

УДК 534.784+612.85+81.342

ВЛИЯНИЕ ВИЗУАЛЬНОЙ ОЦЕНКИ РАССТОЯНИЯ ДО СЛУШАТЕЛЯ НА ЧАСТОТУ ОСНОВНОГО ТОНА ГОЛОСА ДИКТОРА В УСЛОВИЯХ ШУМА

© 2023 г. А. М. Луничкин¹, *, А. П. Гвоздева¹, И. Г. Андреева¹

¹ФГБУН Институт эволюционной физиологии и биохимии
имени И.М. Сеченова РАН, Санкт-Петербург, Россия

*E-mail: BolverkDC@mail.ru

Поступила в редакцию 30.11.2022 г.

После доработки 24.01.2023 г.

Принята к публикации 31.01.2023 г.

Ломбардная речь представляет собой непроизвольные адаптивные изменения голосообразования под влиянием шума. В данной работе была проверена гипотеза о взаимодействии непроизвольного слухоречевого контроля, характерного для нее, и произвольного контроля фонации, который возникает в результате визуальной оценки расстояния до слушателя. Оценивали частоту основного тона голоса (ЧОТ) при визуальной оценке расстояния между диктором и слушателем (1 и 4 м) в условиях шума многоголосия (60 и 72 дБ) и при фронтальном расположении последнего. В работе принимали участие 9 русскоязычных дикторов — женщин 20–35 лет с нормальным слухом. Получили повышение значений ЧОТ дикторов при усложнении коммуникативной ситуации как в случае усиления уровня фонового шума, так и при увеличении расстояния между диктором и слушателем. В тишине и при двух уровнях шума после увеличения расстояния до слушателя прирост значений ЧОТ (Δ ЧОТ) составил 14, 18 и 15 Гц, соответственно, и достоверно не различался ($p > 0.05$, $n = 288$). При сохранении коммуникативной дистанции шум многоголосия уровнем 60 и 72 дБ приводил к достоверно различающимся значениям Δ ЧОТ: при 1 м — 14 и 32 Гц ($p < 0.001$, $n = 288$), а при 4 м — 18 и 33 Гц ($p < 0.001$, $n = 288$), соответственно. Полученные данные свидетельствуют о независимом и аддитивном влиянии факторов шума и коммуникативного расстояния на фонацию.

Ключевые слова: акустика речи, шум многоголосия, слухоречевой контроль, эффект Ломбарда, речевая коммуникация, коммуникация на расстоянии, частота основного тона голоса.

DOI: 10.31857/S0131164622600987, **EDN:** GBDNUW

В процессе общения дикторы изменяют параметры своей речи как непроизвольно, так и произвольно, применяя коммуникативную стратегию. Оба способа контроля речи направлены на то, чтобы облегчить передачу речевого сигнала собеседнику. Согласно современным представлениям управление голосовой моторикой описывают при помощи модели нейронной сети, которая включает в себя две структурно и функционально различающиеся части, каждая из которых реализует один из способов контроля речи [1]. При этом непроизвольный контроль обеспечивается филогенетически древней первичной голосовой моторной сетью, в основном расположенной в подкорковых структурах.

Непроизвольные адаптивные изменения речи, которые возникают в шумной среде, получили название эффекта Ломбарда [2, 3]. Причем наряду с амплитудными изменениями характеристик голоса, происходят и его спектральные изменения [4–6]. Эти изменения состоят в увеличении

частоты основного тона (ЧОТ) голоса и первых формант гласных звуков, а так же в разнонаправленном изменении вторых и последующих формант. Латеральная ретикулярная формация и периакведуктальное серое вещество (PAG) являются возможными слуховыми центрами в первичной вокально-моторной сети. Основным аргументом их участия в формировании эффекта Ломбарда являются многочисленные данные, полученные на млекопитающих [3, 7, 8].

Вторая часть нейросети представляет собой произвольную артикуляционную моторную сеть, берущую начало в вентролатеральной части префронтальной коры [1]. Взаимодействие двух частей нейросети будет приводить к тому, что изменения голоса диктора в условиях шума могут быть скорректированы произвольно. М. Garnier et al. обнаружили, что дикторы проявляют более сильный эффект Ломбарда в том случае, когда вовлечены в задачу интерактивного общения, и более слабый — при неинтерактивном чтении [9]. Эф-

эффект Ломбарда может быть существенно снижен или даже заблокирован при произвольном подержании уровня своего голоса в шуме с помощью обратной связи, которая позволяет визуализировать усилие голоса [10, 11]. У профессиональных хоровых певцов наблюдают более слабый эффект Ломбарда, чем у непрофессионалов, что обусловлено тренировкой [12, 13]. Таким образом, кортикальные процессы могут модулировать эффект Ломбарда как в направлении его активации, так и торможения.

По-видимому, на произвольный контроль может также оказывать влияние зрительная и слуховая информация о взаимном расположении диктора и слушателя. Эта информация позволяет оценить расстояние между ними и соответственно произвольно (осознанно) скорректировать силу голоса, поскольку увеличение коммуникативного расстояния приводит к снижению уровня звукового давления, создаваемого голосом диктора в месте расположения слушателя [14, 15].

