Научная статья
УДК 347.78.01
DOI: 10.17323/tis.2025.27359

Original article

ПРОСОДИЯ КАК ЭЛЕМЕНТ ПРАВОВОЙ ОХРАНЫ ГОЛОСА ГРАЖДАНИНА

PROSODY IN THE LEGAL PROTECTION OF THE HUMAN VOICE

Артем Русланович БУДНИК
Московский государственный технический университет имени Н.Э. Баумана, Москва, Россия,
budnikart@gmail.com,
ORCID: 009-0005-6623-9668

Информация об авторе
А.Р. Будник — студент 3-го курса Московского государственного технического университета имени Н.Э. Баумана

Аннотация. Законопроект об охране голоса человека как объекта личных неимущественных прав в условиях его копирования с помощью искусственных нейронных сетей определил цель и задачи настоящего исследования. Цель работы — системный анализ феномена просодии как юридически значимого параметра. Перечень решенных в исследовании задач охватывает выявление особенностей, параметрический анализ, проверку гипотезы о возможности оцифровки и автоматизированного воспроизведения одной из основных модальностей речи человека — просодии требования. Методика работы построена на нахождении источника качественных образцов данной речевой модальности, параметризации и цифровизации выбранной просодии, а также ее сравнении с нейтральными примерами вербальной коммуникации.

В статье сделан вывод о необходимости правовой детализации категории «голос» в целях выработки эффективного механизмы его охраны и защиты. Полезным итогом работы является постановка эксперимента по измерению и фиксации параметров, которые образуют совокупность повторяющихся признаков, характерных для речевой модальности требования, выполненного с использованием неадаптированных образцов высказываний индивидуумов. Результаты эксперимента получили математическую оценку и лингвистическую интерпретацию. На основании выявленных закономерностей и экспериментальных данных разработан программный модуль генерации речевой просодии требования. Созданный прототип прошел экспериментальную верификацию качества его целевой функции.

Научная новизна работы заключается в конкретизации знаний о феномене и параметрах просодии, необходимых для законодательного регулирования голосовой идентичности граждан в качестве нематериального блага, а также в доказательстве возможности автоматического воспроизведения вербальной экспрессии человека программно-аппаратными средствами. Практическая значимость исследования обусловлена применимостью его результатов для правовой охраны голоса человека, глубокого обучения больших лингвистических моделей навыкам речевой коммуникации и придания вербальным элементам характеристик определенной просодии, в том числе с помощью интеграции разработанного прототипа.

Ключевые слова: правовая охрана голоса, голос как личное благо, просодия, параметры просодии, оцифровка речи, лингвистическая модель, генеративная языковая модель, искусственная нейронная сеть, разработка программного продукта

Для цитирования: *Будник А.Р.* Просодия как элемент правовой охраны голоса гражданина // Труды по интеллектуальной собственности (Works on Intellectual Property). 2025. Т. 53, № 2. С. 49–66; DOI: 10.17323/tis.2025.27359

Artjom R. BUDNIK
Bauman Moscow State Technical University, Moscow, Russia,
budnikart@gmail.com,
ORCID: 0009-0005-6623-9668

Information about the author
A.R. Budnik — 3rd year student of Bauman Moscow State Technical University

Abstract. The draft law on the protection of the human voice as an object of personal non-property rights when

it is generated by artificial neural networks defined the objectives of this study to identify the features, parametric analysis and digitalization of one of the main modalities of human speech-demand prosody. The methodology of the work is based on a systemic analysis of the phenomenon of prosody as a legally significant parameter, putting forward and testing a hypothesis about which type of speech prosody is the most pronounced. The selected methodology included a search for a source of high-quality samples of this speech modality, parameterization and digitalization of the selected prosody, as well as its comparison with neutral examples of verbal communication.

The results of the work include a conclusion on the need for legal detailing of the category of voice in order to develop effective mechanisms for its protection and defense. A useful result of the work was the setting up of an experiment on measuring and recording parameters that form a set of recurring features characteristic of the speech modality of demand, using unadapted, natural examples of individual statements. The results of the experiment were mathematically assessed and linguistically interpreted. Based on the identified patterns and experimental data, a software module for generating speech prosody of the requirement was developed. The created prototype passed experimental verification of the quality of its target function.

The scientific novelty of the work lies in the specification of knowledge about the parameters of prosody, important for its legislative regulation as an intangible benefit, as well as in proving the possibility of automatic reproduction of human verbal expression by software and hardware. The practical significance of the study is due to the applicability of its results for the legal protection of the human voice, deep training of large linguistic models in speech communication skills and imparting characteristics of a certain prosody to verbal elements, including through the integration of the developed prototype.

Keywords: legal protection of voice, personal right, prosody, prosody parameters, speech modality digitization, linguistic model, generative language model, artificial neural network, software product development

For citation: *Budnik A.R.* Prosody in the legal protection of the human voice // Trudi po Intellectualnoy Sobstvennosti (Works on Intellectual Property). 2025. Vol. 53 (2). P. 49–66; DOI: 10.17323/tis.2025.27359

ВВЕДЕНИЕ

В настоящее время в сфере искусственных нейронных сетей наблюдается стремительное развитие больших языковых моделей, в частности авторегрессионных генеративных инструментов на основе технологии GPT (Generative Pre-trained Transformer) [1]. Большие GPT-подобные лингвистические модели также используются для анализа и обработки естественного языка людей (Natural Language Processing, NLP) [2]. Прогресс NLP-технологий позволил разработчикам двинуться в направлении обучения искусственных нейронных сетей характерной для человека речевой коммуникации [3].

Эти методы получили наименование Generative Audio (далее — GA) либо Generative AI for Audio (далее — GAIA). Второй вариант подчеркивает применение технологии искусственного интеллекта в данном процессе. GAIA представляют собой инструменты человекоподобного озвучивания сгенерированного нейронной сетью текста, а также один из перспективных векторов развития языковых моделей и технологий обработки естественного языка. GAIA-технология нацелена на повышение качества взаимодействия пользователя с результатами нейросетевой генерации контента, обогащая их более реалистичными и захватывающими звуковыми эффектами. Развиваясь параллельно с языковыми моделями, GAIA предлагает инновационные методы озвучивания контента с использованием пространственного звука, контекстно определенных и персонализированных аудиоэффектов. Это не только повышает качество восприятия сгенерированного материала, обогащая мультимедийный опыт пользователей, но и придает синергию прогрессу языковых компьютерных моделей.

Задача повышения качества озвучивания текста остается одной из наиболее важных при разработке систем речевого синтеза, особенно в контексте интенсивного развития инструментов узкого искусственного интеллекта[1], и осложняется рядом следу-

[1] Узкий искусственный интеллект (его также называют слабым ИИ или просто ИИ) — это состояние компьютерного программного обеспечения на сегодняшний день. Компьютерные программы в основном разрабатываются программистами-людь-

ющих факторов [4]. Лингвистические особенности, такие как произношение звуков, интонация и ритм речи, могут существенно варьироваться в различных языках и диалектах. Синтезатор речи должен учитывать эти моменты для достижения естественности и воспринимаемости результатов его работы [5]. Фонетические особенности, включая различия в произношении звуков, требуют тщательного моделирования. Тембральные характеристики, связанные с уникальным звучанием каждого голоса, также представляют собой вызов для машинного синтеза речи [6]. Локально-обусловленные особенности, такие как акцент или диалект, могут влиять на фонетические и акустические элементы речи. Качественная модель синтеза речи должна быть способна адаптироваться к разнообразию акустической информации. Речевой синтезатор необходимо строить с учетом просодических элементов, таких как ритм, интонация и тон голоса[2]. Совокупность атрибутов речевой просодии обеспечит передачу эмоциональной окраски и намерений говорящего, что важно для генерирования естественной и понятной речи [7].

Оборотная сторона прогресса лингвистических моделей, голосовых генераторов и вокальных синтезаторов проявилась в увеличении количества случаев неавторизованного использования голосов граждан в коммерческих [8] и противоправных [9] целях. Быстрее других на этот вызов отреагировало государство, в котором расположен самый крупный мировой кластер производства развлекательного контента; лоббисты оценили конкурентный потенциал и масштаб экономических последствий применения технологий имитации голоса, репликации внешности и других узнаваемых черт умерших артистов. В результате был принят специальный закон[3], призванный ограничить использование этой этически сомнительной практики, которая к тому же способна нанести непоправимый ущерб индустрии. В России также выросло количество случаев так называемой кражи голоса, что отмечено в прессе[4] и целевых исследованиях специалистов [10]. По этой причине в нашей стране разработан проект правовой нормы об охране голосов граждан, которую планируется включить в Гражданский кодекс Российской Федерации в качестве отдельной статьи[5].

Цели настоящего исследования разбиты на три этапа. Первый этап заключается в юридическом анализе законопроекта об охране голоса гражданина и выявлении того недостатка, который может препятствовать его эффективному применению на практике. Здесь же будут приведены обоснование необходимости параметрического описания категории «голос» и аргументы в пользу включения речевой просодии в перечень этих характеристик. Второй этап состоит в анализе и классификации особенностей модальностей речи человека. Цели второго этапа — углубленное понимание речевой просодии и разработка методики для обработки естественного языка компьютерными средствами. Достижение целей второго этапа формирует основу для реализации программно-инженерного решения по автоматическому модулированию речевого образца просодией определенного типа на следующей стадии работы. На третьем этапе исследования будет разработан программный модуль, выполняющий функцию автоматического преобразования речи, сгенерированной в нейтральной модальности, для ее озвучивания в заданной просодии. Разработка программного модуля позволит сделать шаг в направлении создания более совершенных и адаптивных систем генерации речи на основе искусственного интеллекта, а также подчеркнет необходимость более четкого и параметрически конкретного определения такого объекта правовой охраны, как голос гражданина.

