АКУСТИЧЕСКИЙ ЖУРНАЛ, 2023, том 69, № 3, с. 340–350

АКУСТИКА ЖИВЫХ СИСТЕМ. БИОМЕДИЦИНСКАЯ АКУСТИКА

УДК 534.784,612.85,81.342

ИЗМЕНЕНИЕ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК ГЛАСНЫХ ЗВУКОВ В РУССКОЙ РЕЧИ НА ФОНЕ ШУМА

© 2023 г. А. М. Луничкин[a], *, И. Г. Андреева[a], Л. Г. Зайцева[a], А. П. Гвоздева[a], Е. А. Огородникова[b]

[a]*Федеральное государственное бюджетное учреждение науки Институт эволюционной физиологии и биохимии им. И.М. Сеченова Российской академии наук, пр. Тореза 44, Санкт-Петербург, 194223 Россия*

[b]*Федеральное государственное бюджетное учреждение науки Институт физиологии им. И.П. Павлова Российской академии наук, наб. Макарова 6, Санкт-Петербург, 199034 Россия*

**e-mail: BolverkDC@mail.ru*

Поступила в редакцию 24.12.2021 г.
После доработки 29.09.2022 г.
Принята к публикации 16.03.2023 г.

В контексте проблемы слухового анализа сложной сцены исследованы акустические характеристики русской речи в условиях шума многоголосия и проявления эффекта Ломбарда. Сравнивали спектры ударных гласных звуков [a], [u], [i] в словах, произнесенных шестью женщинами в тишине и на фоне диотически предъявляемого речеподобного шума уровня 60 дБ, имитирующего многоголосие. В шуме, по сравнению с тишиной, получили повышение частоты основного тона голоса (F_0) и первой форманты (F_1) для всех выделенных гласных. Общей закономерности в изменениях второй форманты (F_2) обнаружено не было. При произнесении гласного звука [i] в шуме F_2 понижалась у всех дикторов, при произнесении гласных звуков [u] и [a] она могла как понижаться, так и повышаться. Таким образом, в основном характер выявленных изменений спектральных характеристик гласных звуков русской речи в шуме соответствовал особенностям ломбардной речи для ряда европейских и азиатских языков. При этом впервые была показана обратно пропорциональная зависимость между F_0 диктора в тишине и ее изменениями в шуме: чем выше F_0 в тишине, тем меньше ее увеличение на фоне шума. Выявленные спектральные изменения отражают процессы адаптивной коррекции артикуляции, направленные на выделение голоса диктора и повышение разборчивости его речи на фоне речеподобного шума.

Ключевые слова: акустика речи, шум многоголосия (речеподобный), слухоречевой контроль, эффект Ломбарда, речевая коммуникация, характеристики голоса

DOI: 10.31857/S032079192110018X, **EDN:** QRABNB

ВВЕДЕНИЕ

В условиях речевого общения его участники решают несколько задач, связанных с приемом и передачей информации. При этом для слушателя наиболее важной из них является выделение голоса и распознавание речи диктора, а для диктора – голосовое оформление высказывания, обеспечивающее минимальную потерю его информационного содержания. В повседневной жизни эти задачи могут осложняться присутствием в среде шума, маскирующего речевой сигнал. Таким образом, физиологическая проблема, исследуемая с середины прошлого века и именуемая “cocktail-party problem” [1], наряду с аспектом восприятия речи слушателем, включает и аспект “подстройки” или адаптации диктором собственной речи к условиям шума для повышения ее разборчивости [2]. При автоматическом обнаружении и распознавании целевого речевого сигнала в условиях многокомпонентной акустической сцены, в том числе в шуме многоголосия, наиболее сложной является ситуация, в которой маскирующий и целевой речевой сигналы оказываются сходными по своим спектральным характеристикам [3].

Для диктора наличие шума многоголосия снижает возможности слухового контроля речи по принципу обратной связи и обнаружения речевых ошибок [4, 5]. По этой причине диктор вынужден изменять параметры собственного голоса. Непроизвольное адаптивное изменение параметров речи в зашумленной обстановке носит название эффекта Ломбарда, а саму речь обозначают как ломбардную [2, 6, 7]. На данный момент центральные механизмы речеобразования, отвечающие за изменение голоса в шуме, точно не установлены [2, 8]. Нейрофизиологические исследования слуха млекопитающих (*Felis catus* и *Saimiri sciureus*)

позволили связать эффект Ломбарда с работой структур, расположенных в стволе мозга. Среди них выделяют серое околоводопроводное вещество среднего мозга [9], верхнеоливарный комплекс и парамедианную ретикулярную формацию моста [10]. Среди других структур, которые возможно вовлечены в организацию ломбардной речи, называют латеральную зону ретикулярной формации и слухоречевые зоны первичной слуховой коры [8].

Эффект Ломбарда обнаруживается в присутствии шума, начиная с уровня интенсивности 43 дБ [11]. Он может играть двойную роль в речевом общении: облегчать распознавание слов слушателем, с одной стороны, и улучшать условия слухового контроля речи диктора по принципу обратной связи (самопрослушивание) – с другой [7, 12, 13]. Исследование этих процессов и их влияния на восприятие речи в условиях многоголосия необходимо для решения актуальных задач повышения помехоустойчивости автоматических систем распознавания речи, создания голосозависимых технологий и автоматических голосовых определителей [14, 15]. Оно также имеет важное значение для развития речевых тренажеров, процедур слуховых тренировок, технических средств слухопротезирования [16], методов диагностики слухоречевых нарушений [17] и биометрической идентификации дикторов [2].

При сравнении с речью в тишине, ломбардная речь отличается повышением интенсивности и частоты основного тона голоса (F_0) [3, 6, 7, 13, 18, 19], увеличением длительности и изменением спектральных характеристик гласных, в первую очередь их основных частотных максимумов – первой и второй формант (F_1 и F_2) [7, 20]. Увеличение F_0 наблюдали в присутствии шума уровнем от 43 до 84 дБ, причем с ростом уровня шума прирост частоты также увеличивался [7]. Значения F_1 в шуме для всех гласных сдвигались в область более высоких частот, а закономерность изменений F_2 не обнаруживалась [7, 12, 21]. Благодаря этим изменениям гласные оказываются более различимыми на фоне шума. Это облегчает распознавание слов слушателем и улучшает процесс самопрослушивания у диктора [7, 12, 21]. Таким образом, адекватным методом оценки особенностей речи в зашумленной среде выступает изучение изменений в формантной структуре гласных звуков, прежде всего, в отношении спектральных максимумов, соответствующих частотам F_1 и F_2, которые в речи взрослого человека различны для разных категорий гласных [22] и являются основой для их фонемной идентификации [23-25].

В русском языке базовые гласные звуки [a], [i], [u] в координатной плоскости F_1–F_2 образуют вершины фонетического “треугольника гласных”, внутри которого расположены остальные гласные фонемы [26, 27]. Поэтому изучение формантных характеристик гласных звуков [a], [i], [u] позволяет оценить изменения, которые в целом происходят при голосообразовании и артикуляции вокальных звуков русской речи в шуме. При такой оценке необходимо учитывать положение гласного в речевом сигнале, в частности, ударность его позиции в слогоритмической структуре слова (словесное ударение), когда формантные максимумы спектра гласных звуков наиболее выражены, благодаря четкости артикуляции [26].