В отличие от ситуации увеличения коммуникативного расстояния, рост уровня окружающего шума не только ухудшает разборчивость речи для слушателя за счет снижения соотношения речь/шум в месте прослушивания, но и затрудняет слухоречевой контроль собственной речи диктором. Таким образом, для диктора фактор шума может быть более сильным и оказывать большее влияние на параметры его голоса, чем фактор коммуникативного расстояния. Это предположение частично подтверждается результатами работы [14], в которой было продемонстрировано, что собеседники в большей степени склонны изменять амплитудные характеристики голоса в соответствии с уровнем окружающего шума, чем с изменением коммуникативного расстояния. Вместе с тем, вопрос о соотношении влияния факторов шума и расстояния при их совместном воздействии на частотные характеристики голоса остается открытым.

Целью настоящего исследования было сопоставить непроизвольный контроль голоса диктора в шуме и произвольный контроль его голоса при визуальной оценке расстояния до слушателя, находящегося в условиях сложной акустической сцены. Оценивали влияние на процесс фонации, в котором основную роль играют связи между первичной вокально-моторной сетью и первичным слуховым путем [16, 17]. Изменения фонации отражаются в ЧОТ, представляющей собой частоту колебания голосовых связок. По этой причине ЧОТ применяют в работах, направленных на оценку отдельных эффектов коммуникативного расстояния [14, 15] или окружающего шума [9, 18]. Визуальная оценка расстояния в настоящей работе была обеспечена при помощи метода “вызванной речи”, при котором диктор, об-

ращаясь к слушателю, повторял набор строго определенных слов. По сравнению со свободной коммуникацией данный метод исключал эмоциональную составляющую речи, неизбежно возникающую в диалоге.

МЕТОДИКА

В записи речи принимали участие 9 дикторов-женщин в возрасте 25–35 лет. Все дикторы являлись носителями русского языка и не имели нарушений слуха и дефектов речи.

Все дикторы, принимавшие участие в записях, проходили обследование слуха, которое включало тональную пороговую аудиометрию и тест обнаружения паузы. Последний позволял оценить временную разрешающую способность слуха дикторов для чистых тонов частотой 0,5, 1, 2 и 4 кГц, и широкополосных шелчков [19]. У всех дикторов пороги обнаружения паузы не превышали 20 мс, а пороги слуха на основных аудиометрических частотах — 20 дБ. Эти показатели соответствовали нормам тонального слуха и временной разрешающей способности.

Речевой материал. В качестве речевого материала применяли 8 двусложных слов русского языка: с ударением на первый слог — “почва”, “ручка”, “выпуск”, “плата”; с ударением на второй слог — “рубеж”, “строка”, “набор”, “кредит”.

Маскирующий сигнал. В качестве маскирующего сигнала использовали шум многоголосия (далее — шум), который создавали микшируя записи слов, описанных в разделе Речевой материал, произнесенных четырьмя дикторами, которые не принимали участие в данном исследовании (двумя мужчинами — 30 и 65 лет, и двумя женщинами — 19 и 60 лет). Частоты основного тона голоса для дикторов-мужчин составляли 117 ± 8 и 139 ± 9 Гц, соответственно, а для женщин — 208 ± 30 и 234 ± 34 Гц, соответственно. Длительность записей слов составляла от 400 до 800 мс. Для каждого слова каждого диктора создавалась звуковая дорожка с многократным повтором слова без пауз. Таким способом были получены 32 звуковые дорожки (8 слов $\times$ 4 диктора). Для создания шума они были микшированы в файл длительностью 40 с. Полученный шум многоголосия нормализовали по уровню, после чего сформировали линейные фронты нарастания и убывания интенсивности по 1 с. Данная методика формирования шума в подробностях описана в работе [20]. Шум многоголосия, при его предъявлении в наушниках, создавал у слушателя ощущение нахождения в комнате, где одновременно говорит множество людей. Измеренный в А-взвешенном режиме уровень звукового давления шума при монофоническом воспроизведении составлял 52 и 64 дБ (А). При диотическом предъявлении шума в условиях экс-

перимента этот уровень соответствовал воспринимаемому уровню громкости моноаурально подаваемого шума 60 и 72 дБ, что обусловлено эффектом бинауральной суммации громкости [21]. Далее по тексту уровни шума указаны с учетом этого эффекта.

Оборудование и экспериментальное помещение. Исследование проводили в анэхоидной звукоизолированной камере объемом 62,5 м³. 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 аналогово-цифровой преобразователь (звуковую карту) с частотой дискретизации 44100 Гц и разрядностью 16 бит. Устройство *Rode NT-USB* было оснащено выходом, к которому подключали головные телефоны закрытого типа *Sennheiser HD-380-Pro*, использовавшееся для подачи диктору шумового сигнала и организации слуховой обратной связи (самопрослушивания, см. далее). Измерение уровня звукового давления шума осуществляли при помощи шумомера *RFT 000014* в А-взвешенном режиме при моноауральной подаче звука.

Настройка самопрослушивания. Головные телефоны уменьшали воспринимаемую диктором громкость собственной речи, поэтому их применение могло влиять на параметры речи диктора (например, ее интенсивность и ЧОТ). Для снижения этого эффекта перед началом эксперимента диктора просили выполнить регулировку самопрослушивания. С этой целью диктор настраивал усиление подаваемого в наушники 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. Настроенное диктором перед началом эксперимента усиление собственного голоса (самопрослушивание) 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 губ, устанавливали микрофонное устройство на напольной стойке. Жесткая фиксация положения головы не применялась, однако во время эксперимента диктора просили держать затылок прижатым к подголовнику, а также не наклонять и не поворачивать голову. На дикторе были надеты головные телефоны закрытого типа. После настройки са-