МЕТОДИКА И МАТЕРИАЛЫ ИССЛЕДОВАНИЯ

В качестве основы для правового анализа используется проект Федерального закона № 718834-8 «О вне-

ми, причем процесс от ввода до вывода открыт для проверки. Программы обычно специализируются в одной области, и может быть математически доказано, что они безопасны или дружественны. Даже в модельно-ориентированных архитектурах, где производство исходного кода частично автоматизировано, результирующий код разрабатывается человеком, и результаты его выполнения в значительной степени предсказуемы.

[2] Просодия — раздел фонетики, в котором рассматриваются такие особенности произношения, как высота, сила/интенсивность, длительность, придыхание, глоттализация, палатализация, тип примыкания согласного к гласному и другие признаки, являющиеся дополнительными к основной артикуляции звука. Элементы просодии реализуются в речевом потоке на всех уровнях речевых сегментов (в слоге, слове, синтагме, фразе и т.д.), выполняя при этом смыслоразличительную роль. В рамках просодии изучается как субъективный уровень восприятия характеристик суперсегментных единиц (высота тона, сила/громкость, длительность), так и их физический аспект (частота, интенсивность, время).

[3] Поправка в гражданское законодательства Штата Калифорния. An act to amend Section 3344.1 of the Civil Code, relating to intellectual property. Assembly Bill No. 1836. Use of likeness: digital replica.

[4] См. *Ященко Г.* Что делать, если голос украли и используют без разрешения? VC.ru., 19.12.2023. — URL: https://vc.ru/legal/959152-chto-delat-esli-golos-ukrali-i-ispolzuyut-bez-razresheniya

[5] Система обеспечения законодательной деятельности Российской Федерации. Законопроект № 718834-8. — URL: https://sozd.duma.gov.ru/bill/718834-8

сении изменений в часть первую Гражданского кодекса Российской Федерации», предполагающий его дополнение ст. 152.3 «Охрана голоса гражданина». Предлагаемые положения оцениваются с точки зрения их достаточности для решения вопросов, которые проявились в судебных разбирательствах об охране голоса и иных звуковых персонифицирующих характеристик человека. Далее выполняется анализ научной литературы в целях исследования тех параметров, которые используются, чтобы описывать голосовую, звуковую и речевую идентичности индивидуума для их учета при конструировании соответствующих регуляторных норм.

Для упрощения задачи выявления характерных признаков, их оцифровки, генерации и разработки механизма автоматического воспроизведения речевой модальности выбирается ее наиболее ярко выраженный, показательный пример. Установление закономерностей, обеспечение чистоты эксперимента, повторяемости и верификации выводов требует наличия не менее десяти типовых образцов. Известно, что наиболее ярко выраженной речевой модальностью является просодия требования [11]. В частности, Ю. Лотман отмечал: «Речевой акт требования может вызывать различные эмоциональные реакции у адресата, в зависимости от тонового окраса и контекста. Например, требование, выраженное сильным эмоциональным зарядом, может вызвать негативную реакцию или конфликт» [12]. Эту мысль в эссе «Речевые акты» развил Дж. Сёрл: «Речевой акт требования может быть эффективным только в том случае, если адресат принимает его как легитимный и обоснованный. Для этого необходимо учитывать социальные нормы и ожидания, а также отношения между говорящим и адресатом» [13]. Таким образом, для достижения целей исследования необходимо найти и проанализировать не менее десяти речевых образцов предъявления требования.

Материал исследования удалось найти на канале видеохостинга [14], где в режиме реального времени показаны моменты работы офицеров полиции США в различных ситуациях. Пользовательский контент регулярно применяется для осуществления социокультурных научных исследований вследствие его высокой информативности и широты эмпирической базы. Так, лингвистические работы Н. Мэдзлен с соавторами [15, 16], посвященные изучению просодических характеристик выражения различных видов отношений в видеоблогах, основаны на сообщениях частных видеоблогеров, составивших собой рафинированную базу релевантных материалов для решения поставленных задач. Публикации, которые были выбраны автором этой статьи из указанного источника, имеют достаточно высокое отношение сигнал/шум — от 60дБ и выше, что сделало их пригодными для оцифровки, параметризации и анализа.

В ходе исследования был проведен анализ 11 высказываний полицейских, которые включали требования и нейтральные обращения. Данные речевые образцы были измерены с помощью программного комплекса PRAAT[6] по следующим параметрам: частота звука в герцах (Гц), громкость образцов в децибелах (дБ), продолжительность высказываний в миллисекундах (мс).

Закономерности, выявленные в результатах измерений, положены в основу разработки программного решения для модуляции речевого образца с целью его воспроизведения в просодии требования.

РЕЗУЛЬТАТЫ

Просодия как юридически значимый параметр правовой охраны голоса

Проблема охраны голоса как нематериального и неотчуждаемого блага личности в России вышла на уровень законотворческих инициатив, которые в большинстве случаев предполагают введение юридической охраны голоса по аналогии с охраной изображения гражданина, зафиксированной в ст. 152.1 ГК РФ. Следует отметить, что для решения данной проблемы могут быть использованы и более общие правовые нормы ст. 150 ГК РФ. Из состава нематериальных благ, определенных в ч. 1 ст. 150 ГК РФ, голос или вокальная идентичность личности вполне относится к «нематериальным благам, принадлежащим гражданину от рождения или в силу закона, неотчуждаемым и не передаваемым иным способом». В таком случае ч. 2 данной статьи требует защиты этого блага «в соответствии с действующими нормами Гражданского кодекса и другими законами в случаях и порядке, ими предусмотренных…». Работа в направлении охраны голоса или вокальной идентичности личности как неотчуждаемого нематериального блага представляет собой своевременную реакцию законодателя на возникшую проблему. Выбранная юридическая техника нормотворчества, заключающаяся в регулировании подобных отношений подобными средствами, выглядит вполне логичным подходом.

Развитие и доступность технологий имитации голоса послужили причиной роста количества дел

[6] PRAAT — это свободное программное обеспечение для фонетического анализа речи, разработанное и поддерживаемое Полом Бурсма и Дэвидом Венинком из университета Амстердама.

о противоправном использовании голоса гражданина в судах Российской Федерации и других стран. В рамках судебных разбирательств, предъявления и оценки доказательств участники таких процессов сталкиваются с необходимостью определения того, что представляет собой категория «голос человека» в конкретных ее параметрах, признаках и характеристиках. Такая определенность необходима для проведения экспертиз и сравнений голосовых и вокально-артикуляционных образцов в целях установления фактов их копирования или заимствования.

Лапидарность формулировок проекта Федерального закона № 718834-8 «О внесении изменений в часть первую Гражданского кодекса Российской Федерации», связанная, по всей видимости, с намерением сохранить лаконичность правовой конструкции, взятой за образец нормы об охране изображения гражданина, породит проблему, которая неизбежно проявится в судебных спорах об охране прав граждан на их голос. Эта проблема заключается в отсутствии в тексте законопроекта определения и квалифицирующих признаков понятия «голос». Законодатель, предположительно, исходит из того, что граждане и правоприменители имеют некое общее, бытовое представление о феномене голоса человека, которого будет вполне достаточно для решения юридических споров.

Анализ предметных научных исследований феномена голоса человека обнаруживает высокую сложность, полимодальность и междисциплинарность данного явления с точки зрения как множественности параметрических характеристик, применяемых для его описания, так и способов их измерения [17]. Значение знаний о голосе для юриспруденции невозможно переоценить, поскольку голосовые параметры человека высоко информативны для анализа и правовой квалификации поступков, действий и бездействия человека [18–21]. Так, вина как психическое отношение лица к совершаемому общественно опасному действию или бездействию и его последствиям, выражающееся в форме умысла или неосторожности, может быть выявлена, подтверждена или опровергнута с помощью параметрического анализа голоса подозреваемого [22].

Технологии синтеза голоса deepfake и имитации певческого голоса vocaloid породили категорию юридических дел о копировании голосов известных личностей без их разрешения в развлекательных и коммерческих целях. Носители узнаваемых голосов протестуют против практики использования их личного нематериального блага, которая получила наименование «вокальная имперсонация». Суть вокальной имперсонации заключается в том, что профиль голоса целевого артиста оцифровывается нейросетью по ряду заданных параметров, включая индивидуальные характеристики тембра, физиологические особенности артикуляции гласных и согласных звуков, специфику дыхания и голосового звукоизвлечения, а также прочие атрибуты вокала, связанные с отличительными чертами произношения, ритмического и динамического слогового интонирования музыкальных и речевых фраз в зависимости от эмоциональной окраски музыкально-смыслового сообщения, формирующие в комплексе узнаваемую голосовую индивидуальность артиста.

В качестве примера практики вокальной имперсонации можно привести резонансный спор с участием продюсерского агентства Jay-Z Roc Nation, которое в 2020 г. потребовало удаления дипфейковых видеороликов, где артист рэп-жанра Jay-Z исполняет песни Билли Джоэла «We Didn't Start the Fire» и монолог «Быть или не быть» из пьесы Шекспира «Гамлет». Этот же метод был использован при создании нового трека ушедшего из жизни американского артиста Тупака Шакура и также вызвал неоднозначную реакцию в среде профильных юристов [23].

