
ЭКСПЕРИМЕНТАЛЬНЫЕ СТАТЬИ

ОБНАРУЖЕНИЕ ЛОМБАРДНОЙ РЕЧИ ДИКТОРОВ РАЗНОГО ПОЛА
ПРИ ПРОСТРАНСТВЕННОМ РАЗДЕЛЕНИИ ИСТОЧНИКОВ ЦЕЛЕВОГО
СИГНАЛА И ШУМА

© 2024 г. И. Г. Андреева^{1, *}, А. М. Луничкин¹, Е. А. Огородникова^{1, 2}

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

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

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

Поступила в редакцию

После доработки

Принята к публикации

Изучали пространственную избирательность слуха к речевым сигналам при разделении целевого сигнала и помехи по расстоянию между их источниками и слушателем. В работе проверяли гипотезу о том, что смещение спектра голоса диктора в сторону высоких частот, которое возникает при говорении в шуме, может улучшить избирательность слуха за счет более активного включения высокочастотного механизма пространственного слуха. Пороги обнаружения целевого сигнала – двухсложного слова, произнесенного мужским или женским голосом, оценивали в задаче двухальтернативного двухинтервального вынужденного выбора в 4 сериях, которые различались типом целевого сигнала (обычная или ломбардная речь) и расположением источников целевого и маскирующего (шум многоголосия) сигналов. Источники сигналов находились на расстоянии 1 и 4 м напротив испытуемого на уровне его головы. Порог обнаружения определяли через пороговое отношение уровней сигнала и шума в месте расположения слушателя (С/Ш). Результаты показали, что порог обнаружения целевого сигнала (мужской и женский голоса диктора совместно) был одинаков для обычной и ломбардной речи и составил -11 дБ С/Ш. Он не зависел от того, какой из источников – целевого сигнала или помехи, находился ближе к слушателю. При обычной речи пороги обнаружения по средним данным различались для мужского и женского голоса, но различие было недостоверным. При ломбардной речи эти пороги достоверно различались: для мужского голоса порог на уровне обнаружения 0.67 был равен -10 дБ, а для женского – -12.5 дБ.

Ключевые слова: речь, голос, эффект Ломбарда, бинауральный слух, пространственная избирательность слуха, расстояние между диктором и слушателем, обнаружение речевого сигнала

DOI: 10.31857/S0869813924020031, EDN: DJSPNE

ВВЕДЕНИЕ

Успешная речевая коммуникация в условиях большой группы людей реализуется благодаря пространственной избирательности слуха. Бинауральные механизмы обеспечивают настройку слухового внимания на определенную область пространства,

в которой находится целевой диктор [1–3]. Слушатель способен различать источники, разделенные по азимутальному углу в семь и более градусов, и при необходимости переключать внимание между ними [4]. Бинауральные механизмы участвуют также и в локализации источника звука по расстоянию при удаленности источника звука от слушателя менее 4–5 м [5, 6]. Показано, что для локализации источника звука на таких расстояниях важны высокочастотные составляющие звукового сигнала [7, 8]. Был обнаружен эффект пространственной избирательности по расстоянию. Он состоит в том, что при одинаковых уровнях интенсивности в месте прослушивания, разделение помехи и целевого сигнала вызывает одинаковое освобождение от маскировки независимо от того, какой из источников – целевого сигнала или помехи – расположен ближе к слушателю [9].

Речь как звуковой сигнал имеет широкий диапазон частот. Вместе с тем основные характеристики голоса дикторов, включающие формантную структуру звуков речи, расположены в более узких областях спектра. Эти области соответствуют рабочему диапазону низкочастотных бинауральных механизмов и границе спектральных диапазонов, в которых реализуются низкочастотные и высокочастотные бинауральные механизмы. Первый диапазон включает частоты основного тона голоса (F_0) у взрослого человека (100–250 Гц) и первую форманту гласных звуков речи (300–1300 Гц). Второй диапазон находится в пределах 2–4 кГц и содержит спектральные области второй и последующих формант гласных звуков. Роль этих механизмов в обнаружении голоса диктора при разделении источников целевого сигнала и помехи по расстоянию практически не исследована [3]. Однако были получены предварительные данные, указывающие на влияние спектральных характеристик голоса диктора на пространственную избирательность слуха по расстоянию при обнаружении его сигнала на фоне речеподобного шума [10].

При обнаружении и распознавании целевого голоса диктора на фоне шума наиболее эффективной помехой является шум толпы, который обладает спектральными и временными характеристиками, близкими к целевому сигналу [11], что вызывает максимальную величину маскирования [12]. Вместе с тем именно такая ситуация – выделение целевого речевого сигнала при пребывании в общественных местах – на улице, в магазине, на вокзале, в аэропорту и т.д. является наиболее экологически и социально значимой.

