DOI: 10.15514/ISPRAS-2026-38(6)-38

Принципы построения системы просодической аннотации на основе параметрических данных

О.В. Гончарова, ORCID: 0000-0003-1044-6244 <goncharova-oxvl@rudn.ru>

Институт системного программирования РАН,
Россия, 109004, г. Москва, ул. А. Солженицына, д. 25.
Российский университет дружбы народов им. П. Лумумбы,
Россия, 117198, г. Москва, ул. Миклухо-Маклая, 6.

Аннотация. В статье рассматривается проблема разработки воспроизводимой системы просодической аннотации, способной адекватно отражать как лингвистически значимые категории, так и градиентную вариативность речевого сигнала. На основе анализа существующих категориальных и непрерывных подходов, выявлены их ограничения: субъективность, неполный охват просодических параметров, слабая сопоставимость данных от разных дикторов. В качестве решения некоторых ограничений предлагается методология, основанная на комплексной параметризации речи. Система формализована, автоматизируема и обеспечивает сопоставимость результатов между разными говорящими и речевыми ситуациями без потери абсолютных характеристик сигнала.

Ключевые слова: просодическая аннотация; русский язык; интонация; нормализация; частота основного тона; интенсивность; длительность; слоговая сегментация; пороговая категоризация; сопоставительный анализ; речевой сигнал.

Для цитирования: Гончарова О.В. Принципы построения системы просодической аннотации на основе параметрических данных. Труды ИСП РАН, 2026, том 38, вып. 6, часть 2, стр. 285–302. DOI: 10.15514/ISPRAS-2026-38(6)-38.

Principles of Building a Prosodic Annotation System Based on Parametric Data

O.V. Goncharova ORCID: 0000-0003-1044-6244 <goncharova_oxvl@pfur.ru>

Institute for System Programming of the Russian Academy of Sciences,
25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.
Patrice Lumumba Peoples' Friendship University of Russia,
6, Miklukho-Maklaya st., Moscow 117198, Russia.

Abstract. The paper addresses the development of a reproducible prosodic annotation system that can capture both linguistically relevant categories and the gradient variability of the speech signal. A review of existing categorical and continuous approaches identifies recurring limitations, including annotator subjectivity, incomplete coverage of prosodic parameters, and limited comparability across speakers. To mitigate these issues, the study proposes a methodology grounded in comprehensive parametric modeling of speech. Its core component is a normalization procedure for the primary acoustic correlates of prosody-fundamental frequency (F0), intensity, and duration-mapped onto a unified speaker-relative scale (0–100) and subsequently discretized via thresholding into perceptually motivated levels. The proposed framework implements a two-layer representation: a lower layer comprising numerical measurements and derived dynamic metrics, and an upper layer represented by sequences of discrete level labels. The system is formalized and amenable to automation, enabling cross-speaker and cross-situational comparability while preserving the signal's absolute characteristics. The paper details the methodological principles and implementation workflow, provides examples of output tables and graphical visualizations, and discusses the remaining limitations of the approach.

Keywords: prosodic transcription; Russian; intonation; normalization; fundamental frequency (F0); intensity; duration; syllable segmentation; threshold-based categorization; comparative analysis; speech signal.

For citation: Goncharova O.V. Principles of Building a Prosodic Annotation System Based on Parametric Data. Trudy ISP RAN/Proc. ISP RAS, 2026, vol. 38, issue 6, part 2, pp. 285-302 (in Russian). DOI: 10.15514/ISPRAS-2026-38(6)-38.

1. Введение

Одной из важнейших задач в области анализа речи является получение разметки, которая, с одной стороны, обладает необходимой степенью детализации для лингвистической интерпретации, а с другой – обеспечивает сопоставимость результатов. В исследованиях отмечается, что ограничения надёжности разметки обусловлены главным образом вариативностью трёх типов:

(i) изменчивостью распределения просодических элементов и структур в зависимости от грамматического и дискурсивного контекста;

(ii) вариативностью набора акустических коррелятов, используемых для реализации данных элементов;

(iii) колебаниями численных значений самих коррелятов даже в пределах одной и той же функционально определяемой единицы [1-2].

Вследствие этого специалисты, выполняющие транскрипцию, часто оказываются в ситуации, когда акустические основания для принятия решения неоднозначны, требуется обращение к более широкому контексту и рассмотрение альтернативных анализов, а общие правила разрешения неоднозначности с трудом поддаются формализации.