Следует подчеркнуть, что особенности русской ломбардной речи ранее практически не изучались. В 1970-х гг. было показано, что эффект Ломбарда приводит к непроизвольному усилению голоса и увеличению длительности звуков вокальной речи [28], однако методические ограничения не позволили выполнить спектральный анализ ее изменений.

Целью данной работы стало сравнение спектральных характеристик гласных звуков русской речи, находящихся в ударной позиции, при произнесении ограниченного набора слов в условиях тишины и на фоне речеподобного шума, имитирующего многоголосие. Оценивали изменения F_0, отражающей индивидуальную характеристику голоса и интонирования, а также спектральных максимумов, соответствующих формантам F_1 и F_2, как показателей артикуляции гласных звуков. С учетом преобладания в научной литературе данных по эффекту Ломбарда для образцов женского голоса и свидетельства о его большей выраженности у женщин, полученных на материале других языков [3, 7, 19, 21, 29–31], объектом нашего внимания стали гласные звуки русского языка в женской речи.

МЕТОДИКА

Дикторы. В записи речи приняли участие 6 женщин в возрасте 20–32 лет ($n = 3$) и в возрасте 58–59 лет ($n = 3$). Все они являлись носителями нормативного русского языка, не имеющими нарушений слуха и дефектов речи, а также опыта длительного общения с маленькими детьми до проведения исследования. Такой подбор испытуемых позволял исключить выраженность изменений акустических характеристик, типичных для материнской речи [19, 21, 32]. Аудиологическое обследование дикторов включало воздушную тональную пороговую аудиометрию и тест обнаружения паузы, представляющий собой стандартное исследование временной разрешающей способности слуха с применением коротких тональных посылок на частотах 0.5, 1, 2 и 4 кГц, и широкополосных щелчков [33]. По результатам обследования дикторов пороги их слуха не превышали 20 дБ на основных аудиометрических частотах и

соответствовали возрастной норме тонального слуха, а пороги обнаружения паузы не превышали 20 мс для всех типов стимулов, что соответствовало норме временной разрешающей способности. Перед началом эксперимента дикторы подписывали информированное согласие об участии в эксперименте. Все процедуры, выполненные в настоящем исследовании с участием людей, соответствовали требованиям Этического комитета Института эволюционной физиологии и биохимии им. И.М. Сеченова и Хельсинкской декларации 1964 г. с ее последующими изменениями.

Речевой материал. Для записи в тишине и на фоне маскирующего сигнала – речеподобного шума – использовали 9 слов русского языка с гласными звуками [a], [i], [u] в разных ударных позициях: в начале слова “рУчка, Армия, мИна”; в середине – “бумАга, малИна, посУда”; в конце – “кредИт, шалУн, строкА”.

Речеподобный шум (далее шум) создавали путем микширования записей двусложных слов, произнесенных четырьмя дикторами разного пола и возраста – двое мужчин (30 и 65 лет) и две женщины (19 и 60 лет), ни один из которых не участвовал в настоящем исследовании. Средние значения и стандартные отклонения F_0 голосов дикторов составляли 117 ± 8 и 139 ± 9 Гц для мужчин, 208 ± 30 и 234 ± 34 Гц для женщин. Записи представляли собой двусложные слова русской речи длительностью от 400 до 800 мс. Всего применяли восемь слов, ни одно из которых не совпадало со словами из тестового речевого материала. Для каждого слова каждого диктора создавалась звуковая дорожка с многократным повтором слова без пауз. Таким способом были получены 32 звуковые дорожки (8 слов × 4 диктора). Для создания речеподобного шума (многоголосие) они были микшированы в файл длительностью 40 с. Полученный шум нормализовали по уровню, затем сформировали линейные фронты нарастания и убывания интенсивности по 1 с. Данная методика формирования шума была описана в нашей работе [34]. Спектральные характеристики шума представлены ниже в разделе Результаты. Шум при его предъявлении в наушниках создавал у слушателя ощущение присутствия в помещении с большим числом одновременно говорящих людей. Измеренный в А-взвешенном режиме уровень звукового давления шума при монофоническом воспроизведении составлял 52 дБ(А). При диотическом предъявлении шума в условиях эксперимента этот уровень соответствовал воспринимаемому уровню громкости монаурально подаваемого шума 60 дБ. Данный эффект обусловлен бинауральной суммацией громкости [35].

Оборудование и экспериментальное помещение. Исследование проводили в анэхоидной звукоизолированной камере объемом 62.5 м^3. Ослабление уровня наружных шумов в камере составляло не менее 40 дБ в диапазоне частот 0.5–16 кГц. Воспроизведение шума и запись голоса диктора выполняли синхронно с применением ноутбука ASUS Sonic Master и программного обеспечения Adobe Audition 1.6.

Для записи голоса использовали устройство Rode NT-USB, сочетающее в себе конденсаторный микрофон с кардиоидной диаграммой направленности и звуковую карту (частота дискретизации 44100 Гц, 16 бит). Звуковая карта имела раздельные регуляторы громкости входного сигнала с ноутбука и входного сигнала с микрофона. Устройство Rode NT-USB было оснащено выходом, к которому подключали головные телефоны закрытого типа Sennheiser HD-380-Pro. Измерение уровня звукового давления, создаваемого шумом, осуществляли при помощи шумомера RFT 000014 в А-взвешенном режиме при монауральной подаче звука.

Условия самопрослушивания. Головные телефоны уменьшали воспринимаемую диктором по воздушной проводимости громкость собственной речи, поэтому их применение могло влиять на параметры речи диктора (громкость, F_0 и т.п.). Для снижения этого эффекта перед началом эксперимента диктора просили выполнить подстройку условий самопрослушивания. С этой целью диктор менял громкость входного сигнала с микрофона при помощи регулятора на устройстве Rode NT-USB таким образом, чтобы воспринимаемая громкость собственной речи в головных телефонах и без них была одинаковой. Условия самопрослушивания, отрегулированные диктором, сохранялись при произнесении слов в тишине и в шуме.