Пространственная избирательность слуха по отношению к речевым сигналам в мировой научной литературе изучается без учета изменений характеристик голоса при произнесении слов в шуме, так называемого эффекта Ломбарда. В шуме диктор испытывает затруднения в восприятии собственной речи. Такое нарушение работы слуховой обратной связи вызывает непроизвольное усиление голоса, которое создается в результате повышения тонуса мышц, участвующих в голосообразовании. Повышение мышечного тонуса приводит как к повышению уровня голоса, так и к повышению F_0 , частоты основного тона голоса, частот основных спектральных компонентов гласных, сонорных согласных [13, 14] и спектрального центроида мощности (ЦМ) речевого сигнала [15–17].

Смещение спектра голоса диктора в сторону высоких частот может улучшить пространственную избирательность речевого слуха по расстоянию за счет более активного включения высокочастотного механизма пространственного слуха. Целью данной работы являлась проверка этой гипотезы. Поскольку типичные мужские и женские голоса значительно различаются по спектру, то дополнительной задачей исследования было выполнить сравнительный анализ пространственной избирательности для дикторов разного пола, которые имеют типичные гендерные характеристики голоса.

МЕТОДЫ ИССЛЕДОВАНИЯ

В исследовании принимали участие 36 испытуемых (16 мужчин, 20 женщин, возраст 18–35 лет, средний возраст 23 ± 6 года) с нормальным слухом по результатам

тональной пороговой аудиометрии и теста обнаружения паузы [18]. Запись речевых стимулов и их последующее прослушивание выполняли в анэхоидной звукоизолированной камере объемом 62.5 м^3 , ослабление уровня внешних шумов составляло не менее 40 дБ на частотах 0.5–16 кГц. Речевые стимулы представляли собой записи восьми двусложных слов русской речи: «кредит», «набор», «рубезж», «строка», «выпуск», «плата», «почва», «ручка». Эти записи были выполнены при участии двух дикторов – носителей русского языка, которые не входили в число испытуемых, участвовавших в экспериментах по обнаружению речи. Первым диктором был мужчина 30 лет (обычная речь F_0 : 113 ± 4 Гц, ЦМ: 899 Гц; ломбардная речь 130 ± 2 Гц, ЦМ: 1186 Гц), вторым – женщина 28 лет (F_0 : обычная речь 198 ± 16 Гц, ЦМ: 1021 Гц; ломбардная речь 214 ± 17 Гц, ЦМ: 1108 Гц). Речь записывали при надетых головных телефонах в тишине (обычная речь) и при диотической подаче дикторам в наушники шума многоголосия уровнем 60 дБ (ломбардная речь). Для снижения влияния заглушающего эффекта наушников на восприятие дикторами собственного голоса использовали обратную связь: сигнал с микрофона, при помощи которого велась запись голоса, в режиме реального времени подавали диктору в наушники с таким усилением, при котором воспринимаемая им громкость собственного голоса соответствовала таковой без головных телефонов.

В качестве маскирующих сигналов применяли отрезки шума многоголосия продолжительностью 2 с, имевшие линейные фронты нарастания и убывания амплитуды продолжительностью по 0.1 с каждый. Шум многоголосия включал записи речи четырех дикторов (двух мужчин и двух женщин). Он был создан в результате микширования 32 аудиофайлов (4 диктора $\times$ 8 слов) длительностью 10 с, в каждом из которых один и тот же речевой стимул повторялся несколько раз подряд без пауз. Спектральные характеристики шума многоголосия представлены на рис. 1. Амплитудно-частотный спектр имеет характерный для речевого сигнала наклон и по своим характеристикам близок к розовому шуму. На спектрограмме видны фрагменты формантной структуры отдельных гласных звуков, типичные для шума многоголосия и демонстрирующие сходство шума с речевым сигналом по ритмической структуре. При прослушивании этого речевого маскера испытуемый слышал шум многоголосия, когда одновременно говорят много людей. Создание такого шума подробно описано в работе [19]. Из полученного в результате микширования аудиофайла были вырезаны десять случайных фрагментов продолжительностью 2 с.