В этих условиях значение приобретает способ представления просодии в транскрипции, и какая именно информация должна быть представлена в разметке: набор решений зависит от исследовательской цели (описание свойств просодической системы, изучение фонетической реализации и вариативности просодических категорий, формирование размеченного корпуса, сопоставление диалектов/языков, задачи распознавания и синтеза речи и тому подобное). В связи с этим в практиках аннотирования встречаются транскрипции,

ориентированные на фонологический уровень, на фонетический уровень или на промежуточные, или категориальные описания, а также форматы, предполагающие совместное хранение символических меток и измеримых характеристик сигнала [2].

В рамках настоящей работы под просодической транскрипцией понимается не только символическая нотация, но и формализованный способ репрезентации просодической организации высказывания, включающий:

(i) выбор единицы привязки измерений;

(ii) процедуру извлечения акустических коррелятов и их привязки к позициям в профиле;

(iii) правила преобразования параметров в сопоставимый вид,

(iv) правила получения дискретного описания как производного от числовых данных.

Таким образом, предлагаемый подход ориентирован прежде всего на стандартизацию процедуры и формата данных, а не на введение нового инвентаря фонологических единиц интонации.

В связи с этим важно развести уровень транскрипции и уровень интерпретации. На уровне транскрипции фиксируются измеримые характеристики сигнала (F0, интенсивность, длительность) и их нормированные представления, а также дискретные уровневые метки, получаемые по заранее заданным правилам категоризации. На уровне интерпретации решается вопрос о том, как конкретные конфигурации числового профиля и уровневых последовательностей соотносятся с фонологическими категориями, функциями и значениями (в частности, с типами акцентов/границ, коммуникативными типами, дискурсивными и прагматическими параметрами), и какие из различий устойчивы для данной системы/языка и подтверждаются независимыми критериями (перцептивными, распределительными, функциональными). В настоящей работе интерпретационный уровень рассматривается как последующий этап: предлагается инструментальное основание, обеспечивающее воспроизводимое получение сопоставимых профилей и их дискретных представлений, которые далее могут быть использованы для проверки фонологических гипотез.

Исходя из обозначенных границ, работа отвечает на следующие вопросы:

(1) какие ограничения существующих категориальных и непрерывных систем препятствуют воспроизводимости и сопоставимости данных;

(2) каким образом возможно задать единую процедуру нормализации основных акустических коррелятов просодии в шкалу сопоставимости при сохранении исходных физических значений;

(3) как сформулировать правила пороговой категоризации числовых параметров, чтобы дискретные уровни были выводимыми из измерений и могли использоваться в дальнейшем для сопоставления и интерпретации;

(4) каким образом организовать двухуровневое представление (числовой слой + дискретный слой) так, чтобы оно было пригодно для корпусной аннотации и статистического анализа, включая описание не только уровней, но и динамических характеристик.

Для ответа на поставленные вопросы требуется сопоставить существующие практики транскрипции по тому, какая информация о просодии фиксируется и в каком виде она представляется, поскольку именно формат репрезентации во многом определяет воспроизводимость процедур, степень интерпретируемости результата и возможности сопоставления. С точки зрения формата репрезентации методы просодической транскрипции целесообразно разделять на два основных типа:

(1) категориальные, представляющие просодию через фиксированный инвентарь меток, которые соотносятся с предполагаемыми просодическими элементами и структурой: маркируются события и границы, вводятся уровни организации и отношения между ними. Данный тип ориентирован на то, чтобы зафиксировать какой просодический элемент реализован и какую роль он выполняет в организации высказывания, то есть обеспечить лингвистически интерпретируемое представление. При этом категоризация опирается на восприятие и на совокупность признаков, включая F0, длительность, интенсивность и другие характеристики, а также учёт контекста;

(2) непрерывные, описывающие просодию через численные параметры и их траектории: измеряются значения и динамика акустических переменных (прежде всего F0, но также интенсивность, длительность и связанные показатели), после чего данные используются для статистического анализа, моделирования и автоматизированной обработки. Такой формат ориентирован на то, чтобы учесть какие параметры сигнала реализуют просодическую организацию и в каких количественных значениях это проявляется. Для непрерывных подходов важно, как выполнены сегментация и привязка измерений к единицам анализа (слогу, слову, фразе): от этого зависит, какие участки сигнала сравниваются и как интерпретируются полученные значения.