Экспериментальная процедура. Во время эксперимента диктор располагался в анэхоидной камере в кресле с подлокотниками и подголовником. В 20 см от диктора, на уровне его рта, устанавливали микрофонное устройство на напольной стойке. Положение головы не фиксировали жестко, однако диктору давали инструкцию держать затылок прижатым к подголовнику, не поворачивать и не наклонять голову. Таким образом поддерживалось постоянное расстояние между губами диктора и микрофоном. На дикторе были надеты головные телефоны закрытого типа. После подстройки условий самопрослушивания (см. предыдущий пункт) осуществляли запись слов. Каждый диктор участвовал в трех сессиях, в которых записывали по три слова. Сессии различались по набору слов, в которых ударная гласная занимала одну из трех возможных позиций. В течение одной сессии диктор по четыре раза произносил друг за другом три слова в следующих условиях: 1) в тишине, 2) в шуме уровня 60 дБ. При произнесении слов диктор обращался к экспериментатору, который си-

Таблица 1. Изменения частоты основного тона голоса при произнесении гласных звуков [u], [i], [a] на фоне диотического речеподобного шума уровня 60 дБ

Диктор №/возраст (лет)	$F_{0к}$ в тишине (Гц)*	$F_{0ш}$ в шуме (Гц)*	ΔF_0 (%)**		
			[a]	[i]	[u]
1/20	233	220	–0.8	–0.3	–5.6
2/25	179	204	20.3	16.4	6.2
3/32	197	206	6.7	8.1	–0.9
4/58	174	191	17.6	8.5	18.0
5/58	167	187	10.0	9.2	15.6
6/59	171	209	10.4	15.3	24.1
Медиана***	10.2	8.8	10.9		

*Приведены медианы $F_{0к}$, $F_{0ш}$ по всем записям диктора в тишине для 9 слов и 4 их повторений ($n = 36$).
**Даны индивидуальные значения ΔF_0 как медианы, полученные для 3 ударных положений гласного звука и 4 их повторений ($n = 12$).
***Групповые ΔF_0 даны как медианы индивидуальных изменений ($n = 6$).

дел напротив диктора на расстоянии 1 м. Для того чтобы уменьшить различия в интонировании слов, дикторов просили перед каждым словом добавлять местоимение “это”: “Это - ручка, это – армия, это – мина” и т.д. После окончания сессии записи голоса сохраняли в формате “wav” для дальнейшего анализа. Для каждого из шести дикторов было записано 72 звуковых фрагмента (9 слов × 4 повтора × 2 условия). Общий объем речевого материала для шести дикторов составил 432 записи слов.

Анализ записей и методы статистического анализа. Анализ ударных гласных звуков из записей слов включал выделение стационарного участка гласного (не менее 50 мс) для определения F_0, F_1 и F_2 в программе Praat (свободно распространяемое программное обеспечение, www.praat.org). Частотный диапазон для оценки F_0 составлял 75–500 Гц. Для определения значений формант использовали авторегрессионный метод Берга (Burg Linear Predictive Coding Autoregressive Method), реализованный в Praat. Шаг по времени составлял 0.01 с, длительность окна интегрирования 0.025 с, максимальное искомое значение формант 5.5 кГц, коррекция предыскажения (pre-emphasis) выполнялась на частотах свыше 50 Гц. Статистическую обработку полученных данных проводили в программах Excel и Statistica 10. Для сравнения параметров речи дикторов в условиях тишины (контроль) и на фоне шума использовали непараметрические методы, в том числе непараметрический *U*-критерий Манна–Уитни и парный критерий Вилкоксона.

РЕЗУЛЬТАТЫ

Изменение частоты основного тона в условиях диотического речеподобного шума по сравнению с тишиной

При контрольных условиях (тишина) в записях отдельных слов группы из 6 испытуемых частота основного тона ($F_{0к}$) находилась в диапазоне 115–310 Гц. Индивидуальные значения медианы $F_{0к}$, которые были получены для всех гласных звуков независимо от их положения в слове, представлены в табл. 1. При анализе этой спектральной характеристики голоса в тишине оказалось, что у дикторов старшей возрастной подгруппы она была ниже, чем в младшей возрастной подгруппе ($p < 0.01$, непараметрический *U*-критерий Манна–Уитни).

При говорении в шуме индивидуальная величина ($F_{0ш}$) была достоверно выше ($p < 0.01$, непараметрический парный критерий Вилкоксона) по сравнению с условиями тишины у дикторов № *2–5*. У диктора № *1* $F_{0ш}$ оказалась ниже $F_{0к}$, которое было самым высоким в группе – 233 Гц. Индивидуальные данные относительных изменений величины $F_{0ш}$ по сравнению $F_{0к}$ (ΔF_0) для каждого гласного звука представлены в табл. 1. Значения ΔF_0 составляли от –5.6 до 24.1%, причем изменение для всех гласных звуков у одного и того же диктора было близким по величине. Заметим, что у дикторов старшего возраста значения ΔF_0 были выражены сильнее по сравнению с дикторами младшего возраста ($p < 0.05$, непараметрический *U*-критерий Манна–Уитни). В целом по группе увеличение медианных значений F_0 находилось в диапазоне 8.8–10.9%.

Анализ зависимости изменений частоты основного тона в шуме по сравнению с тишиной (ΔF_0) от $F_{0к}$ для трех гласных звуков в группе дикторов показал, что чем выше была $F_{0к}$, тем меньше она увеличивалась на фоне шума (рис. 1а–1б). Для гласного звука [а] уравнение линии регрессии описывалось функцией $y = -0.25x + 66$, для гласного [i] – $y = -0.24x + 62$, для гласного [u] – $y = -0.41x + 94$. Коэффициенты детерминации были равны соответственно $R^2 = 0.30$ для [а], $R^2 = 0.21$ для [i], $R^2 = 0.47$ для [u], что соответствовало проявлению умеренной корреляции в первом случае и слабой – во втором и третьем, согласно шкале Чеддока [36]. Причем при высоких $F_{0к}$ (выше 175 Гц) ΔF_0 в отдельных случаях было отрицательным, что можно проследить на рис. 1а–1б. С повышением $F_{0к}$ число таких случаев нарастало, о чем свидетельствовала полученная линейная зависимость. Наиболее низкие значения ΔF_0 наблюдались у дикторов № *1* и № *3*, которые обладали самыми высокими голосами, у диктора № *1* ΔF_0 имело отрицательные значения для всех гласных (табл. 1).

Изменения формантных частот F_1–F_2 гласных звуков в условиях диотического речеподобного шума по сравнению с тишиной

Индивидуальные изменения значений формантных частот F_1–F_2 (ΔF_1–ΔF_2) гласных звуков в условиях шума (табл. 2) были рассчитаны аналогично ΔF_0. Частота форманты F_1 достоверно повышалась для всех гласных звуков ($p < 0.01$, здесь и далее, если не указано иное – непараметрический парный критерий Вилкоксона). Индивидуальные ΔF_0 варьировали от –1.2 до 22.3% для разных гласных звуков. Отметим, что ΔF_1 всех гласных звуков были близкими по величине для дикторов младшего и старшего возраста. По всей группе испытуемых относительные медианные изменения для гласного звука [а] составляли 5.0%, в то время как для гласных звуков [i] и [u] они были больше и близкими по величине – 13.7% для [i], 13.6% для [u]. При этом выраженных индивидуальных и возрастных различий в группе дикторов не наблюдалось.

На фоне шума индивидуальные изменения второй форманты по сравнению с тишиной (ΔF_2) не превышали 7.4%. Для звука [i], характеризуемого наиболее высоким значением F_2 в тишине по сравнению с остальными гласными звуками, групповое медианное значение в шуме не повышалось, а, наоборот, снижалось на 2.6% ($p < 0.05$). Для гласного звука [а] ΔF_2 достоверно увеличивалось на 1.9% ($p < 0.05$). В случае гласного звука [u] медианное групповое значение F_2 достоверно не изменялось в шуме по сравнению с тишиной ($p = 0.902$).