Для воспроизведения шумовых и речевых стимулов использовали ноутбук с внешней звуковой картой Tascam US-16x08 (44100 Гц, 16 бит) и усилитель мощности Neva Audio SA-3004 с подключенными к нему динамическими громкоговорителями Klipsch R-3800-C. Громкоговорители располагались внутри анэхоидной камеры, на расстоянии 1 и 4 м напротив сидящего испытуемого на уровне его головы. Акустический контроль параметров шумовых и речевых стимулов в месте расположения испытуемого осуществляли при помощи набора калиброванного измерительного оборудования фирмы Brüel & Kjær (микрофон 4145, предусилитель 2639, усилитель 2606) в режиме усреднения среднеквадратичной мощности на интервале времени 1 с. Акустический контроль при записи речи выполняли с помощью комплекта измерительного оборудования для поля звукового давления фирмы CrySound (искусственное ухо CRY318 и анализатор спектра CRY651B).

В первой части исследования 20 испытуемым (10 мужчин и 10 женщин) предлагали выполнить перцептивную задачу обнаружения речевого сигнала в четырех экспериментальных сериях. Пороги обнаружения оценивали методом двухальтернативного двухинтервального вынужденного выбора в 4 сериях, которые различались типом целевого сигнала и расположением источников целевого и маскирующего сигналов. Порядок прослушивания серий для всех испытуемых был одинаковым: 1) обычная речь, конфигурация «речь 1 м – шум 4 м»; 2) обычная речь, «речь 4 м – шум 1 м»; 3) лом-

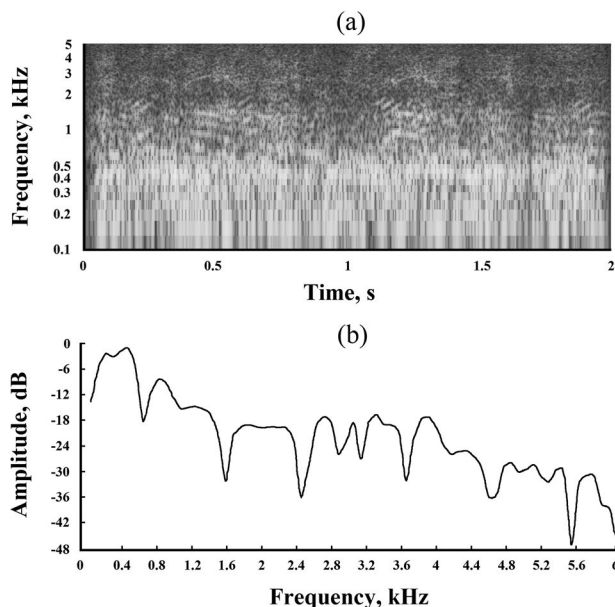


Рис. 1. Спектральные характеристики шума многоголосия. (а) – спектрограмма, (б) – амплитудно-частотный спектр.

бардная речь, «речь 1 м – шум 4 м»; 4) ломбардная речь, «речь 4 м – шум 1 м». Пороги определяли, меняя соотношение сигнал/шум (С/Ш) с шагом в 3 дБ: -8, -11, -14, -17 дБ. Во всех сериях уровень целевого сигнала был равен 60 дБ УЗД в месте прослушивания. В каждой из них количество целевых сигналов составляло 16 (2 диктора х 8 слов) при каждом из отношений С/Ш. В квазислучайных последовательностях серии половину целевых стимулов предъявляли в первом интервале шума, а другую половину – во втором. Рандомизация экспериментальных серий, их воспроизведение и регистрация ответов испытуемых выполнялись с использованием оригинального программного обеспечения, созданного в среде Matlab.

Во второй части экспериментов все 36 человек участвовали в определении порогов обнаружения при определенном положении источников: целевой сигнал прослушивали с расстояния 1 м, а шум многоголосия подавали с расстояния 4 м (серии 1 и 3). Эта часть работы позволила оценить влияние эффекта Ломбарда у дикторов разного пола на пространственную избирательность слуха.

Анализ данных выполняли, вычисляя индивидуальные и среднegrupповые вероятности обнаружения речевых стимулов при разных отношениях С/Ш. Сравнение вероятностей обнаружения речевых стимулов при одинаковых отношениях С/Ш в сериях выполняли с использованием биномиального теста в пакете программ Statistica. Порог обнаружения целевого речевого сигнала оценивали на уровне вероятности обнаружения 0.67.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В первой части исследования оценивали вероятность обнаружения целевого речевого сигнала независимо от пола диктора. По результатам всех экспериментальных серий получили снижение вероятности обнаружения речевого стимула при уменьшении отношения С/Ш до -14 дБ ($p < 0.01$) (рис. 2). Достоверных различий между обнаруже-

