

РЕЧЬ ВЗРОСЛЫХ В РАЗНЫХ ЭМОЦИОНАЛЬНЫХ СОСТОЯНИЯХ: ВРЕМЕННЫЕ И СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ

© 2024 г. А. В. Куражова^{a, b, *}

^aСанкт-Петербургский государственный университет,
Университетская наб. 7/9, Санкт-Петербург, 199034 Россия

^bСанкт-Петербургский государственный педиатрический медицинский университет,
Литовская 2, Санкт-Петербург, 194100 Россия

*e-mail: avk_spb@bk.ru

Поступила в редакцию 21.03.2023 г.

После доработки 21.06.2023 г.

Принята к публикации 19.09.2023 г.

Работа направлена на определение индивидуальных особенностей речи взрослых в разных эмоциональных состояниях. Изучены акустические характеристики речи 12 взрослых носителей русского языка. Проведена аудиозапись речи информантов при произнесении текста-бесмыслицы в эмоциональных состояниях: радость, гнев, печаль, страх, нейтральное. Временные и спектральные характеристики речи анализировали в звуковом редакторе Cool Edit Pro. В речи мужчин максимальный диапазон частоты основного тона выявлен в высказываниях, произнесенных в нейтральном состоянии и состоянии радости, минимальный — в состоянии печали. Для женской речи — максимальный диапазон частоты основного тона в состоянии радости и в состоянии гнева, наименьший — в состоянии печали и в нейтральном состоянии. Диапазон частоты основного тона в женской речи больше, по сравнению с мужской. Для 7 информантов показано, что длительность высказываний в состоянии печали была больше по сравнению с другими состояниями, а в состоянии радости — минимальна. Как мужские, так и женские высказывания в состоянии радости характеризовались максимальными значениями диапазона частоты основного тона, в состоянии печали, наоборот минимальными, паузы между словами в высказываниях в состоянии печали выявлены как у мужчин, так и у женщин. Таким образом, выявлены различия во временных и спектральных характеристиках высказываний в разных эмоциональных состояниях. Определены индивидуальные особенности проявления эмоционального состояния в речи взрослых.

Ключевые слова: акустика речи, эмоциональное состояние, частота основного тона, диапазон частоты основного тона, длительность высказывания, длительность пауз

DOI: 10.31857/S0320791924010124 EDN: ZNRKAG

ВВЕДЕНИЕ

Одним из важных факторов, определяющих эффективность общения между людьми, является способность точно определять эмоции по речи [1]. Распознавание эмоций играет важную роль при взаимодействии человека и компьютера — во многих приложениях эмоции пользователя распознаются только по голосовой модальности на основании паралингвистической информации [2].

Разрабатываются интеллектуальные системы комплексного паралингвистического анализа речи взрослых [3]. В настоящее время идет работа по созданию базы данных эмоциональной речи, и строятся модели для распознавания эмоциональных состояний русскоязычных детей по речи [4, 5].

В классических работах, посвященных изучению проявления эмоций в речи, представлены разные модели эмоциональных состояний: имитация эмоций профессиональными актерами, певцами, воспроизведение эмоций в речи под гипнозом, эмоциональные состояния в условиях психиатрической клиники, естественное эмоциональное состояние человека в экстремальных ситуациях [6–12].

Показано, что изменение эмоционального состояния отражается в акустико-фонетических параметрах речевого сигнала. Характеристиками речевого сигнала, позволяющими определить эмоциональное состояние говорящего, являются громкость, изменение частоты основного тона (ЧОТ), величина “изрезанности” мелодического контура, изменение темпа речи [7].

При изучении певческой речи выявлено, что каждая из эмоций — радость, горе, гнев, страх — выражается изменением совокупности свойств звука: силы, высоты, тембра, темпо-ритмических характеристик [11]. Например, горе характеризуется наибольшей длительностью слога, медленным нарастанием и спадом силы звука. Гнев — резкими переходами, обрывками звука, большой силой голоса, зловещим звенящим или шипящим тембром. Для страха отличительными оказались резкие перепады силы голоса, сильное нарушение ритма мелодии, резкое увеличение пауз [11]. При этом для каждого исполнителя отмечают индивидуальную специфику комбинации голосовых средств для выражения эмоций.