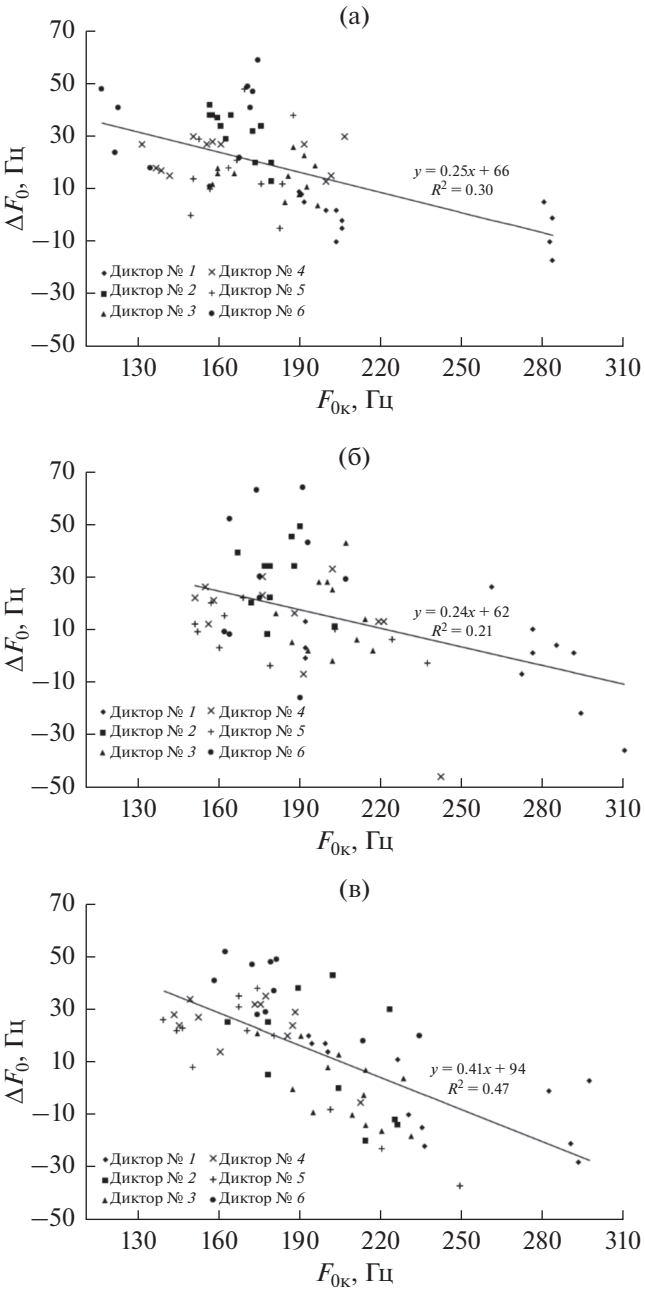

Рис. 1. Зависимость изменений частоты основного тона ΔF_0 на фоне диотического речеподобного шума уровня 60 дБ по сравнению с тишиной, от $F_{0к}$ в тишине для разных ударных гласных звуков. (а) – Для гласного звука [а], (б) – для гласного звука [i], (в) – для гласного звука [u]. Прямая линия – линия регрессии. Разными символами обозначены индивидуальные значения для 6 дикторов-женщин.

В групповых данных проявились те же различия в характере изменений F_1 и F_2 для разных гласных звуков (рис. 2а–2в). Увеличение медианных значений F_1 в шуме было достоверно значимым ($p < 0.01$): для [а] на 40 Гц, для [i] на 47 Гц, для [u] на 51 Гц. Изменения медианных значений F_2

Таблица 2. Относительные изменения частоты формант F_1–F_2 гласных звуков [u], [i], [a] при произнесении на фоне диотического речеподобного шума уровня 60 дБ по сравнению с тишиной

Диктор №	[a]		[i]		[u]	
	ΔF_1	ΔF_2	ΔF_1	ΔF_2	ΔF_1	ΔF_2
1	2.6*	–1.0	16.1	–2.8	12.1	6.1
2	12.6	7.4	10.5	–3.0	16.2	1.1
3	7.4	6.6	7.5	–2.1	10.0	–6.2
4	0.2	3.5	16.0	–1.9	15.1	2.2
5	21.5	0.2	11.9	–4.4	22.3	0.3
6	–1.2	–4.1	22.2	–2.4	10.6	0.8
Медиана**	5.0	1.9	13.7	–2.6	13.6	1.0

*Даны индивидуальные значения ΔF_1 и ΔF_2 в процентах как медианы, полученные для 3 ударных положений гласного звука и 4 их повторений ($n = 12$).
**Групповые изменения спектральных характеристик даны как медианы индивидуальных изменений ($n = 6$).

были различны для трех гласных звуков: для [a] частота F_2 повышалась на 26 Гц ($p < 0.01$), для [i] – понижалась на 69 Гц ($p < 0.01$), а для [u] практически не менялась ($p = 0.68$).

С использованием теста Левена было проведено сравнение дисперсий генеральных совокупностей значений F_1 и F_2 в двух экспериментальных условиях. Сравнение показало, что вариабельность значений F_1 и F_2 гласного звука [a] в речеподобном шуме снижалась ($p < 0.05$). На плоскости F_1–F_2 это выражалось в том, что область положения звуков становилась более компактной. Для гласных звуков [i] и [u] вариабельность значений не менялась ([i] – $p = 0.22$ для F_1, $p = 0.22$ для F_2; [u] – $p = 0.66$ для F_1, $p = 0.16$ для F_2).

На рис. 3 групповые данные представлены формантными треугольниками в тишине и в речеподобном шуме уровня 60 дБ. Вершины треугольников соответствуют положению в координатной плоскости F_1–F_2 гласных звуков [u], [i], [a]. Сдвиг вершин треугольника в шуме относительно соответствующих вершин в тишине отражает изменения F_1 и F_2 в координатном пространстве. Эти изменения привели к уменьшению площади треугольника гласных с 324495 до 301013 Гц^2, расчет которой выполняли согласно формуле [27, 37]. Уменьшение площади треугольника на 7% свидетельствовало о тенденции к централизации гласных, что могло быть проявлением ухудшения слухоречевого контроля.