В классической науке голосоведения выделяют певческие и разговорные голоса, характеризуемые тремя основными параметрами: высотой, тембром и регистром голоса [24, 25]. Однако в современных тематических исследованиях предлагаются расширенные перечни атрибутов индивидуализации голосов [26]. Это, вероятно, связано с тем, что современная музыкальная культура породила множество стилей и направлений, в которых вокальная составляющая имеет немного общего с традиционным академическим и эстрадным пением. Среди новых подходов к голосовому звукоизвлечению стоит упомянуть бэлтинг, тванг и фрай в поп-музыке; скриминг, гроулинг, грайндкор и гуттурал в тяжелой металлической музыке; речитатив, скэт и разнообразные артикуляционные трюки в рэп- и хип-хоп-жанрах.

Дополнительные параметры голосовой индивидуализации включают в себя фонетические акустико-артикуляционные признаки звукообразования, которые могут определять особенности разговорных голосов и составлять художественный метод артистов вокального жанра. Артикуляция представляет собой сложный механизм, функционирование которого обусловлено скоординированной работой дыхательной системы, рта и носоглотки, голосовых связок, языка и губ. Результаты совместной работы этих систем носят строго индивидуальный характер и в значительной мере определяют голос человека. В число артикуляционных параметров голосовой индивидуализации включают специфику произнесения шумных согласных — обструэнтов; особенности звукообразования сонорных согласных, аппроксимантов, носовых, од-

ноударных и дрожащих фонем; придание нетипичного звучания гласным звукам в ударных и безударных слогах. Именно эти дополнительные характеристики используются для индивидуализации голосов упомянутых выше артистов в спорах о правомерности вокальной имперсонации.

Такой атрибут голоса и речи человека, как просодия, логично продолжает ряд параметров, определяющих голос человека, поскольку она является важной фонетической характеристикой. Пересечение объемов понятий «голос человека» как будущего объекта правовой охраны и «речевая просодия» становится очевидным при анализе определений обеих категорий. С.И. Ожегов определяет голос как совокупность звуков, возникающих в результате колебания голосовых связок. Под просодией понимают систему озвучивания слогов речи, интонационные и фонетические особенности произношения и артикуляции звуков. Как видим, обе категории используются для описания способности человека произносить множество разнообразных звуков для осуществления голосовой коммуникации. По этой причине такой параметр, как просодия, должен быть изучен и учтен при дальнейшей детализации правовой нормы об охране голоса гражданина, сформулированной в упомянутом законопроекте в самом общем виде. Конкретизация признаков объекта регулирования необходима для выработки работоспособного механизма регламентации отношений вовлеченных в них субъектов. Без такого уточнения предлагаемая норма может оказаться неэффективной для практического разрешения соответствующих споров.

Обеспечение эффективной правовой охраны голоса с учетом речевой просодии приобретает еще бóльшую актуальность в условиях бурного развития генеративных нейростетевых инструментов. Имитация голоса, окрашенного специфически узнаваемой просодией, позволяет приблизить подражательный результат нейросетевой генерации к оригиналу, повысить степень его убедительности и усилить впечатление достоверности, что, однако, может использоваться не только в позитивных и легальных целях, но и в противоправной практике для осуществления обманных и мошеннических действий.

АНАЛИЗ, ОЦИФРОВКА И АВТОМАТИЗАЦИЯ ГЕНЕРАЦИИ РЕЧЕВОЙ ПРОСОДИИ

По материалам [27, 28] и из собственного опыта автором этой статьи и его коллегами был установлен ключевой фактор, который обуславливает чистоту эксперимента и релевантность измеренных данных. Этот фактор представляет собой использование повторяющихся однотипных образцов речевых фрагментов. Повторяющиеся однотипные образцы речевых фрагментов снижают рассеивание значений данных. Данный подход позволяет более точно выделить и проанализировать те элементы просодии, которые относятся к выражению требования [29]. В противном случае разнообразие ситуаций и контекстов увеличивает зашумленность и дисперсию результатов измерений, осложняя интерпретацию результатов.

Образцы строго регламентированной речи, в частности высказывания исполняющих служебные обязанности полицейских, — наиболее подходящий материал для исследования просодии долженствования. Это утверждение опирается на тот факт, что регулирование поведения и речи официального лица снижает вариативность форм его выражения, делает высказывания более стандартизированными, уменьшает случайные влияния и как следствие предоставляет более качественные данные для структурного анализа [30].

Обнаруженная на видеохостинге база данных образцов, соответствующих критериям строго регламентированной однотипной коммуникации, выступает основой повторяемости характера обращений и качества результатов исследования. Найденные материалы, содержащие четко выраженные требования полицейских в типовых ситуациях, обеспечивают консистентность экспериментальных данных. Эти материалы позволяют выявить и оцифровать характерные признаки просодии конкретного типа — просодии долженствования, которая проявлена в однородных условиях.

Анализ одиннадцати образцов высказываний позволил выявить, что требования, выраженные сотрудниками полиции, имели различные частоты, что отражено на графике в герцах (Гц). Некоторые требования имели высокую частоту, указывающую на энергичность и интенсивность коммуникации, а другие — низкую частоту, свидетельствующую о более спокойном и сдержанном выражении требований.

Что касается громкости, измеренной в децибелах (дБ), то графики показали различные ее уровни, связанные с требованиями и спокойными обращениями. Некоторые требования сопровождались высоким уровнем громкости, что указывает на эмоциональность и интенсивность коммуникации, а другие имели более низкий уровень громкости, свидетельствующий о спокойном и непринужденном настрое.

Кроме того, была проанализирована продолжительность каждого высказывания в секундах. Графики позволили наглядно представить различия в продолжительности требований и спокойных высказываний. Некоторые требования были более длительными, что может указывать на более подробное и развернутое их изложение, а другие оказались краткими и сжатыми.

Графическое представление и расшифровка результатов эксперимента

Частота основного тона (далее — ЧОТ), громкость и длительность высказывания являются определяющими параметрами просодии, которые позволяют вывить ее тип, оцифровать и воспроизвести. Приведу аргументы, которые подчеркивают значимость этих критериев.

ЧОТ может быть связана с эмоциональным состоянием говорящего. Повышенная ЧОТ указывает на возбуждение или стресс, тогда как пониженная — на спокойное состояние говорящего. В контексте речевых просодий ЧОТ может указывать на специфическую интонационную модальность, характерную для высказываний определенного типа [31].

Идентификация акцентов[7] и эмфазы[8]. Высокая ЧОТ может также указывать на характерный акцент или эмфазу, что важно для стандартизации речевых образцов, где определенные слова или фразы подчеркиваются согласно сложившейся практике [32].

Громкость и выделение важных частей высказывания. Громкость играет ключевую роль в выделении элементов речи. При типизации просодии классификация элементов речи по громкости может помочь определить вид обращения среди таких, например, как инструкция, требование или предупреждение [33].

Эмоциональная окраска. Громкость может передавать эмоциональную окраску высказывания. В регламентированных образцах коммуникации подчеркнутая громкостью экспрессивная окраска выступает инструментом передачи эмоционального оттенка, который особенно важен в специфических сценариях, например в требовательных обращениях, произносимых служащими полиции [34].

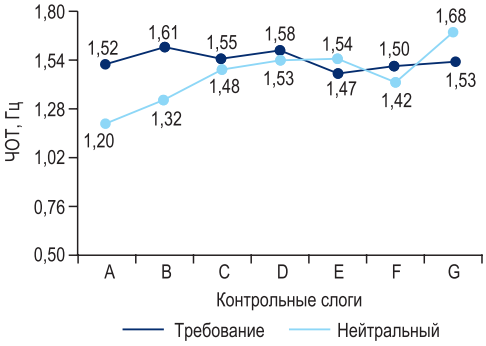
Длительность и структура высказывания. Длительность определяет темп и ритм речи, что важно для структурирования высказывания. Для стандартных просодий характерны определенный темп и длительность отдельных элементов речи [35].

Поддержание стандартов коммуникации. Стандартизированная длительность элементов речи направлена на поддержание определенных стандартов коммуникации, что особенно важно в профессиональных областях, где точность и ясность высказывания имеют большое значение [36].

Многомерный синтетический анализ просодии с учетом ЧОТ, громкости и длительности высказывания позволяет описать и в дальнейшем интерпретировать образцы стандартизированной коммуникации, а также воссоздать нюансированный речевой портрет говорящего [37].

График на рис. 1 отображает соотношение ЧОТ двух видов речевых образцов — требования и нейтрального высказывания. На графике представлены две линии, одна соответствует требованию, а другая — нейтральному обращению.

Рис. 1. Соотношение ЧОТ высказывания требования и нейтрального обращения



Диапазон тона нейтрального высказывания выше параметра данного признака требования в 5 раз (0,5/0,1). Диапазон требования имеет свою особенность: он развивается преимущественно в пределах четвертого и пятого уровней. Диапазон тона нейтрального высказывания развивается от третьего уровня до пятого, происходит градуальный рост кривой ЧОТ, но с пятого контрольного слога кривая ЧОТ резко идет вниз, затем стабилизируется и останавливается на пятом уровне. Максимальные значения зафиксированы в разные контрольные слоги: у первой фразы — в завершении предтакта[9], а у вто-

[7] Акцент в лингвистическом контексте представляет собой особенности произношения звуков, интонации, ритма и других элементов речи, которые отличают говорящего от носителя стандартного или иного диалекта. Акцент является результатом влияния языковой среды, в которой говорящий вырос, и может проявляться в форме различий в звуковом строе, ударении слов, в интонации и даже в выборе лексики.