мопрослушивания выполняли запись слов. Схема экспериментальной процедуры представлена на рис. 1. Каждый диктор участвовал в четырех сессиях записи, различавшиеся набором слов и дистанцией до слушателя (экспериментатора). Между сессиями диктор молчал. В течение одной сессии диктор, обращаясь к экспериментатору, лицо которого мог видеть, произносил по четыре раза друг за другом четыре слова в следующих условиях: 1) в тишине, 2) в шуме многоголосия при его уровне 60 дБ, 3) в шуме многоголосия при его уровне 72 дБ. Во время первых двух сессий экспериментатор располагался напротив диктора на расстоянии 1 м, во время вторых двух сессий — на расстоянии 4 м. Для того, чтобы уменьшить различия в интонировании слов, дикторов просили перед каждым словом добавлять местоимение “это”: “это — кредит, это — набор, это — выпуск, это — плата” и т.д. Во время записи слов слушатель поддерживал зрительный контакт с диктором, сохраняя молчание и не комментируя каким-либо образом произнесение диктором слов. После окончания сессии ее сохраняли в формате “wav”. Для каждого из девяти дикторов было записано 192 звуковых фрагмента (8 слов × 4 повтора × 3 условия × 2 расстояния). Общий объем речевого материала для девяти дикторов составил 1728 записей слов.

Анализ записей и методы статистического анализа. Определение ЧОТ дикторов производили в программе *Praat* (свободно распространяемое программное обеспечение, www.praat.org). В окне программы в каждом записанном звуковом фрагменте по рисунку звуковой волны определяли начало и конец произнесенного диктором слова, игнорируя местоимение “это” и любые следовые шумы, создаваемые диктором после озвучивания слова. Значение ЧОТ рассчитывали при помощи функции “*get pitch*”, как среднее всех значений в слове. Частотный диапазон для оценки ЧОТ составлял 75–500 Гц. Статистическую обработку и визуализацию полученных данных проводили в программах *Excel* и *Statistica 10*. Оценку достоверности различий величины ЧОТ у отдельных дикторов в разных условиях шумовой нагрузки и при разных коммуникативных расстояниях выполняли при помощи непараметрического парного критерия Вилкоксона, внося поправку Бонферрони на множественные сравнения. Контрольным условием считали говорение в тишине при расстоянии между диктором и слушателем соответствовавшем 1 м. Выполняли расчет изменений частоты основного тона (Δ ЧОТ) по сравнению с контролем в следующих условиях: 1) шум 60 дБ, расстояние 1 м; 2) шум 72 дБ, расстояние 1 м; 3) тишина, расстояние 4 м; 4) шум 60 дБ, расстояние 4 м; 5) шум 72 дБ, расстояние 4 м. Кроме того, вычисляли изменения ЧОТ с увеличением расстояния при одинаковой шумовой нагрузке или в

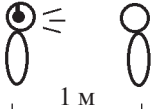
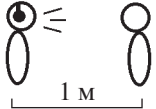
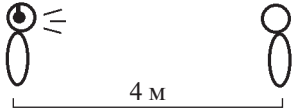
	Тишина	Шум 60 дБ	Шум 72 дБ	
Сессия 1	(C1, C2, C3, C4) × 4	(C1, C2, C3, C4) × 4	(C1, C2, C3, C4) × 4	
Сессия 2	(C5, C6, C7, C8) × 4	(C5, C6, C7, C8) × 4	(C5, C6, C7, C8) × 4	
Сессия 3	(C1, C2, C3, C4) × 4	(C1, C2, C3, C4) × 4	(C1, C2, C3, C4) × 4	
Сессия 4	(C5, C6, C7, C8) × 4	(C5, C6, C7, C8) × 4	(C5, C6, C7, C8) × 4	

Рис. 1. Схема эксперимента по записи речи дикторов.

Каждый эксперимент состоял из четырех сессий записи двусложных слов русского языка (слова обозначены как C1, C2...C8). В сессиях 1 и 2 расстояние до слушателя составляло 1 м, в сессиях 3 и 4 — 4 м. В каждой сессии дикторы повторяли четыре слова друг за другом по четыре раза в тишине, в шуме 60 дБ и в шуме 72 дБ, обращаясь к слушателю.

ее отсутствии (тишина) — $\Delta\text{ЧОТ}(r)$, а также с увеличением шумовой нагрузки от 0 до 60 и от 0 до 72 дБ при неизменном коммуникативном расстоянии — $\Delta\text{ЧОТ}(l)$. Массив данных о ЧОТ и $\Delta\text{ЧОТ}$ проверяли на нормальность распределения при помощи теста Колмогорова–Смирнова и на однородность дисперсии при помощи теста Левена. Различия между значениями ЧОТ, $\Delta\text{ЧОТ}$, рассчитанными относительно контрольного значения, между $\Delta\text{ЧОТ}(r)$ и между $\Delta\text{ЧОТ}(l)$ для всей группы дикторов оценивали при помощи непараметрического парного критерия Вилкоксона. Вносили поправку Бонферрони на множественные сравнения.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для коммуникативных ситуаций, характеризующихся разной степенью сложности и ролью произвольного и непроизвольного контроля голоса, были определены значения ЧОТ голоса русской речи девяти дикторов-женщин (табл. 1). Были вычислены изменения ЧОТ относительно контроля ($\Delta\text{ЧОТ}$), изменения уровня шумовой нагрузки $\Delta\text{ЧОТ}(r)$ и изменения фактора расстояния $\Delta\text{ЧОТ}(l)$. При проверке тестом Колмогорова–Смирнова массивов данных для ЧОТ и $\Delta\text{ЧОТ}$ установили отсутствие нормального распределе-

ния ($p < 0.01$, $n = 1728$), а тест Левена показал отсутствие в выборках соответствующих разным условиям равной дисперсии ($p < 0.01$, $n = 288$), поэтому применение дисперсионного анализа ANOVA представлялось некорректным.