Описанные выше изменения спектральных максимумов гласных звуков, возникающие в речеподобном шуме, можно обобщить схемой, представленной на рис. 4. Согласно схеме, направления сдвигов спектральных характеристик в шуме могли быть разнонаправленными для F_0, F_1, F_2, но все они соответствовали смещению к локальным минимумам спектра шума. Так, частота основного тона голоса увеличивалась на фоне шума для всех гласных звуков, приближаясь к ло-

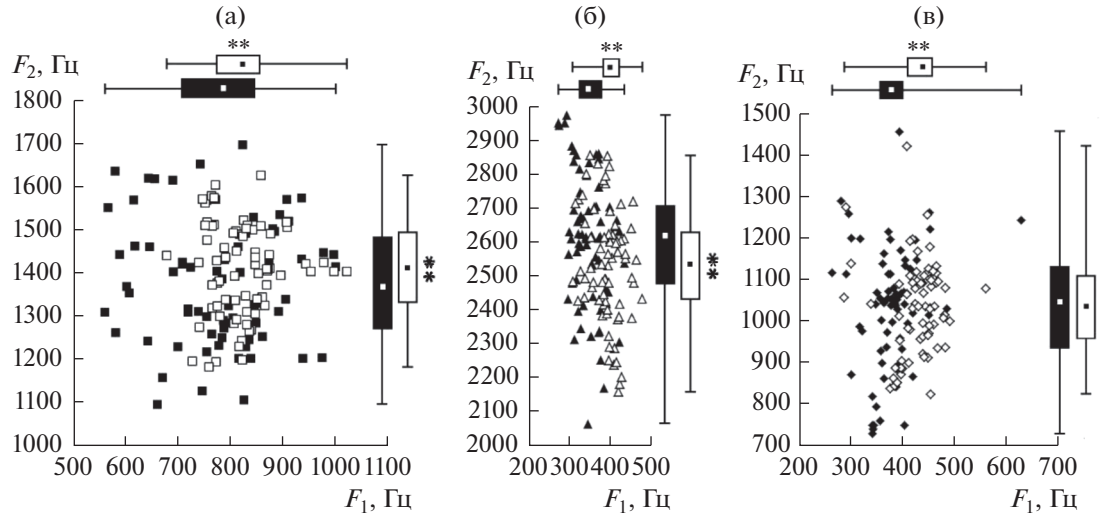

Рис. 2. Положение ударных гласных звуков [a], [i], [u] на плоскости первой и второй формант (F_1–F_2) в тишине и на фоне диотического речеподобного шума. Приведены данные для каждой гласной в трех ударных положениях, произнесенных четыре раза шестью дикторами-женщинами ($n = 72$). (а) – Звук [a], (б) – звук [i], (в) – звук [u]. Черными символами обозначены положения гласного звука в тишине, белыми – в шуме уровня 60 дБ; Показаны минимальное значение, первый квартиль, медиана, третий квартиль и максимальное значение F_1 и F_2; ** – достоверные изменения медианных значений гласных звуков, произнесенных в шуме, по сравнению с условиями тишины ($p < 0.01$, непараметрический парный критерий Вилкоксона).

кальному минимуму порядка 200 Гц. Форманты F_1 всех трех гласных звуков и F_2 звука [a] повышались, F_2 звука [i] снижалась. Все эти изменения приводили к тому, что спектральные параметры речи смещались в частотные области с меньшей шумовой нагрузкой. Отметим, что F_2 гласного звука [u] в шуме практически не менялась. Это может быть связано с тем, что контрольные значения F_2 гласного звука [u] в тишине близки к значениям локального минимума около 1100 Гц и смещение F_2 в шуме могло привести к ухудшению соотношения сигнал/шум. Таким образом, описанные выше изменения спектральных характеристик гласных звуков на фоне речеподобного шума способствовали решению перцептивной задачи – выделению целевого речевого сигнала путем улучшения соотношения сигнал/шум.

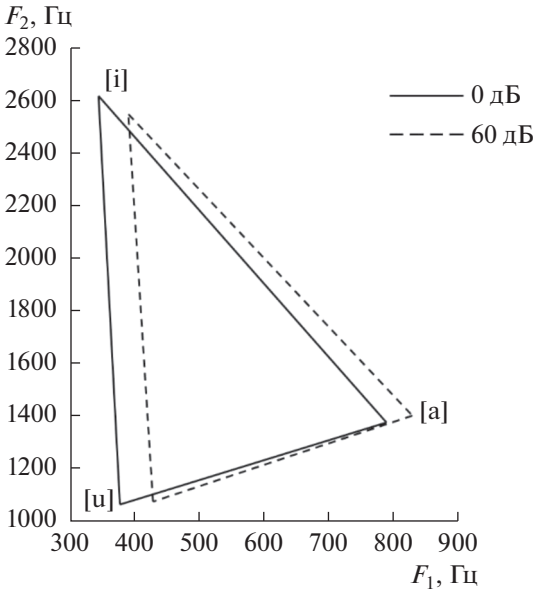

Рис. 3. Смещение треугольника гласных звуков на плоскости формант F_1–F_2 в тишине и на фоне диотического речеподобного шума. Вершины треугольников представлены медианами F_1 и F_2 гласных звуков [a], [i], [u] для группы из шести дикторов-женщин. Сплошная линия – треугольник гласных в тишине, пунктирная линия – треугольник гласных на фоне шума уровня 60 дБ.

ОБСУЖДЕНИЕ

Полученные данные свидетельствуют о том, что в присутствии речеподобного шума F_0 в русской женской речи увеличивается, причем увеличение F_0 в относительных величинах для трех гласных звуков [u], [i], [a] оказывается практически одинаковым. Сходные изменения частоты основного тона, обусловленные эффектом Ломбарда, были описаны ранее для ряда европейских и азиатских языков (табл. 3). Поскольку проявления эффекта Ломбарда в значительной мере зависят от задачи, которая стоит перед диктором, от типа и уровня шума [7, 38], а также от пола диктора [3, 29, 39], в таблице приведены условия записи ломбардной речи для каждого из исследований.

Отметим, что во всех этих работах показатели F_0 в шуме увеличивались. При этом прирост частоты основного тона значительно варьировал даже для близких по интенсивности шумов. Так, при воздействии шума в 60–70 дБ диапазон уве-

Таблица 3. Сводная таблица исследований, в которых определяли изменения частоты основного тона женского голоса на фоне шума по сравнению с тишиной

Авторы	Язык	Тип шума	Объем выборки	Задача диктора	Увеличение F_0
Letowski et al., 1993 [39]	Английский	Речеподобный шум, 70 и 90 дБ	5 женщин	Чтение текста	3 Гц для 70 дБ и 18 Гц для 90 дБ
Patel, Schell, 2008 [40]	Английский	Речеподобный шум, 60 и 90 дБ	8 женщин	Спонтанная речь	14.5 Гц для 60 дБ и 54.7 Гц для 90 дБ
Stowe, Golob, 2013 [38]	Английский	Широкополосный (0.02–20 кГц) и узкополосный (0.5–4 кГц) шум, 75 и 90 дБ	8 женщин	Спонтанная речь	5–20 Гц для широкополосного шума, 4–8 Гц для узкополосного шума
Alghamdi et al., 2018 [41]	Английский	Речеподобный шум, 80 дБ	55 мужчин и женщин	Чтение текста	41–63 Гц
Garnier et al., 2006 [12]	Французский	Речеподобный шум, 85 дБ	1 женщина	Чтение текста	50–70 Гц
Marcoux, Ernestus, 2019 [31]	Нидерландский	Речеподобный шум, 83 дБ	30 женщин	Чтение текста	10–30 Гц
Kleczkowski et al., 2017 [42]	Польский	Речеподобный шум, 82 дБ	3 женщины	Спонтанная речь	75 Гц
Van Ngo et al., 2017 [7]	Японский	Тип шума неизвестен, 66, 72, 72, 78, 84, 90 дБ	1 женщина	Чтение текста	40–110 Гц.
Zhao et al., 2019 [43]	Японский	Речеподобный шум, интенсивность неизвестна	6 женщин	Чтение текста	50–90 Гц
Наше исследование	Русский	Речеподобныйшум, 60 дБ	6 женщин	Спонтанная речь	19 Гц