[8] Эмфаза в лингвистическом контексте представляет собой явление акцентуации или усиления определенных элементов в речи для достижения эффекта выделения, подчеркивания важности или передачи эмоционального оттенка в высказывании. Эмфатическое ударение, или выделение, может воздействовать на различные компоненты речи, такие как звуковая интонация, громкость, длительность и даже выбор лексики.

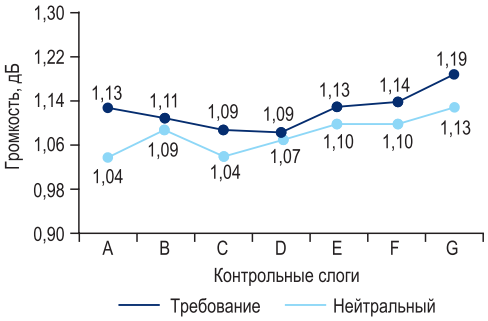
[9] Предтакт в лингвистическом контексте представляет собой феномен, характеризующийся предварительным фонетическим воздействием на артикуляционные аспекты речи перед началом произнесения конкретного фонемного элемента. Этот явный префонетический эффект взаимодействия с артикуляционной системой говорящего может влиять на последующие фонемные выражения, оказывая воздействие на их артикуляционные параметры, интонацию и ритмические характеристики. Артикуляционные приготовления, предше-

рой — в завершении затакта[10]. Консенсуса в отношении минимальных значений также не наблюдается, потому что у первой фразы они зафиксированы в участке каденции[11], а у второй в предтакте.

Что касается тональных уровней слогов фраз, то у первой фразы в предтакте сначала происходит рост, который прерывается и идет на спад, во второй же фразе идет постоянный рост. Похожая ситуация — и в ритмическом корпусе[12], у первой фразы также сначала рост, затем падение, у второй фразы постоянный рост, но уже в меньшей степени. На участке каденции первая фраза выходит на плавный рост, который дублируется в затакте. Вторая фраза в каденции показывает резкое изменение, но в отрицательную сторону, в затакте же наблюдается рост.

График на рис. 2 показывает соотношение громкости двух видов высказываний в ситуации требования и нейтрального тона. На графике представлены две линии, одна соответствует требованиям, а другая — нейтральным обращениям.

Рис. 2. Соотношение громкости двух видов высказываний



Диапазон интенсивности[13] обеих фраз находится примерно на одном уровне. Диапазон интенсивности первой фразы охватывает третий и четвертый уровни, а кривая интенсивности второй фразы развивается на втором и третьем уровнях. Максимальные показатели фраз находятся на разных уровнях, у первой фразы максимальный показатель равен 1,2, у второй — 1,1, но они расположены в одной зоне фонетической сегментации[14]. Динамическая вершина требования и нейтрального высказывания зафиксирована в затакте. В минимальном значении фразы схожи — оба значения находятся в ритмическом корпусе, но у нейтрального обращения минимальное значение находится в предтакте.

Кривая требования фиксирует показатели падения от начала предтакта до начала ритмического корпуса, в ритмическом корпусе происходит стагнация, за которой следует постоянный рост до конца затак-

ствующие фонеме, рассматриваются как фаза предтакта, включающая в себя координацию мускулатуры артикуляционных органов и установление определенных физиологических конфигураций в преддверии произнесения целевой фонемы. Предполагается, что предтакт может служить механизмом оптимизации артикуляции и обеспечивать более эффективное выполнение фонетических задач, таких как достижение оптимальной аккомодации артикуляционных аппаратов для предстоящего речевого элемента.

[10] Затакт — это предшествующая активация артикуляции перед началом произношения конкретного фонемного элемента в рамках высказывания. Затакт включает в себя координацию мускулатуры артикуляционных органов и предварительную организацию параметров перед выражением следующего фонемного сегмента. Он направлен на оптимизацию артикуляционных процессов для точного и эффективного произнесения последующего фонетического элемента.

[11] Каденция в лингвистическом контексте представляет собой структурно-функциональный элемент, характеризующийся завершением или паузой в высказывании, который, в свою очередь, вносит определенный моментарный пунктуационный отсчет в рамках лингвистической формы. Данный феномен демонстрирует свою важность в контексте ритмической и интонационной организации языка, предоставляя акустические и семантические индикаторы завершения или перехода к следующей фразе. В лингвистических исследованиях каденция рассматривается как средство структурирования и маркировки лингвистических единиц, что подчеркивает ее сущностную роль в оркестрации синтактической и прагматической организации речевого материала. Разъяснение многозначных аспектов каденции обогащает наше понимание механизмов фразовой организации и взаимодействия элементов в рамках вербальной коммуникации.

[12] Ритмический корпус — это структурированное агрегирование ритмических элементов в музыкальном или лингвистическом контексте, представляющее собой набор данных, охватывающий временны́е отношения между звуковыми или лингвистическими сегментами. В музыкальном смысле ритмический корпус может включать в себя информацию о длительностях, акцентах, темпе и других параметрах, связанных с организацией времени в музыкальном произведении. В лингвистическом контексте ритмический корпус может содержать данные о длительности и интонационных аспектах речи, предоставляя базу для анализа ритмических особенностей языка. Ритмический корпус служит инструментом для исследования и описания временны́х структур в музыке или языке, а также для проведения сравнительного анализа исследуемых ритмических явлений.

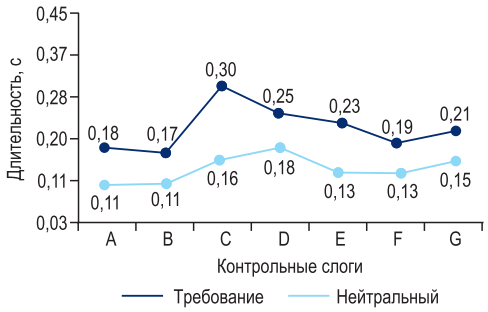
[13] Под интенсивностью в анализе речи понимается степень выраженности, энергичности или силы, с которой произносится или выражается какой-либо элемент речи в контексте. Интенсивность может относиться к различным аспектам речи, включая громкость, эмоциональную окраску, ударение, артикуляцию, а также использование средств паралингвистической коммуникации.

[14] Под фонетической сегментацией в анализе речи понимается процесс разбиения непрерывного потока звуков, образующих высказывание, на дискретные фонетические единицы. Этот процесс является неотъемлемой частью фонетического исследования и направлен на выделение индивидуальных фонетических единиц, таких как фонемы, дифтонги, согласные и гласные, а также их последовательности и взаимодействия внутри произносимого текста.

та. Кривая нейтрального тона показывает другую картину: в предтакте рост, переходящий в падение, затем в ритмическом корпусе — рост до участка каденции, в котором происходит стагнация, далее продолжается рост.

График на рис. 3 отображает соотношение длительности двух видов высказываний в ситуации требования и нейтрального обращения. На графике представлены две линии, одна соответствует требованиям, а другая — нейтральным заявлениям.

Рис. 3. Соотношение длительности двух видов высказываний



Средняя скорость фонации[15] обеих фраз находится на уровне семи слогов в секунду. Требование располагается на втором, третьем и даже четвертом уровнях, а нейтральное высказывание — лишь на первом и втором уровнях. Расположение максимальных значений фраз находится в ритмическом корпусе, но присутствует дифференциация в точках. Минимальные показатели зафиксированы в предтакте с разницей лишь в местоположении точек.

Похожие контуры слоговой долготы[16] прослеживаются практически на всем протяжении развертывания сопоставляемых фраз. Исключение составляет лишь ритмический корпус, где кривая длительности речи нейтрального обращения начинает идти вниз на одну точку позже по сравнению с темпом речи требования. Различия проявляются в уровневых темпоральных параметрах[17], как видно из графика на рис. 3, скорость и замедление речи нейтрального высказывания более плавные и не такие контрастные.

Лингвистическая интерпретация результатов эксперимента

Частота основного тона высказывания может изменяться в зависимости от различных факторов, включая выделение определенных слов и выражение эмоций. В случае требования ЧОТ может быть выше, чем при нейтральном изречении. Обычно, когда люди выражают требование, они придают ему большее значение и силу. Голос становится более энергичным и громким, чтобы подчеркнуть важность и привлечь внимание к требованию. ЧОТ требования оказывается повышенной, чтобы придать выражению большую динамику и убедительность. В нейтральных высказываниях собеседники не выражают эмоционального отношения к сказанному, поэтому ЧОТ обычно остается стабильной и значительно не изменяется. Однако в конце фразы может происходить незначительное увеличение ЧОТ. Это может быть связано с небольшим изменением интонации для логического завершения изречения с явным заключением, подчеркивания важности последнего слова либо идеи. Итак, в требовании уровень ЧОТ может быть выше, чтобы придать силу обращению, а в нейтральном высказывании ЧОТ может лишь незначительно возрастать в конце фразы для ее акцентированного завершения.

Если говорить о различии громкости (интенсивности) фраз двух исследуемых видов, то громкость требования в среднем выше, чем громкость нейтральной сентенции, так как это обусловлено спецификой работы служащих полиции. Использование повышенной громкости в требованиях позволяет им подчеркнуть свой авторитет и напомнить гражданам о необходимости выполнять их инструкции. Громкие команды помогают установить контроль и привлечь внимание находящихся вблизи лиц, чтобы предотвратить возможные нарушения или пресечь преступления. В некоторых случаях, особенно в шумной обстановке или быстро развивающихся ситуациях, громкость требований полицейских способствует более эффективной коммуникации. Это позволяет полицейским передать инструкции и указания быстро и четко, чтобы избежать недоразумений, снизить риск ошибочной интерпретации и неправильных реакций.