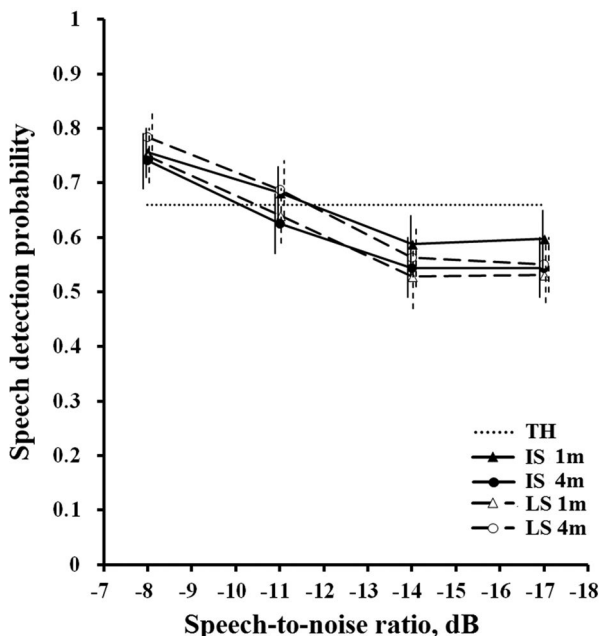


Рис. 2. Обнаружение обычного и ломбардного речевого сигнала при разных положениях источника целевого сигнала и шума многоголосия. По абсциссе – соотношение сигнал/шум (дБ); по ординате – вероятность обнаружения речевого стимула на фоне шума многоголосия; TH – порог обнаружения стимула; IS – стимул с обычной речью; LS – стимул с ломбардной речью; 1 м – источник целевого сигнала расположен на расстоянии 1 м от слушателя; 4 м – источник целевого сигнала расположен на расстоянии 4 м от слушателя. Вертикальные линии – доверительный интервал ($n = 320$).

нием речевого стимула при отношениях С/Ш -14 и -17 дБ не было выявлено ($p > 0.05$), а вероятность обнаружения при этих уровнях зашумления находилась в диапазоне от 0.60 до 0.53 и была ниже пороговой. Сравнение результатов серий с разным положением целевого и шумового источников, а также с обычным и ломбардным целевым речевым сигналом не показало достоверных различий ни при одном из отношений С/Ш. Порог на уровне вероятности обнаружения 0.67 был достигнут при обнаружении стимула во всех сериях при отношении С/Ш равном -11 дБ. Таким образом, пространственная избирательность для целевого сигнала при ломбардной речи оказалась такой же, что и при обычной речи в задаче обнаружения голосов дикторов обоего пола.

Во второй части работы проверяли предположение о том, что с учетом пола диктора ломбардная речь будет обнаруживаться лучше при одинаковых уровнях интенсивности целевых сигналов. Целевой речевой сигнал, источник которого находился на расстоянии 1 м от слушателя, обнаруживали 18 мужчин и 18 женщин на фоне источника шума, удаленного на 4 м от слушателя. Были сопоставлены результаты по обнаружению обычной и ломбардной речи для речевых стимулов, произнесенных мужским и женским голосами. Стимулы, содержащие обычную и ломбардную речь одного и того же диктора мужчины, имели одинаковую вероятность обнаружения при всех отношениях С/Ш ($p > 0.05$) (рис. 3). Такой же результат получили для стимулов, произнесенных женским голосом. Обычная и ломбардная речь диктора женщины обнаруживались с одинаковыми вероятностями при всех отношениях С/Ш ($p > 0.05$). Сравнение обнаружения на фоне шума двусложных слов, произнесенных женским и мужским голосами в обычной речи, не показало достоверных различий ни при одном из соотно-