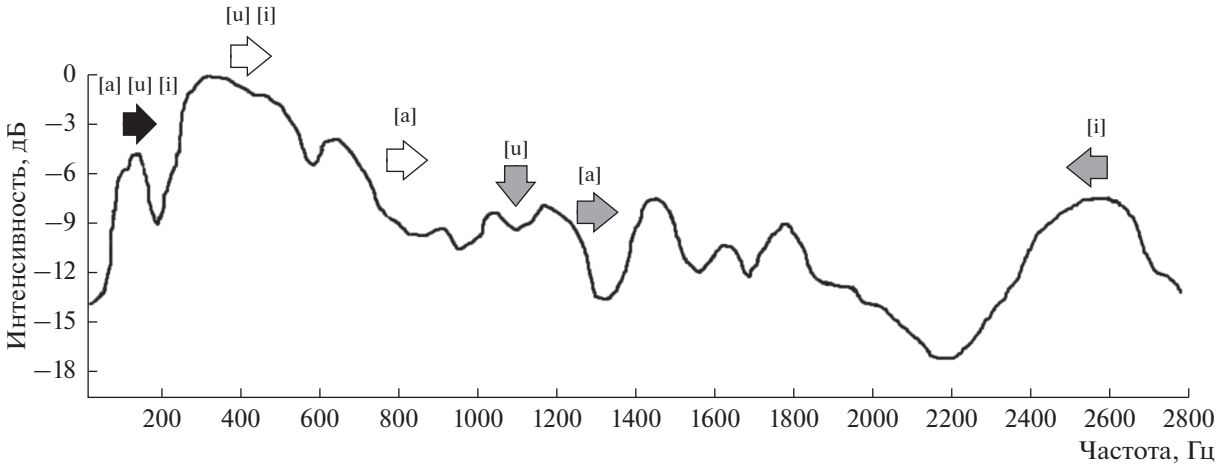

Рис. 4. Направления сдвигов спектральных характеристик гласных звуков [a], [i], [u] относительно амплитудно-частотного спектра речеподобного шума, представленного в диапазоне формант F_0–F_2. Черная стрелка – сдвиг F_0, белые стрелки – сдвиг F_1, серые стрелки – сдвиг F_2.

личения F_0 составлял от 3 до 40 Гц. Для русской речи мы получили средний прирост F_0 в шуме на 19 Гц, что сопоставимо с данными работы [40]. При этом диапазон индивидуальных значений в нашем исследовании составлял 9–34 Гц.

Известно, что показатель F_0 зависит от физиологических характеристик голосовых связок диктора – их длины, толщины, эластичности [24, 44], и может модифицироваться благодаря работе вокальных мышц [45]. С возрастом происходят изменения голоса (пресбифония), которые в основном отражаются в снижении F_0 и в ослаблении силы голоса [44, 46, 47]. При этом эффект понижения основной частоты в большей степени выражен у женщин, что связано с характерной для них атрофией голосовых связок [47]. У мужчин возрастные изменения в большей степени определяются увеличением плотности голосовых складок, поэтому понижение F_0 с возрастом у них менее выражено и может наблюдаться обратный эффект повышения высоты голоса [47]. Эти данные определили интерес к сравнению изменений в шуме с учетом показателей $F_{0к}$ (в тишине) и возрастных характеристик в подгруппах наших дикторов: младшей – возраст 20–35 лет (n = 3), и старшей – возраст 58–59 лет (n = 3).

Результаты показали, что у дикторов-женщин наблюдается следующая зависимость: чем выше F_0 в тишине, тем меньше она увеличивается в условиях многоголосия. Эта тенденция подтверждается статистически и согласуется с необходимостью подстройки голоса под спектральные характеристики применявшегося в исследовании шума (рис. 4). Как было показано в работе [3], в речеподобном шуме F_0 сдвигается в область спектра, в которой энергия маскера имеет локальные минимумы. Такая стратегия увеличивает контраст между речью диктора и фоновым шумом и зависит как от F_0 диктора в тишине, так и от спектральных характеристик шума. Кроме того, ΔF_0 в шуме у старших дикторов были выражены сильнее, чем у более молодых. При этом в младшей подгруппе наблюдались случаи понижения частоты основного тона голоса. В итоге, F_0 дикторов разного возраста на фоне речеподобного шума оказались в спектральной области с пониженной энергией маскера (200 Гц).

В отношении формантной структуры гласных, определяемой конфигурацией ротовой полости и положением языка, проявления эффекта Ломбарда носят более сложный характер [7, 12, 21]. Однако работ, посвященных изменению первой и второй формант гласных звуков в шуме значительно меньше, по сравнению с работами, в которых оценивали частоту основного тона голоса. При этом большинство работ выполнено при анализе речи только 1–2 дикторов.

На рис. 5 показаны направления изменений положения гласных в плоскости F_1–F_2 по нашим данным и по результатам, полученным при изучении французской [12] и японской [30] речи, в которых анализировали женскую речь при воздействии речеподобного шума разных уровней. Во всех этих исследованиях обнаружен прирост F_1 в присутствии шума, выраженный для всех гласных. При этом изменения второй форманты F_2 были незначительны по величине и могли иметь разное направление: в гласном [i] ее частота снижалась; в гласных [a] и [u] – увеличивалась в разной степени. Как и для F_0, увеличение F_1 и разнонаправленный характер изменения F_2 могут определяться смещением в область, где спектральная энергия шумового маскера понижена (рис. 4). Так, следует отметить, что значения F_2

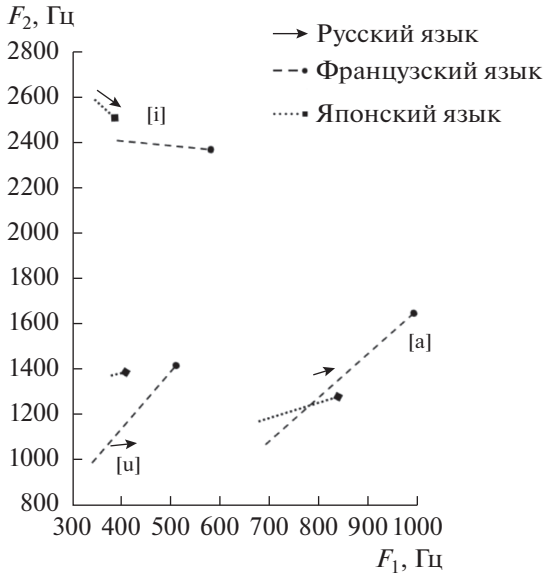

Рис. 5. Сдвиги положений трех гласных в плоскости F_1–F_2 в шуме многоголосия по отношению к соответствующим положениям в тишине в русской, французской и японской речи. Сплошные линии – сдвиг гласных звуков русской речи (наши групповые данные); пунктирные линии – французской речи [12]; точечные линии – японской речи [30].

гласного звука [u], которые уже в тишине находились в зоне сниженной энергии маскера, не обнаруживают достоверного изменения в условиях зашумления.