Полицейские часто оказываются в опасности или в потенциально рисковых ситуациях, в которых

---

[15] Скорость фонации в лингвистическом контексте представляет собой меру темпа произношения речи, измеряемую количеством произносимых фонем, слов или слогов в единицу времени. Этот параметр является ключевым аспектом в акустическом анализе и описывает, с какой быстротой речь произносится говорящим.

[16] Под слоговой долготой в анализе речи понимается мера продолжительности произносимого слога в речи. Этот параметр измеряется величиной времени, отражая длительность времени, в течение которой артикулируется определенный слог внутри произносимого слова или фразы.

[17] Под темпоральными параметрами в анализе речи понимают характеристики, связанные с временны́ми аспектами произношения и организации речевого потока. Эти параметры охватывают различные временные аспекты, такие как длительность, темп и темпоральная структура, и являются ключевыми для изучения ритма, темпа и синхронизации в речи.

их безопасность зависит от способности привлечь внимание и обеспечить исполнение их требований. Громкость обращений помогает решить эти задачи, а также транслировать уверенность и поддерживать безопасность. Эти и многие другие аспекты обусловливают и обосновывают более высокую интенсивность требования. В начале требования его громкость может несколько снижаться, потому что говорящий стремится быть дипломатичным или ненавязчивым, особенно в случаях, когда межличностные отношения связаны с уважением и тактичностью. Он может стараться убедить собеседника с помощью аргументов и логики, а не с помощью силы или настойчивости. Возрастание громкости может происходить с течением времени по нескольким причинам. Во-первых, если требование не было сразу же выполнено, говорящий может увеличить его громкость для трансляции настойчивости и готовности добиться своей цели. Во-вторых, у говорящего могут накапливаться раздражение, возмущение, негодование и другие отрицательные эмоции из-за несоответствующей реакции адресата послания, и он готов проявить твердость намерений через повышение громкости требования.

Нейтральные высказывания могут иметь скачок громкости в начале сентенции, чтобы привлечь внимание слушателя к актуальной информации. Это повышение громкости может быть вызвано сознательным выбором говорящего или автоматической реакцией вследствие важности передаваемых данных. Однако затем, если нейтральное изречение не требует немедленной реакции или эмоционального отклика, громкость может снизиться и стабилизироваться на обычном уровне. Так происходит по мере того, как говорящий выполняет свою роль информатора, не требуя дальнейшего внимания слушателя. Если же состояние невозмутимости говорящего нарушается либо информация становится более важной или эмоционально заряженной, громкость может постепенно возрастать, чтобы сфокусировать внимание адресата на смысловой нагрузке событий или фактов.

Длина фразы в ситуации требования может быть больше, чем в нейтральном высказывании, поскольку в таких случаях обычно используются более формальные и точные выражения. Например, вместо простого и короткого «Come out» («Выходите») сотрудник сказал «Come out here for me» («Выходите сюда ко мне»). Это связано с тем, что в ситуациях требования необходимо быть ясным и точным в выражениях, чтобы избежать недопонимания и конфликтов. Кроме того, в требовании часто используются сильнее эмоционально окрашенные выражения, такие как «Don't be swearing at the cops» («Не ругайтесь на полицейских»), «You better come out and show me some hands right now» («Вам лучше выйти и показать мне руки прямо сейчас»), «Why don't you go and step out of your vehicle?» («Почему бы вам не выйти из машины?»). Дополнительные конструкции и выражения увеличивают длину фразы, поскольку уточняют смысл и сообщают эмоциональный посыл требования.

Относительная скорость произнесения фразы может зависеть от интонации, эмоций и контекста. Например, если требование высказано с яростной интонацией, то оно может быть произнесено быстрее, но с большей энергией и убедительностью. С другой стороны, если требование произнесено спокойно и уверенно, то реплика может быть сказана медленнее, но с большей ясностью и точностью. Как правило, полицейские в ситуациях требования используют более формальные и точные выражения, чтобы избежать недопонимания и конфликтов. Однако они также могут использовать эмоционально окрашенные выражения, чтобы убедительнее донести свою позицию. Например, при задержании подозреваемого они могут использовать составные, усиливающие требование конструкции, такие как «Положите руки на голову и не двигайтесь» или «Вы арестованы за нарушение закона». В таких ситуациях фраза может быть длиннее, чем в нейтральном высказывании, чтобы убедительнее и детальнее довести свое требование и избежать непредвиденных ситуаций.

Если сравнить длительность высказываний на рис. 3, то можно заметить, что к точке C идет замедление, а затем ускорение. Это может быть связано с тем, что человек начинает разговор с целью привлечь внимание. Когда он убедился, что собеседник его слушает, он может перейти к более мягкому и понятному общению, чтобы добиться нужного результата без лишней конфронтации. Ускорение речи может быть связано с тем, что человек хочет достичь быстрого и эффективного решения проблемы или выполнения задачи, но при этом не хочет создавать напряжения или конфликта.

Сравнение ЧОТ на обоих графиках показывает, что у нейтрального высказывания более выраженный разброс значений. Это может быть обусловлено следующими факторами. Интерпретация контекста: нейтральное высказывание в зависимости от ситуации может быть воспринято по-разному. Некоторые люди склонны воспринимать его характер как позитивный, а другие, наоборот, — как негативный. Индивидуальные предпочтения: люди могут иметь разные предпочтения в отношении тона и стиля выражения своих изречений. Например, одни отдают предпочтение более официальному и ровному тону, в то время как другие — более эмоциональным и выразительным интонациям. Культурные различия: разные культуры имеют разные стандарты в отношении тональности речи. На-

пример, в некоторых культурах более прямолинейное и открытое высказывание может быть воспринято как нейтральное, в то время как в других культурах нейтральным считается более дипломатичное и осторожное выражение мнения. В то же время тональность требования должна быть ровной, чтобы обеспечить четкость и однозначность его смысла. Если тональность требования будет меняться или колебаться, это может привести к недопониманию или неправильному выполнению задачи. Поэтому важно формулировать требования ясно и точно, придерживаясь одной тональности на всем протяжении сообщения.

В табл. 1 сведены полученные в результате эксперимента данные.

Таблица 1. Значения измеренных параметров высказывания требования

1	*Фраза*	*You better come out and show me some hands right now*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,19	0,23	0,34	0,43	0,52	0,26	0,27
	Гц		307	343	342	327	295	335	290
	дБ		61	64	65	64	64	63	63
2	*Фраза*	*Come to my voice*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,05	0,16	0,06	0,08	0,14	0,14	0,17
	Гц		330	362	290	380	355	335	290
	дБ		60	62	60	63	64	63	61
3	*Фраза*	*Alright both of you take a breath*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,21	0,15	0,07	0,11	0,18	0,7	0,25
	Гц		170	95	175	231	303	211	150
	дБ		69	70	72	68	68	63	65
4	*Фраза*	*Just do you*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,15	0,17	0,04	0,13	0,14	0,16	0,27
	Гц		148	157	133	188	207	183	162
	дБ		66	70	65	70	67	67	65
5	*Фраза*	*Don't be swearing at the cops*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,25	0,09	0,10	0,19	0,12	0,10	0,33
	Гц		162	225	176	187	178	165	159
	дБ		66	66	62	68	65	56	57
6	*Фраза*	*Come out here for me*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,05	0,10	0,05	0,09	0,07	0,12	0,20
	Гц		184	186	172	189	180	163	123
	дБ		58	68	65	68	60	68	61
7	*Фраза*	*Put your right foot in front of our left and then just stand there*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,11	0,06	0,41	0,23	0,55	0,28	0,68
	Гц		182	150	195	219	180	150	167
	дБ		70	62	68	65	65	66	65
8	*Фраза*	*Why don't you go and step out of your vehicle*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,14	0,22	0,11	0,16	0,27	0,14	0,53
	Гц		250	280	287	214	241	243	184
	дБ		69	64	68	62	65	63	65
9	*Фраза*	*Gotta be under here somewhere*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0.14	0.07	0.10	0.17	0.11	0.08	0.32
	Гц		155	134	116	120	114	106	118
	дБ		72	66	66	68	66	60	63

Окончание табл. 1

10	*Фраза*	*Why are we here*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,13	0,03	0,04	0,03	0,10	0,10	0,12
	Гц		216	281	290	281	256	295	326
	дБ		74	74	74	70	68	67	66
11	*Фраза*	*These pants are green*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
	Секунды		0,16	0,06	0,15	0,06	0,016	0,13	0,21
	Гц		167	223	179	150	129	220	140
	дБ		68	64	74	70	68	74	66

РАЗРАБОТКА МОДУЛЯ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ РЕЧЕВОЙ ПРОСОДИИ

Программирование прототипа модуля генерации речевой просодии

Полученные результаты позволяют создать прототип программной утилиты для автоматического модулирования сигнала в целях придания ему просодии долженствования, выбранной для проведения эксперимента. Разработанная программа для ЭВМ может быть интегрирована в генеративную искусственную нейронную сеть (ИНС), чтобы реализовать речевую коммуникацию в соответствующей просодии. При решении задачи был задействован языковой агент ChatGPT. Данный инструмент был использован для создания кода программы, которая преобразует аудиосигнал.