2. Категориальные просодические транскрипции

Разработанные в британской фонетической традиции первой половины XX века системы интонационной разметки, такие как модели М.А.К. Халлидея или Дж.Д.О'Коннора и Г.Ф.Арнольда, основывались на слуховом анализе и оперировали ограниченным набором тонов, каждому из которых приписывалось определённое значение [3]. Однако, как отмечал Д. Кристал, анализируя систему М.А.К. Халлидея, выбор конкретного тона часто невозможен без понимания коммуникативного намерения говорящего [4]. Таким образом, интерпретация в рамках данных систем не была автономной и для определения тона требовалось учитывать контекст и смысл высказывания, что существенно снижало объективность транскрипции [4]. Еще одним ограничением выступала проблема надёжности. Поскольку разметка целиком зависела от слухового восприятия, уровень согласованности между транскрипциями, выполненными разными специалистами, часто оказывался низким [5]. Кроме того, инвентарь тонов в данных системах был ограниченным и не отражал всего разнообразия реальной интонационной вариативности. К примеру, описания опирались на несколько основных ядерных тонов, тогда как в реальной речи существует множество сложных, комбинированных и промежуточных мелодических контуров. Совокупность данных ограничений и стимулировала в дальнейшем поиск более формализованных подходов к описанию просодии [6].

В русской фонетической традиции одним из ранних шагов в этом направлении стала система интонационной транскрипции С.В.Кодзасова [7]. В рамках данной модели интонационная организация высказывания описывается как комбинация акцентных тональных движений (восходящих, нисходящих и др.), приуроченных к ударным слогам, и специальных контуров, маркирующих границы фраз и их интонационное членение. Несмотря на то, что подход ориентирован на наблюдаемые особенности русского интонационного строя и фиксирует как локальные акцентные изменения тона, так и пограничные характеристики фразовой структуры, применение данной системы выявляет ряд ограничений. В частности, транскрипция отличается высокой сложностью, используя широкий набор специальных символов и диакритических знаков для обозначения различных типов тональных движений, параметров акцентной позиции и паузации. Также модель предполагает тесную связь с семантикой, что означает, что при разметке исследователь фактически интерпретирует, какое значение (вопрос, утверждение, неуверенность, завершённость и т.д.) выражается данным просодическим сочетанием, и исходя из этого выбирает комбинацию признаков. Такой

подход усиливает элемент субъективности – транскрипция зависит от понимания смысла исследователем. Следует также учитывать, что система разрабатывалась с опорой на интонационные конструкции русского языка и не была изначально ориентирована на межъязыковую переносимость и при попытке применения к другим языкам неизбежно возникнет проблема соответствия инвентаря тональных единиц и правил их сочетания типологически иным просодическим системам.

В этом отношении данный подход отличается от англоязычной традиции, где форматы типа “тоны и индексы разрывов” (Tones and Break Indices, ToBI) предполагают адаптацию к различным языкам. Система ToBI, изначально созданная для английского языка в рамках автосегментно-метрической модели, впоследствии была адаптирована для значительного числа других языков [5]. В основе метода лежит представление интонации в виде последовательности дискретных тоновых событий: тональных акцентов на ударных слогах, обозначаемых символами H (высокий тон) и L (низкий тон) и их комбинациями, а также граничных тонов в конце интонационных фраз, таких как L% или H%. Также применяется шкала разрывов (Break Indices), позволяющая фиксировать границы синтагм разной силы в диапазоне 0–4.