На рис. 2 представлены примеры изменений значений ЧОТ в слове “Рубеж”, произнесенного диктором № 3 при разной шумовой нагрузке и расстоянии до слушателя. По данным рис. 2 видно, что усложнение ситуации коммуникации, по сравнению с контролем за счет усиления шума и/или увеличения коммуникативной дистанции, приводило к росту значений ЧОТ дикторов при произнесении слова. В дальнейшем были проанализированы средние значения ЧОТ в слове. Расчет средних значений ЧОТ всегда осуществляли на всем протяжении слова, несмотря на то, что при усложнении коммуникативной ситуации наблюдалось увеличение длительности слов.

Диктор № 1 менял ЧОТ в незначительной степени, при этом достоверные отличия от контрольного условия были получены только в многоголосии уровнем 72 дБ при коммуникативном расстоянии 4 м ($p < 0.01$, $n = 32$, здесь и далее используется непараметрический парный критерий Вилкоксона). У остальных восьми дикторов любое усложнение коммуникативной ситуации вызывало достоверное увеличение ЧОТ относительно

Таблица 1. Индивидуальные значения частоты основного тона голоса (ЧОТ) дикторов при произнесении слов в условиях с разной шумовой нагрузкой и разным расстоянии до слушателя

Диктор (№/возраст)	Частота основного тона, Гц					
	1 м			4 м		
	тишина 0 дБ	шум 60 дБ	шум 72 дБ	тишина 0 дБ	шум 60 дБ	шум 72 дБ
№ 1/20	229 ± 23*	233 ± 14	237 ± 9	231 ± 7	233 ± 7	237 ± 5
№ 2/20	186 ± 23	192 ± 13	217 ± 15	202 ± 17	220 ± 14	242 ± 14
№ 3/20	196 ± 21	240 ± 10	274 ± 9	235 ± 13	280 ± 14	311 ± 19
№ 4/26	199 ± 13	209 ± 12	225 ± 13	209 ± 13	229 ± 15	237 ± 15
№ 5/28	198 ± 16	214 ± 17	258 ± 8	218 ± 8	244 ± 10	275 ± 13
№ 6/29	186 ± 11	205 ± 5	217 ± 10	203 ± 14	215 ± 8	223 ± 11
№ 7/30	201 ± 29	210 ± 31	214 ± 27	206 ± 23	226 ± 39	227 ± 35
№ 8/33	175 ± 15	187 ± 12	204 ± 13	184 ± 15	201 ± 14	218 ± 13
№ 9/35	157 ± 12	164 ± 11	169 ± 12	169 ± 18	173 ± 13	183 ± 3

Примечание: * — приведено среднее значение и стандартное отклонение по всем записям диктора для 8 слов и 4 их повторений ($n = 32$).

контроля ($p < 0.01$, $n = 32$). Индивидуальное повышение ЧОТ в усложненных условиях коммуникации между дикторами различалось. Наименьшая степень увеличения обнаружена у диктора № 1 — с 229 Гц в контроле до 237 Гц при многоголосии 72 дБ, расстоянии 4 м; наибольшая — у диктора № 3, значения ЧОТ которого в контроле и в шуме многоголосия при расстоянии 4 м составили 196 и

311 Гц, соответственно. Значения ЧОТ дикторов в шуме 72 дБ были больше, чем в условиях многоголосия 60 дБ с равной дистанцией до слушателя ($p < 0.01$, $n = 32$). При одинаковом уровне помехи индивидуальные ЧОТ дикторов № 2–9 были выше в случае, когда расстояние составляло 4 м ($p < 0.01$, $n = 32$), в то время как у диктора № 1 расстояние не влияло на ЧОТ при постоянном уровне шума.