Таким образом, смещение спектральных характеристик гласных русской речи в плоскости F_1–F_2 под действием шума имело сложный характер и различалось у разных звуков. Эти различия могут определяться особенностями артикуляции, которые отражаются и в исходных акустических характеристиках гласных – среднее частотное положение и близость значений формант F_1 и F_2 у гласного [a] в отличие от F-картины для гласных [i] и [u]. В работах [12] и [30] анализировали данные одного диктора, что не позволило оценить вариативность изменений у разных дикторов при изучении гласных других языков. Однако, полученные различия для гласных разных языков принципиально не отличаются от индивидуальной вариабельности, продемонстрированной нашими результатами.

Еще одной отличительной особенностью нашей работы от двух упомянутых являлся тип задачи, стоявшей перед диктором. Она заключалась в том, что диктор произносил слова, обращаясь к слушателю, моделируя, таким образом, ситуацию коммуникации. Тогда как два других исследования регистрировали речь в процессе чтения. За исключением небольшого числа работ [38, 40, 42], наиболее распространенным типом задания при изучении эффекта Ломбарда являлось чтение слов или текста с листа.

Выбранная нами методика приближена к реальной речевой задаче, т.е. обеспечению передачи сигнала между диктором и слушателем, что повышает пластичность речевых программ диктора, включая вероятность ошибок произнесения. Вместе с тем, в ломбардной речи может наблюдаться снижение таких ошибок. В речи в шуме, которую мы регистрировали во второй записи, распределение в плоскости F_1–F_2 индивидуальных положений всех трех гласных звуков у всех дикторов было более компактным по сравнению с контролем. Таким образом, в условиях затрудненного речевого общения и ослабления обратной связи артикуляция гласных становится более четкой, что повышает вероятность их идентификации в шуме. Эти изменения имеют преимущественно непроизвольный характер [2, 12] и демонстрируют определенное сходство с изменением формантных характеристик гласных в зависимости от состояния слуха и результатов слухопротезирования [48].

ЗАКЛЮЧЕНИЕ

Гласные звуки русской женской речи в шуме меняются под действием речеподобного шума (эффект Ломбарда). Характер этого изменения схож с описанным для целого ряда других (английский, французский, нидерландский, японский) языков. При этом, как показывают данные проведенного исследования, общая тенденция фиксируемых сдвигов как основной частоты голоса (F_0), так и первой (F_1) и второй (F_2) формант гласных звуков, отражает смещение значений этих спектральных максимумов в область низкой энергии маскирующего шума, что способствует улучшению разборчивости речи. Впервые обнаружены зависимость прироста F_0 в шуме от контрольного значения основной частоты голоса в тишине. Показаны проявления возрастных особенностей в голосовых характеристиках дикторов-женщин. Данные определяют интерес к продолжению анализа ломбардной речи у дикторов-мужчин, голоса которых в норме характеризуются более низкочастотными значениями основного тона, чем у женщин, и гендерными особенностями пресбифонии.

Получены количественные оценки проявлений эффекта Ломбарда в русской женской речи и анализ их статистической значимости, которые могут быть применены в практических разработках, связанных с системами автоматического распознавания речи и программами реабилитации при нарушениях слухоречевой функции.

Работа поддержана средствами государственного бюджета (тема № 075-00967-23-00).

СПИСОК ЛИТЕРАТУРЫ

1. *Bronkhorst A.W.* The cocktail-party problem revisited: early processing and selection of multi-talker speech // Atten. Percept. Psychophys. 2015. V. 77. № 5. P. 1465–1487. https://doi.org/10.3758/s13414-015-0882-9
2. *Brumm H., Zollinger S.A.* The evolution of the Lombard effect: 100 years of psychoacoustic research // Behaviour. 2011. V. 148. № 11–13. P. 1173–1198. https://doi.org/10.1163/000579511X605759
3. *Garnier M., Henrich N.* Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? // Comput. Speech Lang. 2014. V. 28. № 2. P. 580–597. https://doi.org/10.1016/j.csl.2013.07.005
4. *Ludlow C.L., Cikoja D.B.* Is there a self-monitoring speech perception system? // J. Commun. Disord. 1998. V. 31. № 6. P. 505–510.
5. *Möttönen R., Watkins K.E.* Using TMS to study the role of the articulatory motor system in speech perception // Aphasiology. 2012. V. 26. № 9. P. 1103–1118. https://doi.org/10.1080/02687038.2011.619515
6. *Summers W.V., Pisoni D.B., Bernacki R.H., Pedlow R.I., Stokes M.A.* Effects of noise on speech production: Acoustic and perceptual analyses // J. Acoust. Soc. Am. 1988. V. 84. № 3. P. 917–928. https://doi.org/10.1121/1.396660
7. *Van Ngo T., Kubo R., Morikawa D., Akagi M.* Acoustical analyses of tendencies of intelligibility in lombard speech with different background noise levels // J. Signal Process. 2017. V. 21. № 4. P. 171–174. https://doi.org/10.2299/jsp.21.171
8. *Luo J., Hage S.R., Moss C.F.* The Lombard effect: from acoustics to neural mechanisms // Trends Neurosci. 2018. V. 41. № 12. P. 938–949. https://doi.org/10.1016/j.tins.2018.07.011
9. *Nonaka S., Takahashi R., Enomoto K., Katada A., Unno T.* Lombard reflex during PAG-induced vocalization in decerebrate cats // Neurosci. Res. 1997. V. 29. № 4. P. 283–289. https://doi.org/10.1016/S0168-0102(97)00097-7
10. *Hage S.R., Jürgens U., Ehret G.* Audio–vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey // Eur. J. Neurosci. 2006. V. 23. № 12. P. 3297–3308. https://doi.org/10.1111/j.1460-9568.2006.04835.x
11. *Bottalico P., Passione I.I., Graetzer S., Hunter E.J.* Evaluation of the starting point of the Lombard effect // Acta Acust. United Acust. 2017. V. 103. № 1. P. 169–172. https://doi.org/10.3813/AAA.919043
12. *Garnier M., Dohen M., Lœvenbruck H., Welby P., Bailly L.* The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement? // Yehia H.C., Demolin D., Laboissiere R. (Eds.) Proceedings of ISSP 06. Ubatuba, Brazil. 2006. P. 255–262. HAL Id: hal-00214307
13. *Garnier M., Ménard L., Alexandre B.* Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? // J. Acoust. Soc. Am. 2018. V. 144. № 2. P. 1059–1074. https://doi.org/10.1121/1.5051321
14. *Bořil H., Hansen J.H.L.* Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments // IEEE ACM Trans. Audio Speech Lang. Process. 2010. V. 18. № 6. P. 1379–1393. https://doi.org/10.1109/TASL.2009.2034770
15. *Bollepalli B., Juvela L., Airaksinen M., Valentini-Botinhao C., Alku P.* Normal-to-Lombard adaptation of speech synthesis using long short-term memory recurrent neural networks // Speech Commun. 2019. V. 110. P. 64–75. https://doi.org/10.1109/ICASSP.2017.7953209
16. *Lee J., Ali H., Ziaei A., Tobey E.A., Hansen J.H.* The Lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study // J. Acoust. Soc. Am. 2017. V. 141. № 4. P. 2788–2799. https://doi.org/10.1121/1.4979927
17. *McColl D., McCaffrey P.* Perception of spasmodic dysphonia speech in background noise // Percept. Mot. Ski. 2006. V. 103. № 2. P. 629–635. https://doi.org/10.2466/pms.103.2.629-635
18. *Amazi D.K., Garber S.R.* The Lombard sign as a function of age and task // J. Speech Lang. Hear. Res. 1982. V. 25. № 4. P. 581–585. https://doi.org/10.1044/jshr.2504.581
19. *Tang P., Xu Rattanasone N., Yuen I., Demuth K.* Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech // J. Acoust. Soc. Am. 2017. V. 142. № 5. P. 2823–2835. https://doi.org/10.1121/1.5008372
20. *Junqua J.C., Anglade Y.* Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition // Proc. ICASSP. Albuquerque, NM. 1990. P. 841–844. https://doi.org/10.1109/ICASSP.1990.115969
21. *Tang P., Xu Rattanasone N., Yuen I., Demuth K.* Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech // J. Acoust. Soc. Am. 2017. V. 142. № 2. P. 493–503. https://doi.org/10.1121/1.4995998
22. *Якушев Д.И., Скляров О.П.* Моделирование гласных звуков // Акуст. журн. 2003. Т. 49 № 4. С. 567–569. https://doi.org/10.1134/1.1591305
23. *Кузнецов В.Б.* Спектральная динамика и классификация русских гласных // Акуст. журн. 2002. Т. 48. № 6. С. 849–853. https://doi.org/10.1134/1.1522046
24. *Фант Г.* Акустическая теория речеобразования. М.: Наука, 1964. 284 с.
25. *Чистович Л.А., Венцов А.В., Гранстрем М.П.* Физиология речи. Восприятие речи человеком. Л.: Наука, 1976. 388 с.
26. *Бондарко Л.В.* Фонетика современного русского языка. СПб: Изд-во С.-Петербург.ун-та, 1998. 276 с.
27. *Ляксо Е.Е., Григорьев А.С.* Динамика длительности и частотных характеристик гласных на протяжении первых семи лет жизни детей // Рос. физиол. журн. 2013. Т. 99. № 9. С. 1097–1110. eLIBRARY ID: 20260989
28. *Морозов В.П.* Биофизические основы вокальной речи. Л.: Наука, 1977. 232 с.
29. *Egan J.J.* Psychoacoustics of the Lombard voice response // J. Audit. Res. 1972. V. 12. P. 318–324.