Основным достоинством ToBI является стандартизованный фонологический подход, направленный на фиксацию категориальных различий в интонационных контурах. Тем не менее, опыт практического применения данной системы выявил ряд ограничений. Одна из проблем состоит в том, что система регистрирует только категориальные различия, не отражая вариативность в реализации интонационных параметров. Например, два контура с одинаковой пометкой нисходящего тона H+L% могут отличаться по характеристикам, таким как диапазон или скорость падения частоты основного тона, но различия не фиксируются в разметке, поскольку считаются фонетически нерелевантными [5], что ограничивает возможности анализа, например, эмоционально-экспрессивных реализаций или диалектных и региональных различий, где значимыми оказываются параметры, выходящие за пределы категориальной классификации [8]. Еще одна проблема - воспроизводимость разметки. Уровень согласия между экспертами при определении типа тонального акцента часто остаётся невысоким, трудности возникают при различении смежных категорий, таких как H и L+H, или при анализе нетипичных или промежуточных контуров [5]. Проблемы переносимости также остаются актуальными. Несмотря на то, что ToBI описывается как общий формат разметки, на практике для каждого языка создавались собственные версии с частично различающимся инвентарём меток и правилами их применения. В результате не сформировался единый набор обозначений, сопоставимый по степени унификации с международными фонетическими стандартами, что приводит к тому, что сходные по описанию контуры в разных языках могут получать разные метки, тогда как одинаковые обозначения могут соотноситься с неодинаковыми по форме реализациями, что затрудняет сопоставление и обобщение [5]. Также представляется важным, что ToBI описывает преимущественно изменения частоты основного тона и фразовое членение. Другие просодические параметры, такие как темп, точная длительность пауз, динамика интенсивности или характеристики фонации, в стандартной нотации не отражаются [5], в то время как для исследований, требующих учёта полного комплекса просодических характеристик, этого может быть недостаточно.

Одним из подходов, направленных на преодоление ограничений системы ToBI, стала система “ритм и высота” (Rhythm and Pitch, RaP). Так, система RaP вводит отдельный ритмический уровень [5], что позволяет отмечать в транскрипции не только чередование ударных и безударных слогов, но и различную степень акцентного выделения, достигаемую за счёт изменения длительности или интенсивности, а не высоты тона. Кроме того, RaP использует перцептивный подход: разметка строится прежде всего на событиях, выделяемых на слух, тогда как анализ контура F0 используется только как дополнительное средство. Система была разработана с целью устранения недостатков ToBI, такие как бинарная шкала акцентов и

проставление границ без слышимой паузы [5], но следует отметить, что RaP не стала столь же распространённым стандартом, так как она сложнее в реализации и вводит больше уровней меток. Тем не менее, появление RaP демонстрирует осознание важности учета ритма и разных степеней просодического выделения для полноценной транскрипции.

Существует и ряд других систем, например, система IViE (Intonational Variation in English) разрабатывалась для сравнения диалектных вариаций английской интонации и сочетала фонологический уровень с дополнительным фонетическим (например, сужение диапазона) [9]. Подобные проекты часто пытаются преодолеть ограничения базовых моделей, вводя дополнительные уровни аннотации или комбинируя подходы.

3. Непрерывные просодические транскрипции

Другой подход предложили Д. Хёрст и А. Ди Кристо, разработав систему INTSINT (INternational Transcription System for INTonation) [10-11]. Эта система представляет интонацию как последовательность символов, обозначающих относительные изменения высоты тона (например, U – повышение тона, D – понижение, H – высокий уровень, L – низкий и так далее). Чтобы записать интонацию высказывания методом INTSINT, нужно разбить ее на значимые тональные точки (обычно это точки начала, конца фразы, ядерных движений и другие подобные) и проставить символы INTSINT, соответствующие относительному уровню каждой точки, и знаки выравнивания (alignment), показывающие, где эти точки располагаются.

Таким образом, можно указывать, к какой части слога (или другой фонологической единицы) привязан данный тон. Цепочка тоновых символов с привязкой к определенным слогам дает однозначную интерпретацию интонационного контура. Например, предложение «It’s time to go», прочтенное с нейтральной интонацией, может быть транскрибировано как M:/ɪts/ T:/taɪm tə/ D< B]/ɡoʊ/, что означает: начальный средний тон (M) на слове «It’s», затем Top (самый высокий) на середине фразы «time to», затем Downstepped (сниженный относительно предыдущей точки) на ранней части слога «go», и Bottom (самый низкий) на конце этого же слога «go». Если изменить символы – можно записать интонацию вопроса или восклицания для той же фразы.