Ключевыми акустическими характеристиками речи для определения эмоционального состояния говорящего является изменение во времени ЧОТ и интенсивности [9]. Так, при сильном эмоциональном возбуждении изменение ЧОТ может составлять до двух октав. Показано, что для некоторых эмоций характерны значительные изменения спектральных характеристик речевого сигнала: изменение диапазона ЧОТ резкое или более плавное, изменение формантных частот ударных гласных и их интенсивности [13].

Существует другой подход к оценке работы голосового источника, который предполагает создание математических моделей, позволяющих связать параметры голосового источника с фазово-частотными характеристиками сегментов речевого сигнала, что может быть использовано для идентификации диктора по голосу [14, 15].

В настоящее время создание моделей распознавания эмоций в речи человека является одной из наиболее развивающихся областей компьютерной паралингвистики [16, 17]. Исследователи в области автоматического распознавания речи осуществляют поиск набора признаков для гарантированного распознавания эмоций в речи [18]. Экспертные наборы признаков, основанные на знаниях об акустических свойствах речевых сигналов — просодических, спектральных, энергетических, вокализованных, — используются в качестве основы для решения разных задач компьютерной паралингвистики, в том числе распознавания эмоций [19].

Таким образом, современные работы посвящены автоматическому распознаванию эмоций в речи [18, 20–22], однако исследования характеристик взрослой эмоциональной речи с учетом индивидуальных особенностей демонстрации эмоционального состояния говорящим, которые могли бы служить в качестве контроля для сравнения с характеристиками эмоциональной детской речи, практически отсутствуют.

Данное исследование проводится в рамках работ, направленных на определение временных

и спектральных характеристик эмоциональной речи на материале русского языка [5, 23–24] и является частью проекта, посвященного разработке комплексного методического подхода к оценке эмоциональной сферы детей [25].

Цель работы — определить временные и спектральные характеристики речи взрослых в разных эмоциональных состояниях.

МЕТОДИКА

Участниками исследования стали 12 взрослых носителей русского языка (6 мужчин и 6 женщин, возраст 25.3 ± 4.5 и 24.5 ± 4.8 лет соответственно), не имевших нарушений слуха и речи. Дикторы получили инструкцию произносить фразу Л. В. Щербы “Глѡкая кѹздра штѣко будланѹла бѡкра и курдѣчит бокрѣнка”, демонстрируя разные эмоциональные состояния: радость, гнев, печаль, страх, нейтральное состояние. Проводили аудиозапись речи информантов с использованием магнитофона “Marantz PMD660” с выносным микрофоном “Sennheiser e835S”. Микрофон устанавливали на расстоянии 30–50 см от лица диктора. Аудиофайлы сохраняли в формате wav, 48 000 Гц, 16 бит. Параллельно с аудиозаписью регистрировали частоту сердечных сокращений, которая отражала степень возбуждения диктора.

Для оценки точности выполнения диктором поставленной перед ним задачи проводили перцептивный эксперимент. С этой целью созданы 2 аудиотеста: один включал фразы, произнесенные мужчинами, второй — женщинами. Каждый тест включал по 30 фраз, произнесенных информантами в разных эмоциональных состояниях. Фразы дикторов были замешаны в случайном порядке. Длительность каждого теста составила 6 мин. Тесты предъявляли двум группам аудиторов. Первая группа — специалисты с опытом работы в области речевых исследований — эксперты, 8 человек (4 мужчины, 4 женщины, средний возраст, 35.5 ± 11 лет). Вторая группа прослушивающих тесты без опыта работы в области речевых исследований — 50 студентов 1 курса, обучающихся по специальности “Педиатрия” — аудиторы (11 мужчин, 39 женщин, средний возраст 18.9 ± 2.4 лет). Для всех дикторов и экспертов проводили проверку слуха методом тональной аудиометрии в условиях лаборатории в звукоизолированной камере с использованием аудиометра АА-02. Пороги слуха в пределах возрастной нормы — не превышают 25 дБ. Эксперты получали задания определить эмоциональное состояние диктора и степень выраженности эмоции (описать в баллах от 1 до 3 — нет, слабо выражена, сильно выражена) при прослушивании каждой фразы, аудиторы — определить эмоциональное состояние диктора.