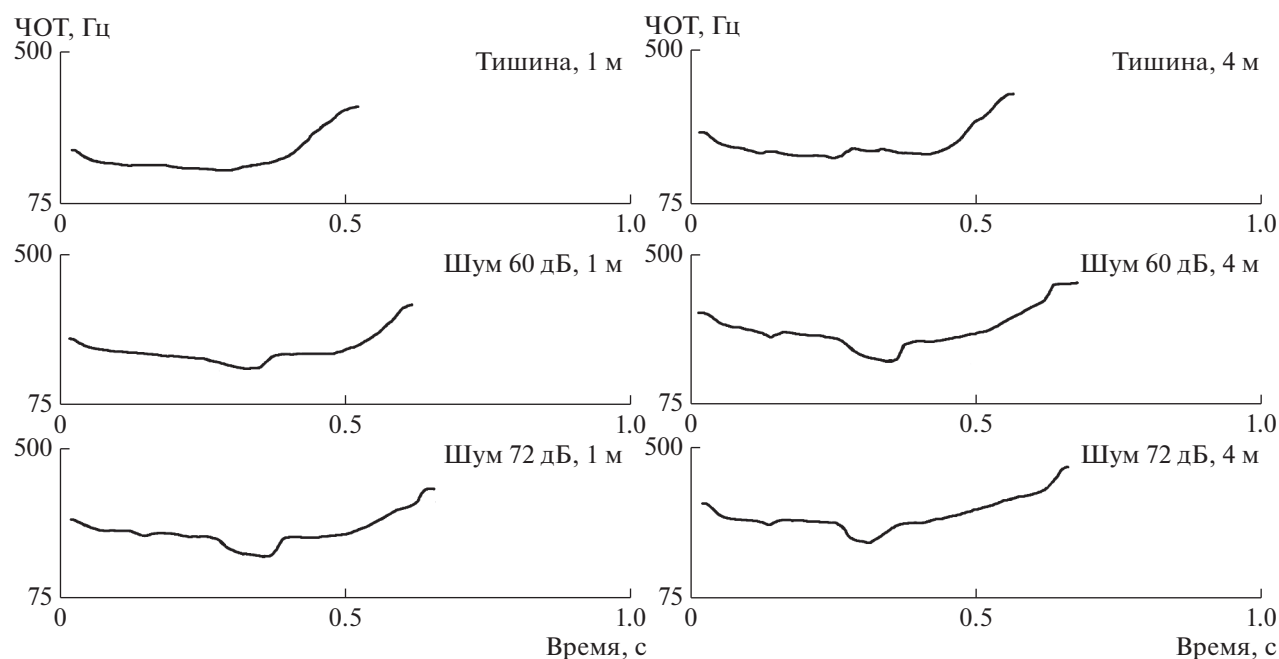


Рис. 2. Примеры изменения значений частоты основного тона при произнесении слова “Рубеж” диктором № 3 в разных экспериментальных условиях.

По оси абсцисс — время произнесения, с; по оси ординат — частота основного тона, Гц.

Индивидуальная вариабельность значений ЧОТ также различалась у разных дикторов. У части из них (например, дикторы № 1, 2, 8) затруднение коммуникативной ситуации приводило к снижению вариабельности показателей частоты основного тона. У других дикторов вариабельность ЧОТ при усилении помехи, напротив, увеличивалась (диктор № 4) или же не имела однозначного характера изменений (дикторы № 3, 7).

При последовательном усложнении акустической сцены для слушателя менялся диапазон значений ЧОТ дикторов. В контроле он составлял от 157 до 229 Гц (72 Гц), тогда как в условиях наиболее сложной коммуникативной ситуации (многоголосие 72 дБ, 4 м) диапазон возрастал на треть по сравнению с контролем — от 218 до 311 Гц (93 Гц). Это расширение диапазона в сложной коммуникативной ситуации было обусловлено тем, что под влиянием факторов разные дикторы меняли голос в различной степени. Учитывая это, влияние факторов на голос определяли не по абсолютным значениям ЧОТ, а по их изменениям в условиях разных акустических сцен.

При сохранении уровня помехи, средний показатель ЧОТ в группе был выше в случае, когда коммуникативное расстояние равнялось 4 м ($p < 0.001$ для тишины, $p < 0.001$ для шума 60 дБ, $p < 0.001$ для шума 72 дБ, $n = 288$). Значения $\Delta\text{ЧОТ}(r)$ по группе составили 14 ± 4 Гц в тишине, 18 ± 4 Гц в многоголосии уровнем 60 дБ, 15 ± 4 Гц в многоголосии уровнем 72 дБ (рис. 3). При этом анализ полученных значений $\Delta\text{ЧОТ}(r)$ показал, что различий между $\Delta\text{ЧОТ}(r)$ в разных условиях не выявлялось ($p > 0.05$ для тишины и 60 дБ; $p > 0.05$ для тишины и 72 дБ; $p > 0.05$ для 60 и 72 дБ, $n = 288$).

Значения $\Delta\text{ЧОТ}(l)$, полученные при усилении шумовой нагрузки с тишины до 60 дБ и с тишины до 72 дБ, достоверно различались между собой как при дистанции 1 м ($p < 0.001$, $n = 288$), так и при дистанции 4 м ($p < 0.001$, $n = 288$) (рис. 4). При расстоянии между диктором и слушателем равном 1 м усиление многоголосия от 0 до 60 дБ приводило к средним значениям $\Delta\text{ЧОТ}(l)$ 14 ± 4 Гц, до 72 дБ — 32 ± 8 Гц. При удалении слушателя на 4 м $\Delta\text{ЧОТ}(l)$ были равны 18 ± 4 и 33 ± 7 дБ, соответственно. Важно отметить, что $\Delta\text{ЧОТ}(l)$ для двух коммуникационных расстояний при говорении в равном шуме не различались (при шуме 60 дБ $p > 0.05$, $n = 288$; при шуме 72 дБ $p > 0.05$, $n = 288$).

Совместное действие факторов шумовой помехи и расстояния по отношению к контрольным условиям приводило к $\Delta\text{ЧОТ}$ равному 32 ± 8 Гц в шуме 60 дБ при дистанции 4 м; 47 ± 11 Гц в шуме 72 дБ при дистанции 4 м. Полученные значения статистически различались ($p < 0.001$, $n = 288$).