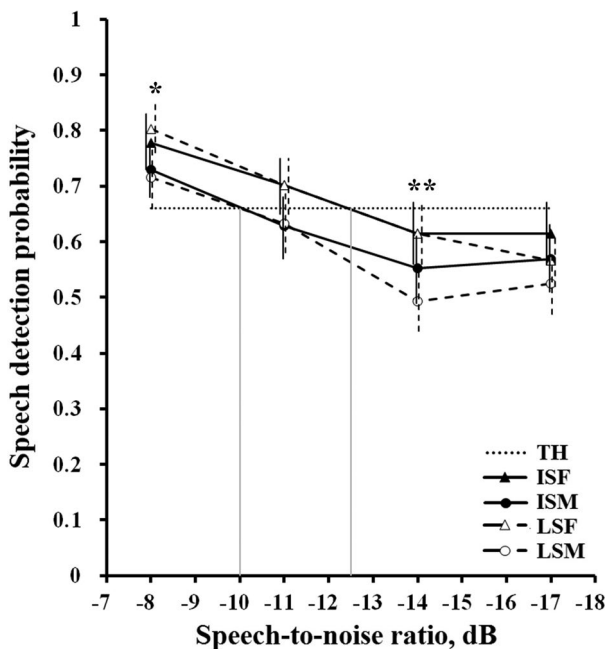


Рис. 3. Обнаружение мужского и женского голоса при обычной и ломбардной речи (источник удален на 1 м от слушателя) на фоне шума многоголосия (источник удален на 4 м от слушателя). По абсциссе – соотношение сигнал/шум (дБ); по ординате – вероятность обнаружения речевого стимула на фоне шума многоголосия; TH – порог обнаружения стимула; ISF – стимул с женской обычной речью; ISM – стимул с мужской обычной речью; LSF – стимул с женской ломбардной речью; LSM – стимул с мужской ломбардной речью. Маркерами показаны достоверные различия вероятности обнаружения женской и мужской ломбардной речи; * – $p < 0.05$, ** – $p < 0.01$ (биномиальный тест, $n = 288$). Вертикальные линии – доверительный интервал ($n = 288$).

шений С/Ш ($p > 0.05$). При С/Ш равном -11 дБ различия выявлялись в виде тенденции ($p = 0.06$).

При ломбардной речи женский голос обнаруживался достоверно лучше мужского при отношении С/Ш = -8 и -14 дБ ($p = 0.014$ и $p = 0.003$ соответственно), а при С/Ш = -11 дБ различия были на уровне тенденции ($p = 0.076$). Таким образом, вблизи порога обнаружения получили тенденцию к лучшему обнаружению двусложных слов, произнесенных женским голосом, причем при ломбардной речи эти различия были статистически достоверны. Этот результат позволил оценить различия в порогах обнаружения мужского и женского голоса. Для мужского голоса порог составил -10 дБ, а для женского – -12.5 дБ.

Таким образом, для обычной речи пороги обнаружения мужского и женского голоса на уровне 0.67 достоверно не различались, а для ломбардной речи со смещенным в сторону высоких частот спектром пороги были различны.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Вероятность обнаружения голоса при пространственном разделении целевого источника и помехи по расстоянию в 3 м была выполнена в диапазоне соотношения С/Ш - 8 – -17 дБ. По сравнению с предыдущим исследованием [9], в котором применяли только обычную речь, дизайн эксперимента был изменен. Этот диапазон был сдвинут на 3 дБ в направлении снижения соотношения С/Ш. Целью такого изменения

было достижение порога обнаружения, поскольку даже при самом низком соотношении С/Ш вероятность распознавания в предыдущем исследовании составляла выше 70%. В результате более высокий уровень неопределенности при обнаружении целевого голоса в измененном диапазоне повлиял на вероятность обнаружения таким образом, что уже для соотношения С/Ш -8 дБ получили значения ниже 0.8 по сравнению со средним значением 0.9, полученным в аналогичной ситуации в предыдущей работе. Поэтому при оценке порога применили критерий 0.67 вероятности обнаружения целевого сигнала. С учетом высокого уровня неопределенности получили такую же величину порога обнаружения целевого сигнала для обычной речи на фоне речеподобного шума (шума многоголосия), что и в предыдущем нашем исследовании. Пороговое соотношение С/Ш составило -11 дБ, а после -14 дБ достоверных изменений вероятности обнаружения не выявлено, т.е. характер ответов испытуемых был случайным.

В представленной работе получили подтверждение полученные ранее данные по положению целевого сигнала и помехи [9]. При равных соотношениях их уровней интенсивности в месте расположения слушателя освобождение от маскировки не зависело от того, какой из источников – целевого сигнала или помехи – находился ближе к слушателю. Достоверных различий в вероятности обнаружения при разном положении целевого сигнала и помехи не было выявлено во всем исследованном диапазоне С/Ш. В упомянутой выше работе освобождение от маскировки при пространственном разделении, по сравнению с пространственным совпадением целевого сигнала и шума многоголосия, составила 2–3 дБ (шум многоголосия). В другом исследовании разница в уровне маскировки при синфазной и противофазной подаче целевого речевого сигнала на фоне непрерывного бинаурального шума равнялась 7 дБ по средним данным в группе испытуемых той же возрастной группы, что и в нашем исследовании [20]. Заметим, что последний результат был получен при максимальном разделении по фазе целевого сигнала и помехи, соответствующем 180° , тогда как освобождение от маскировки в случае пространственного разделения по расстоянию было достигнуто при околопороговых различиях в азимутальных углах, приходящихся на левое и правое ухо – около 2° .