По результатам эксперимента строили матрицы спутывания [16], позволяющие оценить

вероятность правильного распознавания и ошибки, которые допускают аудиторы при определении эмоционального состояния информантов. Матрица спутывания представляет собой таблицу, строки которой соответствуют заданным (прогнозируемым) классам, столбцы – фактическим значениям (реальным классам). Вычисляли: полноту (recall) – долю найденных образцов, принадлежащих к классу, относительно всех образцов этого класса в тестовой выборке; точность (precision) – долю образцов, действительно принадлежащих данному классу, относительно всех образцов, которые отнесли к данному классу; UAR (Unweighted Average Recall) – усредненную полноту для всех эмоциональных состояний, указывающую на среднюю вероятность распознавания.

Анализ временных и спектральных характеристик речи проводили в звуковом редакторе Cool Edit Pro. Для каждого высказывания определяли среднее значение ЧОТ ($F_{\text{ср}}$), максимальное (F_{max})

и минимальное значение ЧОТ (F_{min}), диапазон частоты основного тона (разность между максимальным и минимальным значением ЧОТ, $F_{\text{max}} - F_{\text{min}}$), длительность всего высказывания, длительность пауз между словами. Статистическую обработку данных проводили в программе “Statistica 10.0” с использованием непараметрического критерия Манна–Уитни. Исследование одобрено Этическим комитетом СПбГУ, протокол № 115-02-2 от 06.04.2022.

РЕЗУЛЬТАТЫ

Анализ результатов перцептивного эксперимента показал, что эксперты и аудиторы без опыта работы в области речевых исследований достоверно распознают эмоциональные состояния, демонстрируемые в речи дикторами-мужчинами и дикторами-женщинами (значения полноты для всех эмоциональных состояний 0.58–0.75).

Таблица 1. Распознавание экспертами эмоциональных состояний, демонстрируемых дикторами-мужчинами (% ответов)

	Радость	Нейтральное	Печаль	Страх	Гнев
Радость	60	38	2	0	0
Нейтральное	2	54	24	7	13
Печаль	2	13	79	6	0
Страх	15	7	17	52	9
Гнев	4	15	4	10	67
Total	84	126	126	75	89
Recall	0.60	0.54	0.79	0.52	0.67
Precision	0.72	0.43	0.63	0.70	0.76
F1-score	0.66	0.48	0.70	0.60	0.71
UAR	0.63				

Таблица 2. Распознавание аудиторами эмоциональных состояний, демонстрируемых дикторами-мужчинами (% ответов)

	Радость	Нейтральное	Печаль	Страх	Гнев
Радость	60	32	6	1	1
Нейтральное	5	73	11	8	3
Печаль	3	29	61	6	0
Страх	11	19	20	39	11
Гнев	10	20	7	5	57
Total	90	173	105	59	73
Recall	0.60	0.73	0.61	0.39	0.57
Precision	0.67	0.42	0.58	0.66	0.79
F1-score	0.63	0.53	0.59	0.49	0.66
UAR	0.58				

Таблица 3. Распознавание экспертами эмоциональных состояний, демонстрируемых дикторами-мужчинами (% ответов)

	Радость	Нейтральное	Печаль	Страх	Гнев
Радость	66	23	0	2	9
Нейтральное	0	88	13	0	0
Печаль	0	18	80	1	1
Страх	0	9	7	65	19
Гнев	2	10	0	10	78
Total	68	148	100	78	107
Recall	0.66	0.88	0.80	0.65	0.78
Precision	0.97	0.59	0.80	0.83	0.73
F1-score	0.79	0.71	0.80	0.73	0.76
UAR	0.75				

Таблица 4. Распознавание аудиторами эмоциональных состояний, демонстрируемых дикторами-женщинами (% ответов)

	Радость	Нейтральное	Печаль	Страх	Гнев
Радость	69	26	2	1	3
Нейтральное	1	81	12	3	3
Печаль	0	24	66	3	3
Страх	1	11	10	56	12
Гнев	8	12	1	4	69
Total	79	153	91	66	90
Recall	0.69	0.81	0.66	0.56	0.69
Precision	0.87	0.53	0.73	0.84	0.77
F1-score	0.77	0.64	0.70	0.67	0.73
UAR	0.68				

Аудиторы обеих групп распознают состояния “радость”, “печаль” и “гнев” с большей вероятностью, чем состояния “страх” и “нейтральное” (табл. 1, 2). Эксперты распознают состояния “печаль” и “страх” лучше, чем аудиторы без опыта. Аудиторы без опыта распознают состояния “радость”, “нейтральное” и “печаль” лучше, чем состояния “страх” и “гнев”.

