Dialog, 2008. 273 p. (in Russian).
- Lyashevskaya O.N., Sharov S.A. *Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka)* [Frequency dictionary of the modern Russian language (based on materials from the National Corpus of the Russian Language)]. Moscow: Azbukovnik, 2009. 1090 p. (in Russian).
- Ogorodnikova Ye.A., Labutina O.V., Andreyeva I.G., Gvozdeva A.P., Baulin Yu.A. *Faktor prosodiki v vospriyatii kommunikativnoy stseny s prostranstvennym razdeleniyem istochnikov rechi i rechepodobnoy pomekhi* [Prosody factor in the perception of a communicative scene with spatial separation of speech sources and speech-like interference]. *Tezisy dokladov Mezhdunarodnoy konferentsii "Lingvisticheskiy forum 2020: Yazyk i iskusstvennyy intellekt" / Pod red. A.A. Kibrika, V. Yu. Guseva, D.A. Zalmanova*. Moscow: Institut yazykoznaniya RAN, 2020. P. 127–128. (in Russian).
- Sapogova Ye.Ye. *Psikhologiya razvitiya cheloveka* [Psychology of human development]. M.: Aspekt press. 2001. 460 p. (in Russian).
- Khukhlayeva O.V. *Psikhologiya razvitiya. Molodost', zrelost', starost'* [Developmental psychology. Youth, maturity, old age]. Moscow: Akademiya, 2006. 208 p. (in Russian).
- Andreeva I.G. Spatial selectivity of hearing in speech recognition in speech-shaped noise environment. *Hum. Physiol.* 2018. V. 44(2). P. 226–236. <https://doi.org/10.1134/S0362119718020020>
- Andreeva I.G., Dymnikova M., Gvozdeva A.P., Ogorodnikova E.A., Pak S.P. Spatial separation benefit for speech detection in multi-talker babble-noise with different egocentric distances. *Acta Acustica united with Acustica*. 2019. V. 105. № 3. P. 484–491. <https://doi.org/10.3813/AAA.919330>
- Balling L.W., Mølgaard L.L., Townend O., Nielsen J.B.B. The collaboration between hearing aid users and artificial intelligence to optimize sound. *Seminars in Hearing*. 2021. № 42(3). P. 282–294. <https://doi.org/10.1055/s-0041-1735135>
- Bharathi R., Nalina H.D. Survey of Recent Advances in Hearing Aid Technologies and Trends. *International Research Journal on Advanced Engineering Hub*. 2024. V. 2. I. 2. P. 303–308. <https://doi.org/10.47392/IRJAEH.2024.0046>
- Bregman A.S. *Auditory scene analysis: the perceptual organization of sound*. Cambridge: MIT Press, 1990.
- Bronkhorst A.W. The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception & Psychophysics*. 2015. V. 77(5). P. 1465–1487. <https://doi.org/10.3758/s13414-015-0882-9>
- Cherry E.C. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 1953. V. 25. № 5. P. 975.
- Darwin C.J., Brungart D.S., Simpson B.D. Effects of fundamental frequency and vocal-tract length changes on attention to one or two simultaneous talkers. *J. Acoust. Soc. Am.* 2003. V. 114. P. 2913–2922.
- Davis A., McMahon C.M., Pichora-Fuller K.M., Russ S., Lin F., Olusanya B.O., Chadha S., Tremblay K.L. Aging and Hearing Health: The Life-course Approach. *Gerontologist*. 2016. № 56 (Suppl 2). P. 256–267. <https://doi.org/10.1093/geront/gnw033>
- Fostick L., Ben-Artzi E., Babkoff H. Aging and speech perception: beyond hearing threshold and cognitive ability. *J. Basic Clin Physiol Pharmacol.* 2013. № 24(3). P. 175–183. <https://doi.org/10.1515/jbcpp-2013-0048>
- Gutschalk A., Dykstra A.R. Functional imaging of auditory scene analysis. *Hear. Res.* 2014. V. 307. P. 98.
- Lesica N.A., Mehta N., Manjaly J.G., Deng L., Wilson B.S., Zeng F.-G. Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat. Mach. Intell.* 2021. № 3. P. 840–849. <https://doi.org/10.1038/s42256-021-00394-z>
- Moore B.C.J. *An Introduction to the Psychology of Hearing*. Leiden. Brill., 2012. 442 p.
- Musiek F.E., Chermak G.D. *Handbook of central auditory processing disorder*. San Diego. Plural Publishing, 2014. V. 1. Auditory neuroscience and diagnosis. 768 p.
- Pernet C.R., Belin P. The Role of Pitch and Timbre in Voice Gender Categorization. *Front. Psychol.* 2012.

- Sec. Perception Science. V. 3. <https://doi.org/10.3389/fpsyg.2012.00023>
- Popper A.N., Fay R.R. (Eds). Perspectives on auditory research. *Springer handbook of auditory research*. 2014. 680 p.
- Shamma S.A., Elhilali M., Micheyl C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci*. 2011. V. 34. P. 114.
- Smirnova V.A., Labutina O.V., Gvozdeva A.P. Chapter 9: Speech detection in spatially distributed speech-like noise. In: *Neural Networks and Neurotechnologies* (eds: Yu. Shelepin, E. Ogorodnikova, N. Solovyev, E. Yakimova). St. Petersburg, VVM, 2019. P. 52–60.
- Weston P., Hunter M.D., Sokhi D.S., Wilkinson I. Discrimination of voice gender in the human auditory cortex. *NeuroImage*. 2014. V. 105. P. 208–214. <https://doi.org/10.1016/j.neuroimage.2014.10.056>