Применение в качестве целевого сигнала ломбардной речи не привело к значимым изменениям обнаружения сигнала по сравнению с обычной речью при совместном рассмотрении данных для мужского и женского голоса. Вместе с тем, как показано в разделе «Методы исследования», F0 при ломбардной речи увеличивается на фоне шума по средним данным на 16–17 Гц, а спектральный центроид – на 87 и 287 Гц. Заметим, что голоса дикторов были подобраны нами определенным образом – разница в F0 составила около 100 Гц и была близка к октавному повышению высоты, что характерно для типичных мужских и женских голосов [21]. Таким образом, увеличение доли высоких частот в ломбардной речи оказывается существенным при его сравнении с гендерными различиями голосов дикторов. Оно менее выражено, чем различие типичного мужского и женского голосов, поэтому раздельное рассмотрение предполагаемого эффекта ломбардной речи на обнаружение мужского и женского голосов было обоснованным.

В случае ломбардной речи, то есть при смещении спектральных областей голосов в высокочастотную область по сравнению с обычной речью, достоверные различия были выявлены между результатами обнаружения мужского и женского голосов. Различия в пороге обнаружения для мужского и женского голосов были определены в 2.5 дБ, тогда как для обычной речи пороги различались по средним данным, но достоверных различий выявлено не было. Полученная разница была сопоставима с различием в уровне маскировки при пространственном разделении целевого сигнала и речеподобного шума, она составляла 2–3 дБ (шум многоголосия) [9]. Полученный результат позволяет считать, что найденное различие в величине порогов для мужского и женского голосов при ломбардной речи с пространственным освобождением от маскировки является следствием смещения в высокочастотную область спектра речи этого типа.

Уровень шума многоголосия, при котором выполняли записи речи дикторов, составлял 60 дБ. Известно, что при более высоких уровнях шума эффект сильнее выражен [13, 22, 23]. Использование уровней около 80 дБ, вероятно, привело бы к более существенным различиям в обнаружении обычной и ломбардной речи, однако продолжительное воздействие шума высокого уровня интенсивности при оценке порогов противоречит этическим требованиям безопасности. В качестве косвенного подтверждения усиления эффекта и отличий от обычной речи можно рассматривать данные, полученные в психоакустических опытах и в модельных экспериментах при обнаружении «кричащей» речи [14]. Однако целью нашего исследования было сравнение вероятностей обнаружения обычной и ломбардной речи при уровне шума многоголосия, более характерного для повседневного общения и не вызывающего перцептивного дискомфорта.

В нашей работе были использованы голоса дикторов, которые имеют типичные гендерные различия F0 вблизи одной октавы. Влияние высоты голоса в околопороговых условиях проявилось при сходной схеме расположения источников обычной речи и шума многоголосия [10]. Эти результаты обнаружения обычной речи показали улучшение и значимые различия ($p < 0.05$) для женского голоса с наиболее высокой средней F0 – 226 ± 12 Гц по сравнению с тремя другими (два мужских и низкий женский голос) в случае локализации источника шума на расстоянии 1 м, а источника речи – 4 м от слушателя и соотношения С/Ш -14 дБ [10].

Полученные нами данные существенно дополняют представления об избирательности механизмов выделения речевого сигнала в шуме. Они могут быть использованы для моделирования процессов распознавания речи в шуме, развития технологий слухопротезирования с целью повышения эффективности выделения целевого речевого сигнала в условиях сложных коммуникативных сцен. Вместе с тем исследования голоса последних десятилетий демонстрируют снижение различий по данному признаку и тенденцию к сближению гендерных диапазонов F0, в частности, понижению женских голосов в среднем на 20–22 Гц [25]. На гендерные проявления эффекта Ломбарда может оказывать влияние и возрастной фактор, который определяет изменения в органах речеобразования и, в первую очередь, в голосовых связках пожилого человека [25, 26]. При этом, как правило, диапазоны высоты голоса также сближаются: у мужчин F0 смещается в сторону повышения, у женщин, наоборот, в сторону понижения [26]. В настоящем исследовании это направление сравнительной оценки не рассматривалось, но оно будет включено в дальнейшее изучение проблемы избирательности речевого слуха. Представляется существенным проанализировать возможные гендерные различия в группе аудиторов, т. к. согласно данным Похилайнен и соавт., они наблюдали значительное преимущество в обнаружении «кричащей» речи у аудиторов-мужчин относительно аудиторов-женщин [24].

БЛАГОДАРНОСТИ

Авторы благодарят А. П. Гвоздеву за подготовку программного обеспечения для выполнения процедуры эксперимента и акустические измерения при подготовке методики.

ВКЛАДЫ АВТОРОВ

А. И. Г. – разработка идеи и дизайна работы, написание рукописи; Л. А. М. – выполнение записей речи, проведение экспериментов, анализ данных, работа над текстом рукописи; Е. А. О. – анализ литературных данных, редактирование рукописи.