Главное преимущество INTSINT – применимость для любого языка, по замыслу авторов система подходит для [10-11]. Однако особенность INTSINT заключается в том, что в качестве базового параметра рассматривается исключительно изменение высоты тона [10-11], оставляя вне транскрипции интенсивность и длительность. Соответственно, не отображаются и связанные с ними явления (например, изменение темпа или динамика интенсивности), напрямую не выражаемые в тональной структуре. Ещё одна трудность – символы INTSINT лишены прямой лингвистической интерпретации. В отличие от систем типа ToBI, где метки привязаны к фонологическим категориям, INTSINT фиксирует только относительный рисунок тона, что делает систему полезной для акустических исследований, но менее удобной для описания функциональной нагрузки интонации (например, различия вопроса и утверждения). С другой стороны, это же является достоинством при типологическом сравнении: INTSINT позволяет напрямую сопоставлять мелодические профили высказываний разных языков в унифицированном формате. Например, Д. Хёрсти и А. ДиКристо показали, что во французском языке, в основном, отмечается повышение к концу фразы (много символов H и U в завершении), в английском больше падений (символы D, L) [10-11]. Таким образом, INTSINT решает проблему универсальности снижая уровень лингвистической интерпретации и исключая прочие просодические параметры.

Еще одной системой, описывающей интонацию не через дискретные классы, а через непрерывные параметры является система Tilt [12]. Её исходная установка состоит в том, что интонационная организация может быть представлена как последовательность интонационных событий, для описания которых используются параметры, задающие как

величину и длительность движения, так и его конфигурацию и позиционирование во времени: Amplitude – амплитуда движения тона; Duration – длительность события; Tilt – параметр формы, показывающий баланс между подъёмом и падением тона; Time position – временная позиция события относительно начала высказывания или опорной точки (что позволяет фиксировать раннюю или позднюю реализацию); F0 position – уровень основной частоты (F0) в середине события [12]. Вместе с тем, в литературе отмечается ряд системных недостатков, сужающих применимость Tilt, в частности, локальный подход, ограниченный анализом ударных и конечных слогов, не учитывает глобальную интонационную структуру [13], описание интервалов между событиями остаётся недостаточным [14], невозможность описывать некоторые конфигурации F0 (например, падение с последующим подъёмом в пределах слога) [15] и др.

В целом анализ показывает, что общепринятой системы просодической транскрипции нет: разные подходы решают разные задачи и предполагают неизбежные компромиссы. К ним относятся:

(1) недостаточная полнота охвата просодии – многие системы фокусируются на тональном компоненте, оставляя без явного описания интенсивность, темп, ритм и тембр [10], что упрощает запись, но снижает информативность для коммуникативных параметров (например, эмоциональности или степени уверенности);

(2) ограниченный или, наоборот, чрезмерно расширенный инвентарь категорий – стремление к минимализму (например, ограниченный базовый алфавит H/L и специфические инвентари тональных акцентов/граничных тонов в ToBI-адаптациях) приводит к укрупнению реального разнообразия, из-за чего разные реализации могут обозначаться одинаково [5], тогда как слишком детализированные системы усложняют разметку, ухудшая воспроизводимость;

(3) проблемы надёжности и согласованности разметки – интонационная разметка во многом опирается на восприятие, поэтому оценки разных экспертов нередко расходятся в выделении тонов и границ (например, категория Break Index 2 в ToBI, куда попадают неоднозначные случаи пауз [5]);

(4) теоретическая и языковая несопоставимость - системы опираются на разные фонологические основания, поэтому обозначения, принятые, например, в британской традиции, не всегда могут быть однозначно преобразованы в нотацию ToBI или в систему С. В. Кодзасова;

(5) вариативность часто носит градиентный характер (например, по диапазону/масштабу и др.) [16], при этом нотации типа ToBI подразумевают интерпретацию различий как категориальных противопоставлений, что затрудняет построение адекватной теории соответствия «форма → значение».

4. Методологические основы предлагаемой системы

На основе анализа существующих систем просодической транскрипции и методов нормализации в предлагаемом подходе используются базовые принципы существующих нотаций, а также вводятся дополнительные, призванные устранить выявленные ограничения. Теоретико-методическая основа транскрипции формируется в соответствии с:

(1) фонологическим подходом, исходящим из отсутствия прямого инвариантного соответствия между акустическим сигналом и коммуникативной функцией и требующий введения абстрактных единиц, соотносимых с множеством функций и допускающих вариативную реализацию в контексте;

(2) принятием интонационной фразы в качестве базовой единицы сегментации;

(3) сохранением принципа привязки интонационных событий к ударным слогам.

Также целесообразным представляется сформулировать набор требований к описанию, которое одновременно является

(a) процедурой получения сопоставимых измерений и

(b) нотацией, пригодной для интерпретации и статистического анализа.