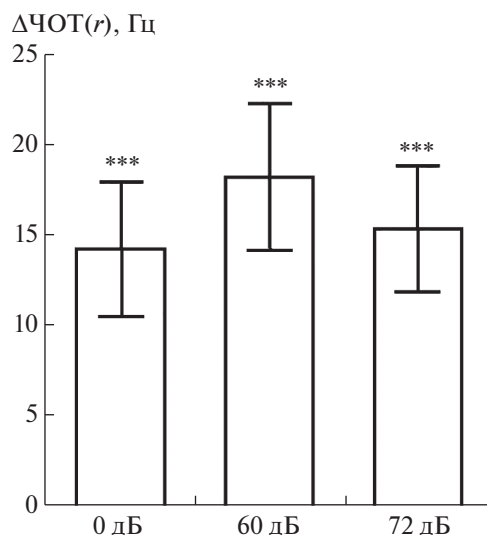


Рис. 3. Средние значения изменения частоты основного тона в группе из девяти дикторов при увеличении расстояния между диктором и слушателем от 1 до 4 м при сохранении уровня помехи ($\Delta\text{ЧОТ}(r)$). По оси абсцисс — условия говорения: 0 дБ — тишина; 60 дБ — многоголосие уровня 60 дБ; 72 дБ — многоголосие уровня 72 дБ. По оси ординат — $\Delta\text{ЧОТ}(r)$, Гц. *** — достоверно значимый прирост $\Delta\text{ЧОТ}(r)$ при расстоянии 4 м относительно 1 м ($p < 0.001$, $n = 288$, непараметрический парный критерий Вилкоксона).

Анализ значений $\Delta\text{ЧОТ}$ в коммуникативных ситуациях, усложненных по сравнению с контролем, показал сходство изменений ЧОТ при ряде сочетаний двух факторов. Значения $\Delta\text{ЧОТ}$, полученные при говорении в многоголосии уровнем 60 дБ при коммуникативном расстоянии 1 м и тишине при расстоянии 4 м, не различались между собой ($p > 0.05$, $n = 288$). Значения $\Delta\text{ЧОТ}$ при предъявлении шума уровнем 72 дБ с дистанцией 1 м и шума 60 дБ дистанцией 4 м, также достоверно не различались ($p > 0.05$, $n = 288$). Наибольшее значение прироста ЧОТ было получено при шумовой нагрузке уровнем 72 дБ и увеличении коммуникативного расстояния до 4 м. В этом сочетании факторов $\Delta\text{ЧОТ}$ было достоверно выше, чем в остальных изученных ($p < 0.001$, $n = 288$).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Изменения частоты основного тона голоса с расстоянием оказались практически одинаковыми в условиях тишины и при обоих уровнях шума. Полученные значения $\Delta\text{ЧОТ}(r)$ в тишине составили по средним данным 14 Гц. При росте коммуникативного расстояния $\Delta\text{ЧОТ}(r)$ увеличивался на 18 и 15 Гц в условиях шума 60 и 72 дБ, соответственно. Различие в $\Delta\text{ЧОТ}(r)$ для всех трех случаев было недостоверным. Таким образом, шум среднего уровня интенсивности не влиял на из-

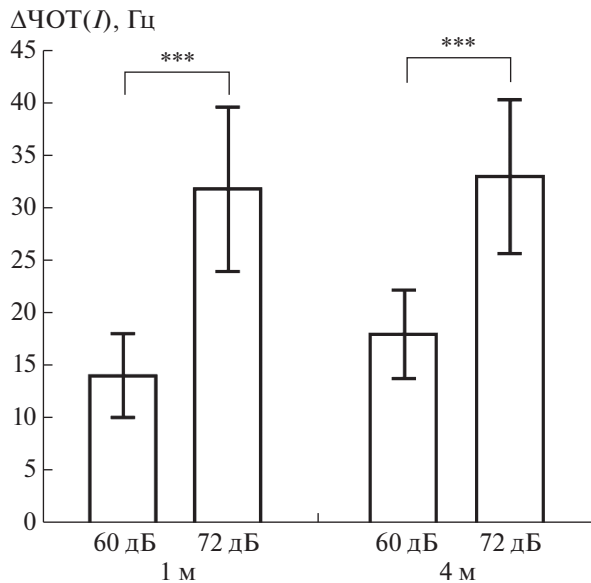


Рис. 4. Средние значения изменения частоты основного тона в группе из девяти дикторов при усилении шумовой нагрузки с тишины до 60 дБ и с тишины до 72 дБ при неизменном коммуникативном расстоянии ($\Delta\text{ЧОТ}(Л)$).

По оси абсцисс — условия говорения: 0 дБ — тишина; 60 дБ — многоголосие уровня 60 дБ; 72 дБ — многоголосие уровня 72 дБ; 1 и 4 м — коммуникативное расстояние. По оси ординат — $\Delta\text{ЧОТ}(Л)$, Гц. *** — достоверные различия между уровнями шумовой нагрузки 60 и 72 дБ ($p < 0.001$, $n = 288$, непараметрический парный критерий Вилкоксона).

менения фонации при увеличении расстояния до слушателя.

При увеличении уровня шума от 60 до 72 дБ получили близкие значения увеличения $\Delta\text{ЧОТ}(Л)$ — 15 и 18 Гц для обоих расстояний 1 и 4 м, соответственно. Эти значения $\Delta\text{ЧОТ}(Л)$ достоверно не различались, что указывает на отсутствие влияния расстояния до слушателя на произвольные изменения фонации при повышении уровня шума.

Таким образом, при росте коммуникативного расстояния и при повышении уровня шумовой помехи регистрировали увеличение ЧОТ голоса, что соответствовало известным данным литературы при раздельной оценке действия этих факторов [9, 14, 15, 18]. Причем, сходство значений $\Delta\text{ЧОТ}(Л)$ и $\Delta\text{ЧОТ}(r)$ свидетельствует о близком по величине воздействию двух разных факторов — шума и увеличения расстояния, затрудняющих коммуникацию и влияющих на фонацию разными путями, т.е. посредством произвольного и произвольного контроля. Значения обоих факторов, усложняющих коммуникацию, были подобраны нами в соответствии с представлением о том, что увеличение расстояния в четыре раза должно уменьшать уровень звука в месте прослушивания на 12 дБ. Согласно данным литературы

удвоение расстояния снижает уровень сигнала на 6 дБ в условиях свободного поля [22].