ФИНАНСИРОВАНИЕ РАБОТЫ

Данная работа финансировалась за счет средств гранта Российского научного фонда (проект № 22-25-00068). Никаких дополнительных грантов на проведение или руководство данным конкретным исследованием получено не было.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Все исследования проводились в соответствии с принципами биомедицинской этики, изложенными в Хельсинкской декларации 1964 г. и последующих поправках к ней. Они также были одобрены Комитетом по этике Института эволюционной физиологии и биохимии им. И.М. Сеченова (протокол № 01-10 от 18 октября 2022 г.)

Каждый участник исследования дал добровольное письменное информированное согласие после получения разъяснений о потенциальных рисках и преимуществах, а также о характере предстоящего исследования.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы данной работы заявляют, что у них нет конфликта интересов.

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. *Andreeva IG* (2018) Spatial selectivity of hearing in speech recognition in speech-shaped noise environment. *Hum Physiol* 44: 226–236.
<https://doi.org/10.1134/S0362119718020020>
3. *Andreeva IG, Ogorodnikova EA* (2022). Auditory Adaptation to Speech Signal Characteristics. *J Evol Biochem Physiol* 58(5): 1293–1309.
<https://doi.org/10.1134/S0022093022050027>
4. *Marrone N, Mason CR, Kidd G* (2008) Tuning in the spatial dimension: Evidence from a masked speech identification task. *J Acoust Soc Am* 124(2): 1146–1158.
<https://doi.org/10.1121/1.2945710>
5. *Haustein BG* (1969) Hypothesen uber die einhorige Entfernungswahrnehmung des menschlichen Gehors. *Hochfrequenztechnik und Electroakustic* 78(2): 45–57.
6. *Mershon DH, Bowers JN* (1979) Absolute and relative cues for the auditory perception of egocentric distance. *Perception* 8(3): 311–322.
<https://doi.org/10.1068/p080>
7. *Kolarik AJ, Moore BC, Zahorik P, Cirstea S, Pardhan S* (2016) Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Atten Percept Psychophys* 78: 373–395.
<https://doi.org/10.3758/s13414-015-1015-1>
8. *Андреева ИГ, Бахтина АВ, Гвоздева АП* (2014) Разрешающая способность слуха человека по расстоянию при приближении и удалении источников звука разного спектрального состава. *Сенс сист* 28(4): 3–12. [*Andreeva IG, Bahtina AV, Gvozdeva AP* (2014) Human's hearing resolution in case of localizing of approaching and withdrawing sound images with various spectral structures. *Sensory Systems* 28(4): 3–12. (In Russ)].
<https://www.elibrary.ru/item.asp?id=22741084>
9. *Andreeva IG, Dymnikova M, Gvozdeva AP, Ogorodnikova EA, Pak SP* (2019). Spatial separation benefit for speech detection in multi-talker babble-noise with different egocentric distances. *Acta Acust United Acust* 105(3): 484–491.
<https://doi.org/10.3813/AAA.919330>
10. *Огородникова ЕА, Лабутина ОВ, Андреева ИГ, Гвоздева АП, Баулин ЮА* (2020) Фактор просодики в восприятии коммуникативной сцены с пространственным разделением источников речи и речеподобной помехи. *Лингвистический форум 2020: Язык и искусственный интеллект*. Москва, 12–14 ноября: 127–128. [*Ogorodnikova EA, Labutina OV, Andreeva IG, Gvozdeva AP, Baulin YA* (2020) Faktor prosodiki v vospriyatii kommunikativnoj sceny s prostanstvennym razdeleniem istochnikov rechi i rechepodobnoj pomekhi [The prosody factor in the perception of a communicative scene with spatially separate sources of speech and speech-like interference. *Linguistic Forum 2020: Language and artificial intelligence*. Moscow, November 12-14: 127–128. (In Russ)].
11. *Kleczkowski P, Żak A, Król-Nowak A* (2017) Lombard effect in Polish speech and its comparison in English speech. *Arch Acoust* 42(4): 561–569.
<https://doi.org/10.1515/aoa-2017-0060>.