Принципами, необходимыми для получения сопоставимых измерений, содержательной интерпретации и статистического анализа являются:

1) воспроизводимость и минимизация интерпретационной вариативности. Нотация должна опираться на формализованные правила извлечения и фиксации параметров, чтобы уменьшить зависимость результата от экспертной интерпретации и повысить согласованность разметки между специалистами;

2) сопоставимость дикторов при сохранении абсолютных характеристик. Нормализация должна обеспечивать сравнимость данных при сохранении информации об абсолютном уровне и диапазоне (высотном регистре, динамическом диапазоне, темпоральных ресурсах). В противном случае сравнение сводится к форме относительного контура ценой потери лингвистически и коммуникативно значимых различий;

3) единая процедура для комплекса параметров (не только для F0). Поскольку многие нотации преимущественно фиксируют F0, оставляя без внимания интенсивность, длительность, темп и другие компоненты, требуется подход, в котором методы измерения и приведения к единой форме применяются одинаково к нескольким просодическим переменным;

4) привязка измерений к лингвистически релевантным позициям, а не к произвольной равномерной временной сетке. Сравнение должно строиться не на основе произвольных точек, а относительно позиций, заданных лингвистической разметкой (например, сегмент, слог, фразовая позиция), что позволит сохранить фазовые соотношения и ритмическую структуру речи;

5) двухуровневое представление данных: числовое и категориальное. Описание должно допускать хранение точных числовых значений и категориальную интерпретацию на основе относительной шкалы, что даст возможность работать с одними и теми же данными как для акустического анализа, так и для качественного описания, удобного для восприятия и сравнения;

6) однозначные правила категоризации. Категориальная запись должна строиться по явно заданным порогам и фиксированному набору меток (уровни, диапазон, регистр и т. п.), чтобы присвоение категорий было однозначным и проверяемым. Таким образом снимается проблема неочевидных правил, которые обычно повышают субъективность разметки и затрудняют использование методики в других исследованиях;

7) устойчивость к аномальным данным. Процедура приведения к единой форме должна минимизировать влияние отклонений в данных и учитывать, что важную информацию могут нести не только сами значения, но и характер их распределения (например, асимметрия, концентрация), а также динамические изменения. Без этого сравнение может быть искажено техническими погрешностями или окажется ненадёжным при сопоставлении разных говорящих;

8) выходные данные, непосредственно пригодные для статистического сопоставления. Результат описания должен иметь формат, позволяющий выполнять количественные сравнения: таблицы, производные характеристики (интервалы, скорости), а также категориальные признаки (уровни/регистр/диапазон), что обеспечивает переход от разметки как описания контура к воспроизводимой модели сопоставления.

5. Реализация предлагаемого подхода

Предлагаемая схема основана на понимании просодической транскрипции как формализованной записи, в которой числовые измерения и их категориальная интерпретация задаются совместно и согласованно. Транскрипция в данном случае представляет собой комплексное представление данных, которое одновременно (1) обеспечивает дальнейшую категоризацию и соотнесение результатов с фонологическими метками и позициями интонационной структуры и (2) сохраняет количественные характеристики, необходимые для акустического анализа (уровень высоты тона и интенсивности, длительность, показатели динамики, форма контуров). Важным является то, что сопоставление между дикторами и построение усреднённых траекторий выполняется в относительной шкале 0–100, вычисляемой индивидуально для каждого диктора по его собственному диапазону.

Данными для анализа служат аудиозаписи и трёхуровневая разметка в формате TextGrid, включающая сегментный и слоговой уровни, а также классификацию типов сегментов (см. Рис. 1), что позволяет сопоставлять параметры по определённым интервалам и точкам (предъядерная/ядерная/заядерная зоны, позиция ударности, слоговая модель и др.).