В случае увеличения шумовой нагрузки снижается слуховой контроль диктором собственного голоса из-за нарушения самопрослушивания, причем изменения фонации возникают произвольно. В обычных условиях, когда шум присутствует в окружающей среде, а не подается в наушники, как в нашем эксперименте, он также затрудняет распознавание речи для слушателя. В случае увеличения коммуникативного расстояния, нарушения обратной слухоречевой связи не происходит, но у слушателя возникают затруднения, обусловленные снижением уровня поступающего сигнала. Это усложнение коммуникативной ситуации может вызывать адаптивные изменения голоса диктора только за счет произвольного контроля при визуальной оценке расстояния, т.к. слушатель по условиям эксперимента молчит на протяжении всего периода регистрации голоса. Таким образом, используя разные способы контроля голоса и, тем самым, задействуя разные нейронные сети, мы при сходном усложнении коммуникативной ситуации получили одинаковые изменения голоса, что свидетельствует о независимом влиянии на фонацию произвольного и произвольного механизма контроля голоса. Обратного эффекта, т.е. снижения ЧОТ при усложнении коммуникативной ситуации, а, следовательно, неспособности диктора менять фонацию адекватно меняющимся условиям, обнаружено не было. Напротив, было получено аддитивное влияние изученных факторов на частоту основного тона, т.е. увеличение шума на 12 дБ при дистанции 1 м давало тот же эффект, что и увеличение расстояния с 1 до 4 м.

Полученные нами значения $\Delta\text{ЧОТ}(r)$ в тишине составили по средним данным 14 Гц и согласовывались с изменениями частоты для сходного диапазона расстояний. Наша оценка приращения частоты основного тона оказалась выше, чем в работе [18] — около 7 Гц, но ниже чем в работе [23] — около 35–39 Гц. В еще одной работе [24] наблюдался широкий разброс индивидуальных значений $\Delta\text{ЧОТ}$ от 8 до 40 Гц, что согласуется с обнаруженным нами высоким уровнем межиндивидуальной вариативности. Различия данных, полученных в разных исследованиях, можно связать с рядом факторов. Во-первых, с условиями записи голоса: в работе [24] она производилась на улице, на фермерском поле, в нашем исследовании и работах [18, 23] в звукозаглушенных анаэодных камерах. В работе [18] показано, что прирост ЧОТ в камере был выше на 4 Гц, чем в других исследованных помещениях. Во-вторых, с полом дикторов — в работе [18] изучали голоса только мужчин, в то время как в работе [23] мужчина увеличивал ЧОТ в меньшей степени, чем женщины. В-третьих, необходимо учитывать индивидуальные особенности

дикторов и малые выборки. В записи голоса [24] принимало участие четыре человека, в [23] — всего три, тогда как нами показана высокая вариативность Δ ЧОТ в группе дикторов. Таким образом, оценка изменений ЧОТ при увеличении расстояния, выполненная в нашем исследовании с применением метода вызванной речи, показала хорошее соответствие с результатами ранее выполненных работ.

ЗАКЛЮЧЕНИЕ

Речепродукция включает в себя согласованный контроль множества нейромоторных систем, включая артикуляцию, фонацию и дыхание. Контроль фонации обеспечивается прямой и обратной связью. Если обнаруживается несоответствие между прямой и слуховой обратной связью, то к речевым мышцам посылаются корректирующие команды для подстройки голоса. Во время фонации в шумной обстановке возникает несоответствие между прямой и слуховой обратной связью [25, 26]. Данные авторов настоящей статьи свидетельствуют о том, что произвольный контроль под влиянием визуальной оценки расстояния может вызывать сходные по величине с действием шума изменения фонации. Причем при средних уровнях шума и типичных коммуникативных расстояниях оба фактора — шум и коммуникативное расстояние, действуют независимо и аддитивно.

Этические нормы. Все исследования проведены в соответствии с принципами биомедицинской этики, сформулированными в Хельсинкской декларации 1964 г. и ее последующих обновлениях, и одобрены локальным этическим комитетом Института эволюционной физиологии и биохимии им. И.М. Сеченова (Санкт-Петербург).

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

Финансирование работы. Данное исследование поддержано средствами РНФ (проект № 22-25-00068).

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

Вклад авторов в публикацию. А.М. Луничкин, И.Г. Андреева, А.П. Гвоздева — идея работы и планирование эксперимента, А.П. Гвоздева — подготовка методики, А.М. Луничкин — сбор и обработка данных, А.М. Луничкин, И.Г. Андреева, А.П. Гвоздева — написание и редактирование манускрипта.