12. *Brungart DS, Simpson BD, Ericson MA, Scott KR* (2001) Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am* 110(5 Pt 1): 2527–2538. <https://doi.org/10.1121/1.1408946>.
13. *Van Ngo T, Kubo R, Morikawa D, Akagi M* (2017) Acoustical analyses of tendencies of intelligibility in lombard speech with different background noise levels. *J Signal Process Syst* 21(4): 171–174. <https://doi.org/10.2299/jsp.21.171>
14. *Tang P, Xu Rattanasone N, Yuen I, Demuth K* (2017) Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *J Acoust Soc Am* 142(2): 493–503. <https://doi.org/10.1121/1.4995998>
15. *Lu Y, Cooke M* (2008) Speech production modifications produced by competing talkers, babble, and stationary noise. *J Acoust Soc Am* 124(5): 3261–3275. <https://doi.org/10.1121/1.2990705>
16. *Lu Y, Cooke M* (2009) Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *J Acoust Soc Am* 126(3): 1495–1499. <https://doi.org/10.1121/1.3179668>
17. *Garnier M, Henrich N* (2014) Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Comput Speech Lang* 28(2): 580–597. <https://doi.org/10.1016/j.csl.2013.07.005>
18. *Keith RW* (2000) Random Gap Detection Test. St Louis: Auditec St Louis 2000.
19. *Gvozdeva AP, Lunichkin AM, Zaytseva LG, Ogorodnikova EA, Andreeva IG* (2022) Joint Changes in First and Second Formants of /a/, /i/, /u/ Vowels in Babble Noise—a New Statistical Approach. In *International Conference on Speech and Computer*. Cham: Springer Int Publ. 252–264. https://doi.org/10.1007/978-3-031-20980-2_22
20. *Strouse A, Ashmead DH, Ohde RN, Grantham DW* (1998) Temporal processing in the aging auditory system. *J Acoust Soc Am* 104(4): 2385–2399. <https://doi.org/10.1121/1.423748>
21. *Puts DA, Gaulin SJC, Verdolini K* (2006) Dominance and the evolution of sexual dimorphism in human voice pitch. *Evol Hum Behav* 27(4): 283–296. <https://doi.org/10.1016/j.evolhumbehav.2005.11.003>
22. *Stowe LM, Golob EJ* (2013) Evidence that the Lombard effect is frequency-specific in humans. *J Acoust Soc Am* 134(1): 640–647. <https://doi.org/10.1121/1.4807645>
23. *Bottalico P, Passione II, Graetzer S, Hunter EJ* (2017) Evaluation of the starting point of the Lombard effect. *Acta Acust United Acust* 103(1): 169–172. <https://doi.org/10.3813/AAA.919043>
24. *Pohjalainen J, Raitio T, Yrttiaho S, Alku P* (2013) Detection of shouted speech in noise: Human and machine. *J Acoust Soc Am* 133(4): 2377–2389. <https://doi.org/10.1121/1.4794394>
25. *Berg M, Fuchs M, Wirkner K, Loeffler M, Engel C, Berger T* (2017) The Speaking Voice in the General Population: Normative Data and Associations to Sociodemographic and Lifestyle Factors. *J Voice* 31(2): 257.e13–257.e24. <https://doi.org/10.1016/j.jvoice.2016.06.001>
26. *Шилenkova BB, Бестолkова ОС* (2013) Пресбифония. Возрастные изменения акустических параметров голоса. *Вестн оториноларингол* 78(6): 24–27. [*Shilenkova VV, Bestolkova OS* (2013) Age-related changes in the acoustic voice characteristics. *Vestn Otorinolaringol* 8(6): 24–27. (In Russ)]. <https://www.elibrary.ru/item.asp?id=21074035>

Lombard speech detection in case of spatial separation between noise source and talkers of different genders

I. G. Andreeva^{a,*}, A. M. Lunichkin^a, and E. A. Ogorodnikova^{a,b}

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

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

^{*}*e-mail: ig-andreeva@mail.ru*

The spatial selectivity of hearing to speech signals was studied when the target signal and interference were separated by distance between their sources and the listener. In the work, the hypothesis about the improvement of hearing selectivity due to more intensive activation of the high-frequency binaural mechanism due to the shift of the speaker's voice spectrum occurs in noisy environment towards high frequencies, was tested. The thresholds for detecting the target signal – a two-syllable word uttered by male or female, were evaluated in the two-alternative two-interval forced choice paradigm in 4 series. Series differed by the type of target signal (ordinary or Lombard speech) and the location of target source and noise (multi-talker noise) one. The both sources were located at a distance of 1 and 4 m opposite the subject at the level of his head. The detection threshold was defined as the ratio of signal and noise levels at the listener's place (S/N). The threshold for detecting the target signal (male and female speaker voices together) was -11 dB S/N for ordinary as well as Lombard speech. It did not depend on which of the sources - the target signal or noise, was closer to the listener. In normal speech, the detection thresholds on average differed for male and female voices, but the difference was not significant. In Lombard speech, these thresholds were significantly different: for a male voice, the threshold at a detection level of 0.67 was -10 dB S/N, and for a female voice – -12.5 dB S/N.

Keywords: speech, spectral characteristics of voice, Lombard effect, binaural hearing, spatial hearing selectivity, talker-to-listener distance, speech detection