Эмоциональные состояния, демонстрируемые дикторами-женщинами, эксперты и аудиторы без опыта (табл. 3, 4) распознают с более высокой вероятностью, чем эмоциональные состояния, демонстрируемые мужчинами (UAR – 0.75 и 0.69 соответственно). Эксперты с высокой вероятностью распознают все состояния. Аудиторы без опыта лучше распознают состояния “радость”, “нейтральное”, “печаль” и “гнев”, с меньшей вероятностью состояние “страх”.

Результаты перцептивного эксперимента показали, что дикторы, несмотря на отсутствие

актерского опыта, выполнили поставленную перед ними задачу.

На основании инструментального спектрографического анализа выявлены различия по значениям разности диапазона ЧОТ, длительности высказываний дикторов при демонстрации разных эмоциональных состояний.

Максимальный диапазон значений ЧОТ в мужской речи выявлен в высказываниях, произнесенных в нейтральном состоянии и состоянии радости ($p < 0.05$), минимальный – в состоянии печали (табл. 5). В речи женщин максимальные значения диапазона ЧОТ выявлены в состоянии радости и в состоянии гнева ($p < 0.05$), наименьшие – в состоянии печали и в нейтральном состоянии (табл. 6). Диапазон значений ЧОТ в разных эмоциональных состояниях в женской речи больше, по сравнению с мужской ($p < 0.05$). Не выявлено значимых различий по средним значениям ЧОТ в высказываниях, произнесенных

Таблица 5. Временные и спектральные характеристики речи мужчин в разных эмоциональных состояниях

Состояние \ Характеристики	Нейтральное	Радость	Гнев	Печаль	Страх
F_{0cp} , Гц	150 ± 7	160 ± 13	156 ± 38	148 ± 5	150 ± 6
F_{0max} , Гц	178 ± 6	206 ± 13	187 ± 42	172 ± 6	169 ± 20
F_{0min} , Гц	103 ± 6	103 ± 6	109 ± 24	109 ± 11	103 ± 7
$F_{0max} - F_{0min}$, Гц	75 ± 13	103 ± 6	125 ± 48	62 ± 10	65 ± 13
Длительность высказывания, мс	4128 ± 395	3764 ± 11	3950 ± 625	$4391^* \pm 5$	4005 ± 62

Примечание: в таблице представлены средние значения и стандартные отклонения каждого параметра, * – $p < 0.05$ – критерий Манна–Уитни.

Таблица 6. Временные и спектральные характеристики речи женщин в разных эмоциональных состояниях

Состояние \ Характеристики	Нейтральное	Радость	Гнев	Печаль	Страх
f_{0cp} , Гц	219 ± 24	250 ± 64	250 ± 64	226 ± 19	256 ± 58
F_{0max} , Гц	265 ± 24	$367^* \pm 80$	$312^* \pm 57$	289 ± 55	305 ± 65
F_{0min} , Гц	188 ± 1	188 ± 19	188 ± 29	188 ± 1	196 ± 19
$F_{0max} - F_{0min}$, Гц	78 ± 23	$186^* \pm 72$	$125^* \pm 48$	101 ± 55	109 ± 48
Длительность высказывания, мс	4429 ± 782	4383 ± 444	4321 ± 273	4795 ± 998	4434 ± 633

Примечание: в таблице представлены средние значения и стандартные отклонения каждого параметра, * – $p < 0.05$ – критерий Манна–Уитни.

Таблица 7. Индивидуальные различия в группах дикторов

Дикторы	UAR экспертов и аудиторов	Степень выраженности эмоционального состояния ¹	Изменение ЧОТ	Временные характеристики	Стратегия
Мужчины					
1	0.57	1.85 ± 0.24	–	+	2
2	0.79	2.48 ± 0.42	+	+	3
3	0.68	1.93 ± 0.31		+	2
4	0.24	1.55 ± 0.34	–	–	–
5	0.66	1.88 ± 0.48	–	+	2
6	0.54	2.1 ± 0.44	+	+	3
Женщины					
7	0.68	2.1 ± 0.35	+	+	3
8	0.69	2.3 ± 0.64	+	–	1
9	0.74	1.7 ± 0.27	–	+	2
10	0.58	1.85 ± 0.24	+	–	1
11	0.92	2.68 ± 0.30	+	+	3
12	0.85	2.4 ± 0.36	+	–	1

¹ – Согласно балльной оценке аудиторов, представлены средние значения и стандартные отклонения.