Созданный лингвистической моделью в результате серии итеративных запросов исследователя программный код изменяет параметры входного аудиосигнала в соответствии с экспериментально установленными значениями для просодии требования, представленными в таблице. Модулированный выходной сигнал сохраняется и воспроизводится для оценки результатов работы утилиты. Данный подход позволяет автоматизировать процесс звуковой обработки речи, обеспечить гибкую настройку параметров генерации и широкие возможности для анализа параметров вербальной коммуникации. Ниже записан показательный фрагмент кода программы, разработанной для генерации речевой просодии.

```
from pydub import AudioSegment
from pydub.playback import play

def modulate_audio(input_file, output_file, duration_factor=1.0, volume_factor=1.0, pitch_factor=1.0):
    # Загрузка звукового файла
    sound = AudioSegment.from_file(input_file)

    # Изменение длительности слога
    duration_changed = sound.speedup(playback_speed=duration_factor)

    # Изменение громкости (интенсивности)
    volume_changed = duration_changed + 10 * volume_factor # Поднимем или понизим громкость

    # Изменение герцовки (высоты голоса)
    pitched = volume_changed._spawn(volume_changed.raw_data, overrides={
        "frame_rate": int(sound.frame_rate * pitch_factor)
    })

    # Сохранение модулированного звукового файла
    pitched.export(output_file, format="wav")

    # Проигрывание оригинала и модулированного звука
    print(«Исходный файл:»)
    play(sound)
    print("Модулированный файл:")
    play(pitched)

# Пример использования
modulate_audio("input.wav", "output_modulated.wav", duration_factor=0.8, volume_factor=1.5, pitch_factor=1.2)
```

Процесс создания программного модуля включил в себя следующие этапы.

1. Выбор инструмента для генерации исходного текста модуля. В качестве инструмента была выбрана ИНС GPT-3 в силу того, что она является одним из наиболее передовых доступных агентов автоматизированного программирования на языке Python.

2. Интеграция. С помощью встроенного в GPT-3 API (интерфейса программирования приложений) мы реализовали возможность аппаратно-программного взаимодействия с этой языковой моделью.

3. Разработка кода модуля. Модуль генерации речевой просодии был запрограммирован итерационно-циклическим методом отправки запросов в GPT-3, обработки полученных ответов и формирования уточняющих заданий. Цикл повторялся до достижения готового к тестированию результата.

4. Извлечение и контейнеризация. Перспективные версии исходного текста GPT-3 извлекались, оформ-

лялись в виде функциональной программной библиотеки и снабжались интерфейсами ввода и вывода данных для последующего тестирования.

5. Тестирование и оценка результатов. Синтезированный модулем речевой сигнал сравнивался с эталонным образцом из базы данных по лингвистическим параметрам, указанным выше, а также на слух непредвзятого пользователя[18]. По результатам сравнения оценивалась эффективность функционирования модуля и формировались корректирующие запросы для ИНС GPT-3 в целях получения следующей, более совершенной версии исходного текста программы. Этот процесс повторялся до сближения параметров эталонного и тестируемого образцов на уровне 70–92%.

Экспериментальная верификация модуля генерации речевой просодии

Необходимость разработки программного модуля генерации речевой просодии обусловлена потребностью в экспериментальной верификации результатов, полученных в ходе исследования. Созданная программная утилита реализована с использованием языка программирования Python и ИНС GPT-3 от компании OpenAI.

Испытания модуля проводились с использованием высококачественного микрофона, программного обеспечения для обработки аудиоинформации, а также плоских, лишенных выраженной просодии речевых фрагментов (табл. 2). Модуль принимал входной сигнал и вносил изменения в ритм, интонацию, эмоциональные нюансы и другие параметры речи в соответствии с предварительно запрограммированными настройками. На выходе модуля снимался аудиосигнал с измененными параметрами, реализующими просодию требования. В каждом сеансе испытаний были использованы идентичные параметры входного сигнала и программной модуляции просодии.

Эксперимент был повторен десять раз для доказательства стабильности показателей. Результаты испытаний оказались корректными и повторяемыми, что свидетельствует об обоснованности выбранной методики исследования и воспроизведения такого лингвистического элемента, как речевая просодия, а также о ее качественной реализации.

Таблица 2. Значения сгенерированных модулем параметров просодии требования

Исходный образец	*Person calling for help*	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды		0,14	0,08	0,15	0,07	0,17	0,14	0,19
Гц		174	213	168	160	131	225	146
дБ		69	67	72	71	65	70	67
Генерация 1		*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды		0,17	0,12	0,22	0,11	0,21	0,17	0,23
Гц		181	222	175	166	140	231	152
дБ		74	69	76	73	67	72	70
Генерация 2		*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды		0,16	0,12	0,21	0,10	0,21	0,18	0,22
Гц		179	220	178	168	137	230	151
дБ		74	68	75	72	67	71	70
Генерация 3		*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды		0,16	0,11	0,22	0,11	0,20	0,17	0,21
Гц		178	221	176	168	141	229	151
дБ		73	69	74	73	66	72	68
Генерация 4		*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды		0,17	0,12	0,21	0,11	0,21	0,18	0,23
Гц		179	222	176	167	139	230	150
дБ		69	67	72	72	66	70	67

[18] Lay Listener — тест непредвзятого пользователя, также известный как Lay listener, представляет собой концепцию, описывающую способность восприятия звуков и аудиоинформации обыденным, далеким от науки пользователем, лишенным специализированных знаний в области аудиотехники или звуковой обработки. Термин «Lay listener» подчеркивает отсутствие технического образования или опыта в сфере звуковых технологий у данного слушателя.

Окончание табл. 2

Генерация 5	*Первая*	Вторая	Третья	*Четвертая*	Пятая	Шестая	Седьмая
Секунды	0,17	0,11	0,20	0,10	0,22	0,17	0,21
Гц	177	222	179	168	137	231	152
дБ	72	69	74	72	67	71	68
Генерация 6	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	Шестая	Седьмая
Секунды	0,17	0,12	0,22	0,11	0,21	0,17	0,23
Гц	181	222	175	166	140	231	152
дБ	74	69	76	73	67	72	70
Генерация 7	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды	0,17	0,13	0,22	0,11	0,21	0,18	0,23
Гц	180	221	176	167	138	231	152
дБ	72	68	74	72	67	72	69
Генерация 8	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды	0,17	0,12	0,22	0,10	0,21	0,17	0,23
Гц	178	222	175	166	140	230	151
дБ	73	69	76	73	67	71	69
Генерация 9	*Первая*	*Вторая*	*Третья*	*Четвертая*	*Пятая*	*Шестая*	*Седьмая*
Секунды	0,17	0,12	0,22	0,10	0,21	0,17	0,21
Гц	179	219	177	169	143	229	151
дБ	73	69	73	71	66	72	68
Генерация 10	*Первая*	Вторая	Третья	Четвертая	Пятая	Шестая	Седьмая
Секунды	0,16	0,12	0,21	0,11	0,21	0,17	0,22
Гц	179	222	176	167	138	228	153
дБ	72	69	74	72	67	73	69

Важным итогом эксперимента стал факт подтверждения гипотезы о возможности автоматической генерации коммуникативных речевых просодий с помощью программно-аппаратных компьютерных средств.

Практическая применимость и масштабирование прототипа

Разработанный прототип представляет собой масштабируемую систему, которая может быть дополнена другими просодиями и интегрирована в генеративную нейронную сеть для реализации функции человекоподобной голосовой коммуникации.

Прототип модуля создан на основании предложенной в статье методики. Это означает, что система может быть обучена и настроена для воспроизведения различных интонационных, ритмических и эмоциональных параметров речи, характерных для просодий других видов. Увеличение номенклатуры поддерживаемых просодий расширит функциональные возможности и полезность прототипа.

Масштабируемость относится к способности системы расширяться и приспосабливаться к новым требованиям, сохраняя при этом эффективность и стабильность [38]. Модульная структура разработанного прототипа призвана обеспечить гибкость и расширяемость его программной системы, а также поддержку библиотеки просодий без существенного изменения архитектуры данной утилиты.

Созданный модуль может быть встроен в нейросетевые структуры с машинным и глубоким обучением. Научно-лингвистическая природа данного проекта открывает возможность использования разработанного прототипа в системах искусственного интеллекта, основанных на больших языковых моделях. Для этого прототип должен быть адаптирован к стандартам и интерфейсам, используемым в данной области, включая настройку взаимодействия с входными и выходными слоями ИНС. Осуществление такой интеграции сделает данный прототип эффективным инструментом для разработки систем синтеза речи с учетом различных лингвистических и эмоциональных контекстов.

ЗАКЛЮЧЕНИЕ

Развитие больших языковых моделей, способных генерировать релевантные ответы на вопросы и поддержи-

вать содержательную коммуникацию с пользователем, ставит перед их исследователями и разработчикам новые задачи. Сегодня сформировался тренд на совершенствование аудиоречевых каналов коммуникации искусственных нейронных сетей, которые призваны достоверно имитировать особенности человеческой речи.

В настоящем исследовании выдвинута гипотеза о возможности параметризации, оцифровки, генерации и последующего воссоздания фонетических особенностей речевого потока, характерного для человека. Для этого выбран наиболее яркий и простой пример просодии требования, которая контрастно высвечивает критерии, определяющие данный вид речевой модальности. Минимально необходимый комплект изученных показателей включил в себя частоту основного тона, громкость и длительность произнесения анализируемых фраз.

Проведенный по указанным выше параметрам анализ одиннадцати речевых образцов показал наличие закономерностей, позволяющих отличать нейтральную тональность обращений от тех, что связаны с предъявлением требования. Вытекающим из данного факта следствием оказалась возможность воссоздавать эти показатели и придавать таким образом нейтральным высказываниям тон и характер требования.

Для проверки данного предположения на практике было осуществлено прототипирование программного модуля, выполняющего задачу изменения параметров тонально нейтрального высказывания так, чтобы после обработки частоты основного тона, длительности и громкости произнесенных слогов оно приобретало просодию требования. Создание прототипа было выполнено с помощью нейросетевого генератора программного кода на языке Python итеративным методом постановки задачи, оценки результата ее выполнения и формулировки следующего запроса, развивающего полученный продукт в необходимом направлении.