30. *Matsumoto S., Akagi M.* Variation of formant amplitude and frequencies in vowel spectrum uttered under various noisy environments // Proc. NCSP2019, Honolulu. 2019. P. 4–7.
31. *Marcoux K., Ernestus M.* Pitch in native and non-native Lombard speech // Proc. ICPhS. Melbourne. 2019. P. 2605–2609.
32. *Ляксо Е.Е.* Некоторые характеристики материнской речи, адресованной младенцам первого полугодия жизни // Психол. журн. 2002. Т. 3. № 2. С. 55–64. eLIBRARY ID: 17315992
33. *Keith R.W.* Development and standardization of SCAN-C Test for Auditory Processing Disorders in Children // J. Am. Acad. Audiol. 2000. V. 11. № 8. P. 438–445.
34. *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 Acust. United Acust. 2019. V. 105. № 3. P. 484–491. https://doi.org/10.3813/AAA.919330
35. *Marks L.E.* Binaural summation of loudness: Noise and two-tone complexes // Percept. Psychophys. 1980. V. 27. № 6. P. 489–498. https://doi.org/10.3758/BF03198676
36. *Koterov A.N., Ushenkova L.N., Zubenkova E.S., Kalininna M.V., Biryukov A.P., Lastochkina E.M., Molodtsova D.V., Wainson A.A.* Strength of association. Report 2. Graduation of correlation size // Med. Radiol. Radiat. Saf. 2019. V. 64. № 6. P. 12–24. https://doi.org/10.12737/1024-6177-2019-64-6-12-24
37. *Sapir S., Ramig L.O., Spielman J.L., Fox C.* Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech // J. Speech. Lang. Hear. Res. 2010. V. 53. P. 114–125. https://doi.org/10.1044/1092-4388(2009/08-0184)
38. *Stowe L.M., Golob E.J.* Evidence that the Lombard effect is frequency-specific in humans // J. Acoust. Soc. Am. 2013. V. 134. № 1. P. 640–647. https://doi.org/10.1121/1.4807645
39. *Letowski T., Frank T., Caravella J.* Acoustical properties of speech produced in noise presented through supra-aural earphones // Ear Hear. 1993. V. 14. № 5. V. 332–338. https://doi.org/10.1097/00003446-199310000-00004
40. *Patel R., Schell K.W.* The influence of linguistic content on the Lombard effect // J. Speech Lang. Hear. 2008. V. 51. P. 209–221. https://doi.org/10.1044/1092-4388(2008/016)
41. *Alghamdi N., Maddock S., Marxer R., Barker J., Brown G.J.* A corpus of audio-visual Lombard speech with frontal and profile views // J. Acoust. Soc. Am. 2018. V. 143. № 6. P. 523–529. https://doi.org/10.1121/1.5042758
42. *Kleczkowski P., Żak A., Król-Nowak A.* Lombard effect in Polish speech and its comparison in English speech // Arch. Acoust. 2017. V. 42. № 4. P. 561–569. https://doi.org/10.1515/aoa-2017-0060
43. *Zhao Y., Ando A., Takaki S., Yamagishi J., Kobashikawa S.* Does the Lombard Effect Improve Emotional Communication in Noise? Analysis of Emotional Speech Acted in Noise // Proc. Interspeech. 2019. P. 3292–3296. DOI: arXiv:1903.12316.
44. *Russell A., Penny L., Pemberton C.* Speaking fundamental frequency changes over time in women: a longitudinal study // J. Speech Lang. Hear. Res. 1995. V. 38. № 1. P. 101–109. https://doi.org/10.1044/jshr.3801.101
45. *Titze I.R., Luschei E.S., Hirano M.* Role of the thyroarytenoid muscle in regulation of fundamental frequency // J. Voice. 1989. V. 3. № 3. P. 213–224. https://doi.org/10.1016/S0892-1997(89)80003-7
46. *Nishio M., Niimi S.* Changes in speaking fundamental frequency characteristics with aging // Folia Phoniatr. Logop. 2008. V. 60. № 3. V. 120–127. https://doi.org/10.1159/000118510
47. *Шиленкова В.В., Бестолкова О.С.* Пресбифония. Возрастные изменения акустических параметров голоса // Вестник оториноларингологии. 2013. Т. 78. № 6. С. 24–27. eLIBRARY ID: 21074035
48. *Коваленко А.Н., Кастыро И.В., Решетов И.В., Попадюк В.И.* Исследование роли слухопротезирования в формировании площади акустического поля гласных // Докл. Акад. наук. Науки о жизни. 2021. Т. 497. № 1. С. 204–208. https://doi.org/10.31857/S2686738921020141