мужчинами и женщинами в разных эмоциональных состояниях.

По длительности высказываний в разных эмоциональных состояниях значимых различий не выявлено.

В речи мужчин паузы отсутствовали в состоянии радости и нейтральном, паузы присутствовали в высказываниях в состоянии страха, печали и гнева, максимальные значения длительности пауз между словами в состоянии страха. У женщин паузы в высказываниях присутствовали в состоянии печали, радости и нейтральном, максимальные значения длительности пауз зарегистрированы в состоянии печали.

В табл. 7 на основе данных перцептивного эксперимента и инструментального спектрографического анализа представлены индивидуальные различия в группах дикторов. Показано, что при высокой точности распознавания эмоциональных состояний, демонстрируемых дикторами, за исключением диктора № 4, наблюдаются различия в стратегиях проявления эмоционального состояния. Стратегия 1 — изменение значений диапазона ЧОТ в разных эмоциональных состояниях, стратегия 2 — изменение временных характеристик высказывания, стратегия 3 — изменение значений диапазона ЧОТ и временных характеристик высказывания. Для дикторов-мужчин № 1, 3, 5 характерна стратегия 2, для дикторов-мужчин № 2 и № 6 — стратегия 3. У трех дикторов-женщин (№ 8, 10, 12) выявлена стратегия 1, у двух — стратегия 3 и только у 1 диктора-женщины — стратегия 2 (№ 9).

Представлены значения разности между максимальным и минимальным ЧОТ диктора-женщины

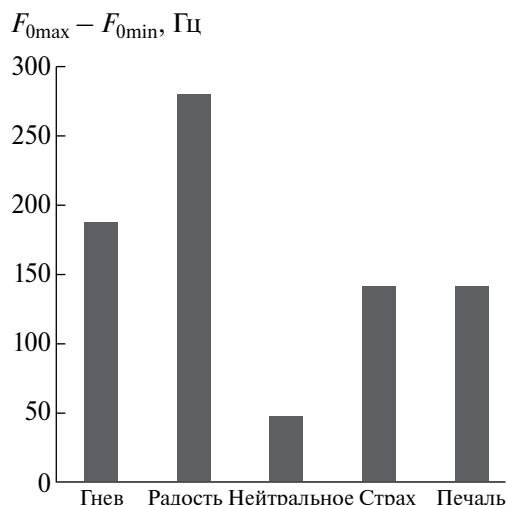


Рис. 1. Значения диапазона ЧОТ в речи диктора-женщины № 12 в разных эмоциональных состояниях.

№ 12, в речи которой эмоции были выражены наиболее ярко (рис. 1). Минимальное значение диапазона ЧОТ характерно для нейтрального состояния — 48 Гц, максимальное для состояния радости — 279 Гц, в состоянии агрессии и страха не выявлено различий по значениям диапазона ЧОТ — 188 Гц, печаль 140 Гц.

У другого участника исследования (диктора № 1) различий по значениям диапазона ЧОТ не выявлено. При этом высказывания в разных эмоциональных состояниях различались по временным характеристикам: по длительности высказывания, по