Разработанная программная утилита прошла тестирование на предмет качества модуляции сигнала в целях приближения его характеристик к искомым значениям. Результаты испытаний оказались корректными и повторяемыми, эксперимент был воспроизведен десять раз для доказательства стабильности показателей, его итоги представлены в виде таблицы измеренных величин.

Факт работоспособности созданного модуля позволил сделать заключение о корректности выдвинутой в статье гипотезы относительно реализуемости моделирования речевой просодии компьютерными средствами, практической применимости теоретических выводов исследования и его материального результата в виде прототипа программного продукта для обучения больших языковых моделей человекоподобной коммуникации.

Полученный результат также подчеркнул актуальность разработки эффективной правовой нормы для охраны звуковой или вокальной идентичности гражданина, именуемой в повседневном обиходе голосом человека, в конкретных параметрах, которые позволят количественно и качественно сравнивать между собой речевые или певческие образцы для информированного и справедливого разрешения связанных с ними юридических споров. Объемы понятий «голос» и «речевая просодия» пересекаются. Просодия объективируется в звуковой форме и по этой причине может интерпретироваться как голос гражданина. Правовое регулирование использования голоса человека требует уточнения этого феномена в определенных характеристиках, одной из которых выступает речевая просодия.

Анализ просодии особенно эффективен в области права и правосудия. В судебных разбирательствах истолкование интонации может использоваться для обнаружения истинности или ложности намерения говорящего, определяющего исход дела. Просодия может служить индикатором эмоционального состояния человека. Эмоции, выраженные через голос, могут указывать на степень искренности и уверенности говорящего (свидетеля, истца, ответчика и т.д.). Так, если гражданин говорит с повышенными интонацией и темпом, это может свидетельствовать о нервозности, неуверенности или надуманности его высказывания. Анализ просодии позволяет точнее понимать контекст и оценивать достоверность показаний. Таким образом, учет извлекаемой из просодии информации становится фактором повышения качества судопроизводства.

В условиях растущей популярности технологий синтеза и распознавания речи знания о просодии и необходимость принимать эту звуковую характеристику индивидуума во внимание становятся еще более актуальной задачей. Современные системы могут не только фиксировать высказывания, но и анализировать их произношение и эмоциональную окраску. Это открывает новые возможности для защиты прав граждан при посягательствах на их интересы со стороны мошенников, поскольку позволяет надежно идентифицировать и интерпретировать их голосовые данные. По этой причине интеграция просодических аспектов в законодательную и правоприменительную практику — это шаг к более эффективной охране прав граждан и справедливому правосудию.

СПИСОК ИСТОЧНИКОВ

1. Language Models are Unsupervised Multitask Learners / A. Radford, J. Wu, R. Child et al. OpenAI, 2019. P. 1–24.

2. *Jurafsky D., Martin J.H.* Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition/ Ed. 3. Lond.: Pearson, 2019. P. 1–16.
3. *Потапов В.В., Казак Е.А.* Речевая коммуникация в сетевых структурах: между глобальным и локальным: сб. науч. трудов. М.: РАН. ИНИОН, Отдел языкознания, 2022. 280 с.
4. *Чубиков А.В.* Исследование методов для повышения качества озвучивания текста в системах речевого синтеза // Вестник Российского университета дружбы народов. 2018. № 4(20). С. 551–558.
5. *Huang L., Yu D.* Linguistic Features for Speech Synthesis: A Review // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2011. Vol. 19. No. 7. P. 1723–1734.
6. *Xu Y., Jiang X.* Voice Synthesis Based on Timbral Characteristics // Proceedings International Conference on Electrical and Control Engineering. Melaka, Malaysia, 2019. P. 124–128.
7. *Tao X., Li S.* Emotional Speech Synthesis Based on Prosodic Attributes // Proceedings IIEEE International Conference on Acoustics, Speech and Signal Processing. New Orleans, LA, USA, 2017. P. 511–515.
8. *Chung D., Kim H., Ahn S.* An integrated study of user acceptance and resistance on voice commerce // International Journal of Innovation and Technology Management. 2022. Vol. 19. No. 7.
9. *Singh R., Jiménez A., Øland A.* Voice disguise by mimicry: deriving statistical articulometric evidence to evaluate claimed impersonation // IET Biometrics. 2017. Vol. 6. No. 4. P. 282–289.
10. *Матвеев А.Г., Мартьянова Е.Ю.* Гражданско-правовая охрана голоса человека при его синтезе и последующем использовании // Ex jure. 2023. № 3. С. 118–131. DOI: 10.17072/2619-0648-2023-3-118-131
11. *Аникян Т.С.* Экспрессивный потенциал просодии в инаугурационной риторике (на материале обращения Джозефа Байдена в 2021 г.) // Litera. 2021. No. 8. DOI: 10.25136/2409-8698.2021.8.36334
12. *Лотман Ю.М.* Семиотика культуры и понятие текста // Ученые записки Тарт. гос. ун-та. 1981. № 515. С. 3–7.
13. *Searle J.R.* Speech Acts: an Essay in the Philosophy of Language. Lond.: Cambridge University Press, 1969. P. 1–220.
14. YouTube-канал A&E, рубрика видео «LIVE PD». — URL: https://www.youtube.com/@AETV
15. *Madzlan N.A., Han J., Bonin F., Campbell N.* Automatic recognition of attitudes in video blogs — prosodic and visual feature analysis // INTERSPEECH-2014. Eds. H. Li, P. Ching. ISCA, 2014. P. 1826–1830. — URL: https://www.isca-speech.org/archive/interspeech_2014/i14_1826.html
16. *Madzlan N.A., Han J., Bonin F., Campbell N.* Towards automatic recognition of attitudes: Prosodic analysis of video blogs // Proceedings 7th International Conference on Speech Prosody 2014 / Eds. N. Campbell, D. Gibbon, D. Hirst. ISCA, 2014. P. 91–94. DOI: 10.21437/SpeechProsody.2014-6
17. *Рафикова А.С., Валуева Е.А., Панфилова А.С.* Голос и психологические свойства человека: обзор современных исследований // Психология Журнал Высшей школы экономики. 2022. Т. 19. № 1. С. 195–215.
18. *Ramsay R.W.* Speech patterns and personality // Language and Speech. 1968. Vol. 11. No. 1. P. 54–63. DOI: 10.1177/002383096801100108
19. *Eisenberg P., Zalowitz E.* Judging expressive movement: III. Judgments of dominance-feeling from phonograph records of voice // Journal of Applied Psychology. 1938. Vol. 22. No. 6. P. 620–631. DOI: 10.1037/h0059457
20. *Stagner R.* Judgments of voice and personality // Journal of Educational Psychology. 1936. Vol. 27. No. 4. P. 272–277. DOI: 10.1037/h0057086
21. *Taylor H.C.* Social agreement on personality traits as judged from speech // Journal of Social Psychology. 1934. Vol. 5. No. 2. P. 244–248. DOI: 10.1080/00224545.1934.9919452
22. *Truesdale D.M., Pell M.D.* The sound of passion and indifference // Speech Communication. 2018. Vol. 99. P. 124–134. DOI: 10.1016/j.specom.2018.03.007
23. *Jones A., Bennett R., Cross S.* Keepin' it real? Life, death, and holograms on the live music stage // The Digital Evolution of Live Music / Eds. Angela Cresswell Jones, Rebecca Jane Bennett. Kingston Upon Hull, UK: Chandos Publishing, 2015. P. 123–138. ISBN 9780081000670. DOI: 10.1016/B978-0-08-100067-0.00010-5
24. *Рудин Л.Б.* Основы голосоведения. М.: Граница, 2009. 104 с. — URL: https://voiceacademy.ru/images/files/nauch_trudi/osnovy-golosovedenija-2009.pdf
25. *Емельянов В.В.* Развитие голоса. Координация и тренинг: учебное пособие. 12-е изд., стер. СПб.: Лань; Планета музыки, 2023. 168 с. — URL: https://reader.lanbook.com/book/316097#4
26. *Busquet F., Efthymiou F., Hildebrand C.* Voice analytics in the wild: Validity and predictive accuracy of common audio-recording devices // Behaviour Research. 2024. Vol. 56. P. 2114–2134. DOI: 10.3758/s13428-023-02139-9
27. *Brown A., Davis K.* The importance of using repetitive speech samples in prosody research // Journal of Speech Sciences. 2018. Vol. 3. No. 27. P. 245–260.