СПИСОК ЛИТЕРАТУРЫ

1. Hage S.R., Nieder A. Dual neural network model for the evolution of speech and language // *Trends Neurosci.* 2016. V. 39. № 12. P. 813.
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.
3. Luo J., Hage S.R., Moss C.F. The Lombard effect: from acoustics to neural mechanisms // *Trends Neurosci.* 2018. V. 41. № 12. P. 938.
4. Liénard J.S., Di Benedetto M.G. Effect of vocal effort on spectral properties of vowels // *J. Acoust. Soc. Am.* 1999. V. 106. № 1. P. 411.
5. Traunmüller H., Eriksson A. Acoustic effects of variation in vocal effort by men, women, and children // *J. Acoust. Soc. Am.* 2000. V. 107. № 6. P. 3438.
6. Koenig L.L., Fuchs S. Vowel formants in normal and loud speech // *J. Speech Lang. Hear. Res.* 2019. V. 62. № 5. P. 1278.
7. Nonaka S., Takahashi R., Enomoto K. et al. Lombard reflex during PAG-induced vocalization in decerebrate cats // *Neurosci. Res.* 1997. V. 29. № 4. P. 283.
8. 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.
9. Garnier M., Henrich N., Dubois D. Influence of Sound Immersion and Communicative Interaction on the Lombard Effect // *J. Speech Lang. Hear. Res.* 2010. V. 53. № 3. P. 588.
10. Pick H.L., Jr., Siegel G.M., Fox P.W. et al. Inhibiting the Lombard effect // *J. Acoust. Soc. Am.* 1989. V. 85. № 2. P. 894.
11. Therrien A.S., Lyons J., Balasubramaniam R. Sensory attenuation of self-produced feedback: the Lombard effect revisited // *PLoS One.* 2012. V. 7. № 11. P. e49370.
12. Tonkinson S. The Lombard effect in choral singing // *J. Voice.* 1994. V. 8. № 1. P. 24.
13. Bottalico P., Graetzer S., Hunter E.J. Effect of training and level of external auditory feedback on the singing voice: volume and quality // *J. Voice.* 2016. V. 30. № 4. P. 434.
14. Weisser A., Miles K., Richardson M.J., Buchholz J.M. Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments // *J. Acoust. Soc. Am.* 2021. V. 149. № 4. P. 2896.
15. Fux T., Feng G., Zimpfer V. Talker-to-listener distance effects on the variations of the intensity and the fundamental frequency of speech / 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 22–27, 2011. P. 4964.
16. Hage S.R., Jürgens U. On the role of the pontine brainstem in vocal pattern generation: a telemetric single-unit recording study in the squirrel monkey // *J. Neurosci.* 2006. V. 26. № 26. P. 7105.
17. Pieper F., Jürgens U. Neuronal activity in the inferior colliculus and bordering structures during vocalization in the squirrel monkey // *Brain Res.* 2003. V. 979. № 1–2. P. 153.

18. *Pelegrín-García D., Smits B., Brunskog J., Jeong C.H.* Vocal effort with changing talker-to-listener distance in different acoustic environments // *J. Acoust. Soc. Am.* 2011. V. 129. № 4. P. 1981.
19. *Keith R.W.* Development and standardization of SCAN-C Test for Auditory Processing Disorders in Children // *J. Acoust. Soc. Am.* 2000. V. 11. № 8. P. 438.
20. *Andreeva I.G., Dymnikowa M., Gvozdeva A.P. et al.* 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.
21. *Marks L.E.* Binaural summation of loudness: Noise and two-tone complexes // *Percept. Psychophys.* 1980. V. 27. № 6. P. 489.
22. *Coleman P.D.* An analysis of cues to auditory depth perception in free space // *Psychol. Bull.* 1963. V. 60. № 3. P. 302.
23. *Shih C., Lu H.Y.D.* Effects of talker-to-listener distance on tone // *J. Phon.* 2015. V. 51. P. 6.
24. *Cheyne H.A., Kalgaonkar K., Clements M., Zurek P.* Talker-to-listener distance effects on speech production and perception // *J. Acoust. Soc. Am.* 2009. V. 126. № 4. P. 2052.
25. *Meekings S., Evans S., Lavan N. et al.* Distinct neural systems recruited when speech production is modulated by different masking sounds // *J. Acoust. Soc. Am.* 2016. V. 140. № 1. P. 8.
26. *Meekings S., Scott S.K.* Error in the Superior Temporal Gyrus? A Systematic Review and Activation Likelihood Estimation Meta-Analysis of Speech Production Studies // *J. Cogn. Neurosci.* 2021. V. 33. № 3. P. 422.

The Impact of Visual Estimates of Talker-to-Listener Distance on Fundamental Frequency in Noise

A. M. Lunichkin^{a,*}, A. P. Gvozdeva^a, I. G. Andreeva^a

^a*Sechenov Institute of Evolutionary Physiology and Biochemistry RAS, St. Petersburg, Russia*

^{*}*E-mail: BolverkDC@mail.ru*

Lombard speech is an involuntary adaptive changes in voicing under the influence of noise. In the current work we examine the relationship between involuntary auditory-speech control, common for the Lombard speech, and voluntary control of phonation which occurs as a result of the visual estimation of the distance to the listener. Fundamental frequencies (F0) were estimated in 9 Russian normally hearing female speakers aged 20–35 years. An increase in F0 was obtained when the communicative conditions became more complex in both increase in surrounding background noise level and growth of talker-to-listener distance cases. In quiet and in noise of 60 and 72 dB the increase in talker-to-listener distance led to F0 increments of 14, 18 and 15 Hz, which did not differ significantly from each other ($p > 0.05$, $n = 288$). When the communicative distance held constant, babble noise of different levels led to significantly different values of $\Delta F0$: for 1 m distance – 14 and 32 Hz ($p < 0.001$, $n = 288$), and for 4 m – 18 and 33 Hz ($p < 0.001$, $n = 288$), respectively. The data obtained evidence independent and additive impact of noise and communicative distance on phonation.

Keywords: speech acoustics, multi-talker babble noise, auditory-speech control, the Lombard effect, speech communication, communication at distance, voice fundamental frequency.