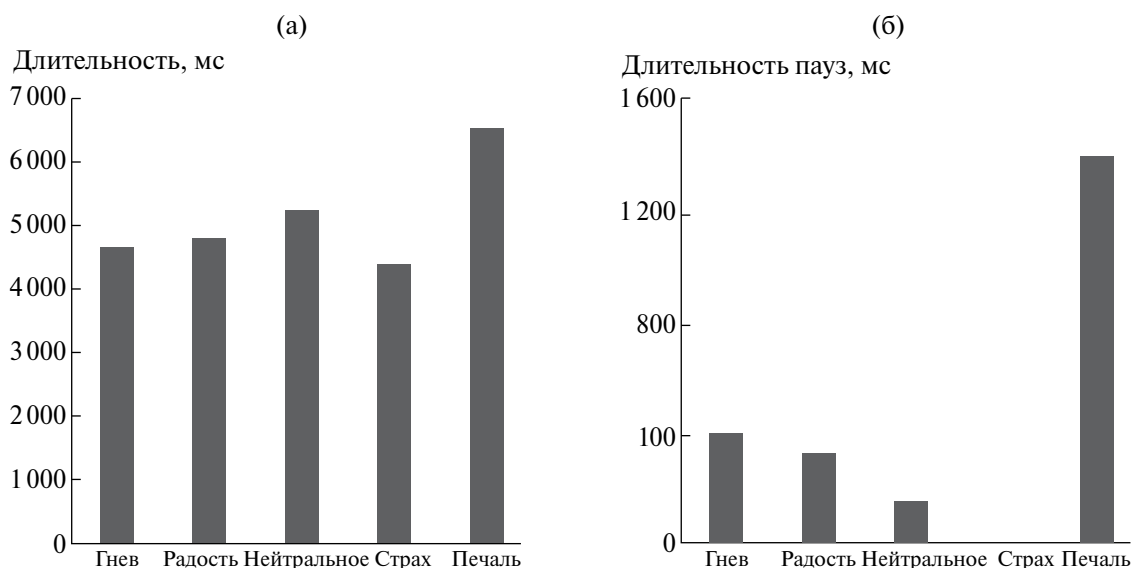


Рис. 2. (а) — Значения длительности высказывания и (б) — пауз внутри высказывания в речи диктора № 1 в разных эмоциональных состояниях.

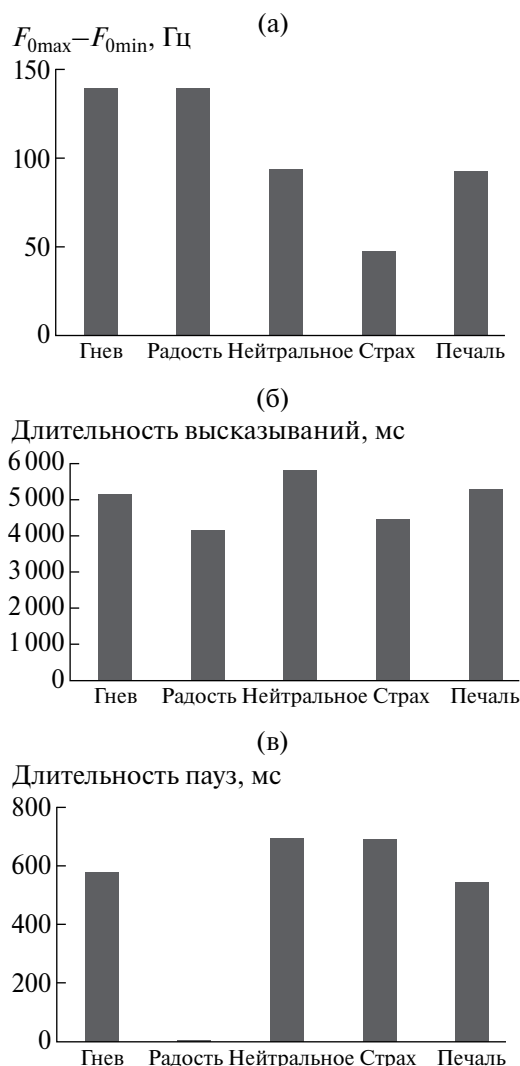


Рис. 3. (а) — (а) — Значения диапазона ЧОТ, (б) — длительности высказывания и (в) — пауз внутри высказывания диктора-женщины № 11 в разных эмоциональных состояниях.

длительности пауз внутри высказывания (рис. 2). Наибольшая длительность высказывания была характерна для состояния печали — 6569 мс. Выявлены большие паузы между словами. Наиболее коротким было высказывание в состоянии страха — 4785 мс, при этом пауз между словами не выявлено. Высказывание в состоянии гнева было незначительно короче, чем в состоянии радости, но длительность пауз в состоянии гнева была больше.

Еще одна стратегия проявления эмоционального состояния в речи — высказывания, произнесенные диктором № 11 в разных эмоциональных состояниях, различаются по совокупности спектральных и временных характеристик (рис. 3).

Максимальные значения диапазона ЧОТ, зарегистрированные в высказываниях диктора № 11

в состоянии гнева и радости (140 Гц), превышают значения диапазона ЧОТ в нейтральном состоянии и в состоянии печали. Минимальное значение диапазона ЧОТ зарегистрировано в состоянии страха. В состоянии гнева общая длительность высказывания больше, чем в состоянии радости. При этом диктор в состоянии гнева делает паузы между словами (574 мс), в состоянии радости паузы между словами отсутствуют. В нейтральном состоянии и в состоянии печали диапазон ЧОТ одинаков. Длительность высказывания в нейтральном состоянии больше за счет более длительных пауз, по сравнению с состоянием печали, в котором в высказывании длительность гласных в конце слова больше, чем в высказывании в нейтральном состоянии.