28. *Кибрик А.Е.* Методика полевых исследований (к постановке проблемы). М.: Изд-во Моск. ун-та, 1972. 181 с.
29. *Lee H., Park S.* Analyzing prosodic elements in speech demands using repetitive speech fragments // Proceedings of the International Conference on Speech and Language Processing. ISCA, 2019. P. 89–102.
30. *Johnson R., Adams M.* The use of regulated speech samples in studying prosody of obligation // Law Enforcement Quarterly. 2020. Vol. 4. No. 15. P. 312–325.
31. *Garcia L., Rodriguez E.* Intonational modalities in specific types of utterances // Journal of Phonetics and Intonation. 2017. Vol. 1. No. 19. P. 56–70.
32. *Cho T., Ladefoged P.* Variation and universals in VOT: evidence from 18 languages // Journal of Phonetics. 1999. Vol. 2. No. 27. P. 207–229. DOI: 10.1006/jpho.1999.0094
33. *Heldner M.* On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish // Journal of Phonetics. 2003. No. 31. P. 39–62. DOI: 10.1016/S0095-4470(02)00071-2
34. *Scherer K.R., Banse R., Wallbott H.G.* Emotion inferences from vocal expression correlate across languages and cultures // Journal of Cross-Cultural Psychology. 2001. Vol. 1. No. 32. P. 76-92. DOI: 10.1177/0022022101032001009
35. *Blicher D.L., Diehl R.L., Cohen L.B.* Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement // J. Phonetics. 1990. No. 18. P. 37–49. DOI: 10.1016/S0095-4470(19)30357-2
36. *Cutler A., Foss D.J.* On the role of sentence stress in sentence processing // Language and Speech. 1977. Vol. 1. No. 20. P. 1–10. DOI: 10.1177/002383097702000101
37. *Elif Bozkurt, Yücel Yemez, Engin Erzin.* Multimodal analysis of speech and arm motion for prosody-driven synthesis of beat gestures // Speech Communication. 2016. No. 85. P. 29–42. DOI: 10.1016/j.specom.2016.10.004
38. *Chen M., Mao S., Liu Y.* Big data: A survey // Mobile Networks and Applications. 2014. Vol. 2. No. 19. P. 171–209. DOI: 10.1007/s11036-013-0489-0

REFERENCES

1. Language Models are Unsupervised Multitask Learners / A. Radford, J. Wu, R. Child et al. OpenAI. 2019. P. 1–24.
2. *Jurafsky D., Martin J.H.* Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition/ Ed. 3. Lond.: Pearson, 2019. P. 1–16.
3. *Potapov V.V., Kazak E.A.* Rechevaya kommunikaciya v setevyx strukturax: mezhdu globalnym i lokalnym: sb. nauch. trudov. M.: RAN. INION; Otdel yazykoznaniya, 2022. 280 p.
4. *Chubikov A.V.* Issledovanie metodov dlya povysheniya kachestva ozvuchivaniya teksta v sistemax rechevogo sinteza // Vestnik Rossijskogo universiteta druzhby narodov. 2018. No. 4(20). P. 551–558.
5. *Huang L., Yu D.* Linguistic Features for Speech Synthesis: A Review // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2011. Vol. 19. No. 7. P. 1723–1734.
6. *Xu Y., Jiang X.* Voice Synthesis Based on Timbral Characteristics // International Conference on Electrical and Control Engineering. Melaka, Malaysia, 2019. P. 124–128.
7. *Tao X., Li S.* Emotional Speech Synthesis Based on Prosodic Attributes // Proceedings IIEEE International Conference on Acoustics, Speech and Signal Processing. New Orleans, LA, USA, 2017. P. 511–515.
8. *Chung D., Kim H., Ahn S.* An integrated study of user acceptance and resistance on voice commerce // International Journal of Innovation and Technology Management. 2022. Vol. 19. No. 7.
9. *Singh R., Jiménez A., Øland A.* Voice disguise by mimicry: deriving statistical articulometric evidence to evaluate claimed impersonation // IET Biometrics. 2017. Vol. 6. No. 4. P. 282–289.
10. *Matveyev A.G., Martyanova E.Yu.* Grazhdansko-pravovaya okhrana golosa cheloveka pri yego sinteze i posleduyushchem ispolzovanii // Ex jure. 2023. No. 3. S. 118–131. DOI: 10.17072/2619-0648-2023-3-118-131
11. *Anikyan T.S.* Ekspressivnyj potencial prosodii v inauguracionnoj ritorike (na materiale obrashheniya Dzhozefa Bajdena v 2021 g.) // Litera. 2021. No. 8. DOI: 10.25136/2409-8698.2021.8.36334
12. *Lotman Yu.M.* Semiotika kultury i ponyatie teksta // Uchen. zapisky Tart. gos. un-ta. 1981. No. 515. S. 3–7.
13. *Searle J.R.* Speech Acts. An Essay in the Philosophy of Language // Cambridge University Press. 1969. P. 1–220.
14. YouTube-kanal A&E, rubrika video "LIVE PD". — URL: https://www.youtube.com/@AETV
15. *Madzlan N.A., Han J., Bonin F., Campbell N.* Automatic recognition of attitudes in video blogs — prosodic and visual feature analysis // INTERSPEECH-2014. Eds. H. Li, P. Ching. ISCA, 2014. P. 1826–1830. — URL: https://www.isca-speech.org/archive/interspeech_2014/i14_1826.html
16. *Madzlan N.A., Han J., Bonin F., Campbell N.* Towards automatic recognition of attitudes: Prosodic analysis of video blogs // Proceedings 7th International Conference on Speech Prosody 2014 / Eds.

N. Campbell, D. Gibbon, D. Hirst. ISCA, 2014. P. 91–94. DOI: 10.21437/SpeechProsody.2014-6
17. *Rafikova A.S., Valuyeva E. A., Panfilova A.S.* Golos i psikhologicheskiye svoystva cheloveka: obzor sovremennykh issledovaniy // Journal of the Higher School of Economics. Psychology. 2022. T. 19. No. 1. S. 195–215.
18. *Ramsay R.W.* Speech patterns and personality // Language and Speech. 1968. Vol. 11. No. 1. P. 54–63. DOI: 10.1177/002383096801100108.
19. *Eisenberg P., Zalowitz E.* Judging expressive movement: III. Judgments of dominance-feeling from phonograph records of voice // Journal of Applied Psychology. 1938. Vol. 22. No. 6. P. 620–631. DOI: 10.1037/h0059457
20. *Stagner R.* Judgments of voice and personality // Journal of Educational Psychology. 1936. Vol. 27. No. 4. P. 272–277. DOI: 10.1037/h0057086
21. *Taylor H.C.* Social agreement on personality traits as judged from speech // Journal of Social Psychology. 1934. Vol. 5. No. 2. P. 244–248. DOI: 10.1080/00224545.1934.9919452
22. *Truesdale D.M., Pell M.D.* The sound of passion and indifference // Speech Communication. 2018. Vol. 99. P. 124–134. DOI: 10.1016/j.specom.2018.03.007
23. *Jones A., Bennett R., Cross S.* Keepin' it real? Life, death, and holograms on the live music stage // The Digital Evolution of Live Music / Eds. Angela Cresswell Jones, Rebecca Jane Bennett. Kingston Upon Hull, UK: Chandos Publishing, 2015. P. 123–138. ISBN 9780081000670. DOI: 10.1016/B978-0-08-100067-0.00010-5
24. *Rudin L.B.* Osnovy golosovedeniya. M.: Granitsa, 2009. 104 s. — URL: https://voiceacademy.ru/images/files/nauch_trudi/osnovy-golosovedenija-2009.pdf
25. *Emelyanov V.V.* Razvitiye golosa. Koordinatsiya i trening: uchebnoye posobiye. 12-e izd., ster. SPb.: Lan; Planeta muzyki, 2023. 168 s. — URL: https://reader.lanbook.com/book/316097#4
26. *Busquet F., Efthymiou F., Hildebrand C.* Voice analytics in the wild: Validity and predictive accuracy of common audio-recording devices // Behaviour Research. 2024. Vol. 56. P. 2114–2134. DOI: 10.3758/s13428-023-02139-9
27. *Brown A., Davis K.* The importance of using repetitive speech samples in prosody research // Journal of Speech Sciences. 2018. Vol. 3. No. 27. P. 245–260.
28. *Kibrik A.E.* Metodika polevyx issledovanij (k postanovke problemy). M.: Izd-vo Mosk. un-t-a, 1972. 181 p.
29. *Lee H., Park S.* Analyzing prosodic elements in speech demands using repetitive speech fragments // Proceedings of the International Conference on Speech and Language Processing. ISCA, 2019. P. 89–102.
30. *Johnson R., Adams M.* The use of regulated speech samples in studying prosody of obligation // Law Enforcement Quarterly. 2020. Vol. 4. No. 15. P. 312–325.
31. *Garcia L., Rodriguez E.* Intonational modalities in specific types of utterances // Journal of Phonetics and Intonation. 2017. Vol. 1. No. 19. P. 56–70.
32. *Cho T., Ladefoged P.* Variation and universals in VOT: evidence from 18 languages // Journal of Phonetics. 1999. Vol. 2. No. 27. P. 207–229. DOI: 10.1006/jpho.1999.0094
33. *Heldner M.* On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish // Journal of Phonetics. 2003. No. 31. P. 39–62. DOI: 10.1016/S0095-4470(02)00071-2
34. *Scherer K.R., Banse R., Wallbott H.G.* Emotion inferences from vocal expression correlate across languages and cultures // Journal of Cross-Cultural Psychology. 2001. Vol. 1. No. 32. P. 76-92. DOI: 10.1177/0022022101032001009
35. *Blicher D.L., Diehl R.L., Cohen L.B.* Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement // J. Phonetics. 1990. No. 18. P. 37–49. DOI: 10.1016/S0095-4470(19)30357-2
36. *Cutler A., Foss D.J.* On the role of sentence stress in sentence processing // Language and Speech. 1977. Vol. 1. No. 20. P. 1–10. DOI: 10.1177/002383097702000101
37. *Elif Bozkurt, Yücel Yemez, Engin Erzin.* Multimodal analysis of speech and arm motion for prosody-driven synthesis of beat gestures // Speech Communication. 2016. No. 85. P. 29–42. DOI: 10.1016/j.specom.2016.10.004
38. *Chen M., Mao S., Liu Y.* Big data: A survey // Mobile Networks and Applications. 2014. Vol. 2. No. 19. P. 171–209. DOI: 10.1007/s11036-013-0489-0