ОБСУЖДЕНИЕ

В работе представлены данные о временных и спектральных характеристиках высказываний взрослых, произнесенных в разных эмоциональных состояниях. В качестве дикторов выбраны взрослые — носители русского языка, не имеющие профессионального актерского опыта, т.е. участники исследования воспроизводили эмоции согласно своему личному опыту. Такой подход позволил получить записи с учетом индивидуальных особенностей речи информантов. Соответствие продемонстрированных дикторами эмоциональных состояний заданным условиям эксперимента подтверждено результатами перцептивного анализа. С одной стороны, показано, что каждая эмоция характеризуется совокупностью отличительных акустических признаков голоса [11], с другой стороны, установлена большая вариативность и индивидуальные особенности в проявлении эмоций разными дикторами [12]. В настоящей работе инструментальный анализ высказываний позволил выявить следующие общие закономерности в изменении акустических характеристик речи говорящего в разных эмоциональных состояниях: максимальный диапазон ЧОТ в состоянии радости, минимальный — в состоянии печали.

Выявлены индивидуальные особенности эмоциональной речи участников исследования и разные стратегии проявления эмоциональных состояний в голосе и речи информантов. Первая стратегия основана на преимущественном изменении ЧОТ в высказывании, вторая на изменении длительности высказывания и пауз внутри высказывания, третья предполагает одновременное изменение ЧОТ и временных характеристик высказывания.

Показано, что радость может проявляться значительным повышением значений диапазона ЧОТ по сравнению с ЧОТ высказываний, произнесенных в других эмоциональных состояниях, в том

числе состоянии гнева. В случае, если значения диапазона ЧОТ в высказываниях информанта практически не изменялись, то длительность высказываний в состоянии радости и гнева была больше, чем в нейтральном состоянии, но при этом в состоянии радости паузы внутри высказывания были короче, чем в состоянии гнева.

ЗАКЛЮЧЕНИЕ

В ходе исследования с использованием инструментального анализа речи взрослых носителей русского языка показано, что высказывания, произносимые мужчинами и женщинами в разных эмоциональных состояниях, могут различаться как по спектральным, так и временным характеристикам. В качестве общих закономерностей можно отметить максимальные значения диапазона ЧОТ в состоянии радости по сравнению с другими состояниями и максимальные значения длительности высказываний в состоянии печали. Выявлены индивидуальные особенности, характерные для разных информантов, которые заключаются в разных стратегиях проявления эмоционального состояния в речи.

Работа выполнена при финансовой поддержке Российского Научного Фонда (проект № 22-45-02007).

СПИСОК ЛИТЕРАТУРЫ

1. *Schwartz R., Pell M.D.* Emotional speech processing at the intersection of prosody and semantics // *PLoS One*. 2012. V. 7(10). P. e47279. <https://doi.org/10.1371/journal.pone.0047279>
2. *Papakostas M., Siantikos G., Giannakopoulos T., Spyrou E., Sgouropoulos D.* Recognizing Emotional States Using Speech Information // *Adv. Exp. Med. Biol.* 2017. V. 989. P. 155–164. https://doi.org/10.1007/978-3-319-57348-9_13
3. *Величко А.Н., Верхоляк О.В., Карнов А.А.* Программная система для распознавания эмоций в речи (ProSpER — Program for Speech Emotion Recognition). 2020.
4. *Matveev Y., Matveev A., Frolova O., Lyakso E., Ruban N.* Automatic speech emotion recognition of younger school age children // *Mathematics*. 2022. V. 10. P. 2373. <https://doi.org/10.3390/math10142373>
5. *Lyakso E., Frolova O., Dmitrieva E., Grigorev A.* EmoChildRu: Emotional Child Russian Speech Corpus // *Ronzhin A., Potapova R., Fakotakis N. (Eds.) Speech and Computer. SPECOM 2015. Lecture Notes in Computer Science*. 2015. V. 9319.
6. *Галунов В.И.* Некоторые проблемы акустической теории речеобразования // *Акуст. журн.* 2002. Т. 48. № 6. С. 845–848.
7. *Галунов В.И.* О возможности определения эмоционального состояния говорящего по речи // *Речевые технологии*. 2008. № 1. С. 60–66.
8. *Галунов В.И., Манеров В.Х., Тарасов В.И.* Слуховой анализ речи, полученной при использовании различных методов моделирования эмоциональных состояний // *Материалы симпозиума “Речь и эмоции”*. Ленинград, 1974. С. 79–83.
9. *Галунов В.И., Манеров В.Х.* Пути решения проблемы создания систем определения эмоционального состояния говорящего // *Вопр. кибернетики*. Вып. 22. М., 1976. С. 95–114.
10. *Морозов В.П.* Акустический язык эмоций в жизни и науке // *Язык эмоций, мозг и компьютер. Вычислительная техника и ее применение*. М.: Знание, 1989. Вып. 9. С. 3–18.
11. *Морозов В.П.* Язык эмоций и эмоциональный слух. Избранные труды 1964–2016. М.: ФГБУН Институт психологии РАН, 2017. 397 с.
12. *Потапова Р.К., Потапов В.В.* Язык, речь, личность. М: Языки славянской культуры, 2006. 491 с.
13. *Михайлов В.Г., Златоустова Л.В.* Измерение параметров речи. М.: Радио и связь, 1987. 168 с.
14. *Сорокин В.Н., Леонов А.С.* Фазовый анализ активности голосового источника // *Акуст. журн.* 2021. Т. 67. № 2. С. 185–202.
15. *Сорокин В.Н., Леонов А.С.* Фазовые модуляции в речевом сигнале // *Акуст. журн.* 2022. Т. 68. № 2. С. 218–232.
16. *Ляко Е.Е., Фролова О.В., Гречаный С.В., Матвеев Ю.Н., Верхоляк О.В., Карнов А.А.* Голосовой портрет ребенка с типичным и атипичным развитием / Под ред. Ляко Е.Е., Фроловой О.В. СПб.: Издательско-полиграфическая ассоциация высших учебных заведений, 2020. 204 с.
17. *Schuller D.M., Schuller B.W.* A Review on Five Recent and Near-Future Developments in Computational Processing of Emotion in the Human Voice // *Emotion Review*. 2021. V. 13. № 1. P. 44–50. <https://doi.org/10.1177/1754073919898526>
18. *Singh J., Saheer L.B., Faust O.* Speech Emotion Recognition Using Attention Model // *Int. J. Environ. Res. Public Health*. 2023. V. 20. P. 5140. <https://doi.org/10.3390/ijerph20065140>
19. *Двойникова А.А., Маркитантов М.В., Рюмина Е.В., Уздяев М.Ю., Величко А.Н., Рюмин Д.А., Ляко Е.Е., Карнов А.А.* Анализ информационного и математического обеспечения для распознавания аффективных состояний человека // *Информатика и автоматизация*. 2022. Т. 21. № 6. <https://doi.org/10.15622/ia.21.6.2>
20. *Ververidis D., Kotropoulos K.* Emotional speech recognition: Resources, features, and methods // *Speech Communication*. 2006. V. 48. № 9. P. 1162–1181.
21. *Xu M., Zhang F., Zhang W.* Head fusion: Improving the accuracy and robustness of speech emotion

- recognition on the IEMOCAP and RAVDESS dataset // IEEE Access. 2021. V. 9. P. 74539–74549.
22. *Потанова Р.К.* Лингвистические знания и новые технологии // Акуст. журн. 2002. Т. 48. № 4. С. 552–559.
23. *Lyakso E., Frolova O.* Emotion State Manifestation in Voice Features: Chimpanzees, Human Infants, Children, Adults // Ronzhin A., Potapova R., Fakotakis N. (Eds.) Speech and Computer. SPECOM 2015. Lecture Notes in Computer Science. V. 9319. Springer, Cham.
24. *Grigorev A.S., Gorodnyi V.A., Frolova O.V., Kondratenko A.M., Dolgaya V.D., Lyakso E.E.* Acoustic and Perceptual Features of the Emotional Speech of Adolescents Aged 14–16 Years // Neurosci Behav Physi. 2020. V. 50. P. 1224–1231.
25. *Lyakso E., Frolova O., Kleshnev E., Ruban N., Mekala M.M., Arulalan K.V.* Approbation of the Child's Emotional Development Method (CEDM) // Companion Publication of the 2022 International Conference on Multimodal Interaction (ICMI '22 Companion). New York, NY, USA. 2022. P. 201.