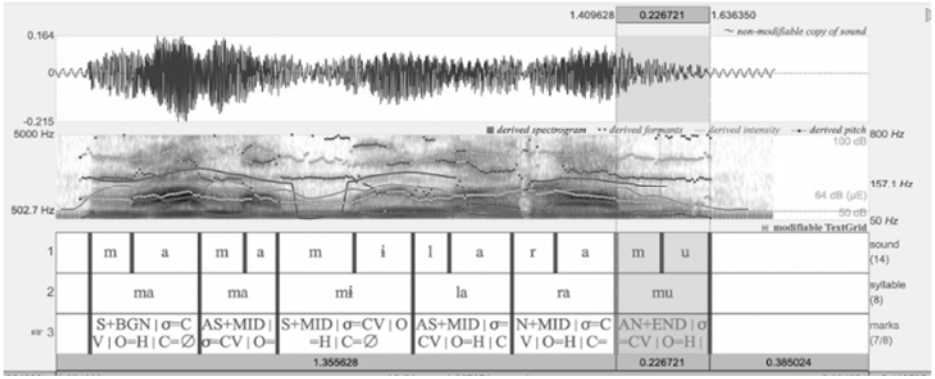


Рис. 1. Пример разметки в Praat TextGrid:
три уровня («sound», «syllable», «marks») и объединённые метки слогов.
Fig. 1. Example Praat TextGrid annotation:
three tiers ("sound", "syllable", "marks") and combined syllable labels.

Процедура реализована как автоматизированный вычислительный конвейер: ручной этап ограничивается подготовкой разметки в Praat TextGrid (сегментный и слоговой уровни, а также уровень marks), все последующие операции выполняются в среде Python. Результат обработки выдаётся в виде набора выходных таблиц и графических представлений. Минимальной единицей сопоставления принимается профиль высказывания, представляющий собой упорядоченный набор значений параметров, рассчитанных для заданной последовательности слогов; профиль формируется так, чтобы он мог быть представлен в двух формах: (а) как таблица числовых значений (исходные единицы и относительные индексы 0–100), (б) как транскрипция - последовательность категориальных меток уровней, полученных по фиксированным пороговым правилам.

Для частоты основного тона исходным преобразованием служит перевод F0 из герц в полутоны (без выравнивания по среднему значению или стандартному отклонению):

$$ST(f_k) = 12 \cdot log_{\,2}\left(\frac{f_k}{F_{ref}}\right).$$

Далее для каждого диктора задаётся индивидуальный диапазон в полутонах по нижней и верхней границам ST_{low} и ST_{high}, при этом размах диапазона $R_{ST} = ST_{high} - ST_{low}$. Любое значение ST переводится в относительную шкалу 0–100 по формуле:

$$PT(ST) = 100 \cdot \frac{ST - ST_{low}}{R_{ST}},$$

где PT интерпретируется как положение текущего значения внутри индивидуального диапазона диктора в данной позиции профиля (0 соответствует нижней границе диапазона, 100 – верхней). Для контроля масштаба преобразования дополнительно вычисляется вес полутона в индивидуальном диапазоне (в процентах шкалы 0-100) для конкретного диктора:

$$w = \frac{100}{R_{ST}}.$$

В таблице результата для каждого слога фиксируются длительность D, исходные значения F0 в герцах (Mean pitch, Max pitch, Min pitch, Start pitch, End pitch) и соответствующие им нормированные индексы (в файле они представлены в столбцах вида Mean pitch (PT), Max pitch (PT), Min pitch (PT), Start pitch (PT), End pitch (PT)), а также их уровневые интерпретации (столбцы вида Mean pitch (PT) Description и др.), что обеспечивает однозначную связь между числовым значением и его категориальной меткой.

Индивидуальные границы диапазона диктора оцениваются автоматически с помощью вероятностной модели, реализованной как классификация по дискретизированным границам. Обучение выполнялось на калибровочных данных, где для каждого диктора, помимо основной записи чтения, дополнительно фиксировались два режима на сопоставимом речевом материале: «тихо» – произнесение с минимально возможной громкостью при сохранении нормальной фонации и разборчивости (без шёпота) и «громко» - произнесение с максимально возможной громкостью при сохранении нормальной фонации и разборчивости (без крика). По калибровочным записям вычислялись целевые числовые значения нижней и верхней границ диапазона: I_L определялось по режиму «тихо», I_U – по режиму «громко».

Для того чтобы оценивать качество предсказания и получать вероятностный прогноз, значения I_L и I_U переводились в классы путём дискретизации на K интервалов. Дискретизация выполнялась по квантилям распределения целевых значений в обучающей выборке отдельно для I_L и I_U: строились границы $b_0^{(L)} < b_1^{(L)} < \cdots < b_K^{(L)}$ и $b_0^{(U)} < b_1^{(U)} < \cdots < b_K^{(U)}$, где b_0 и b_K – минимальное и максимальное наблюдаемое значение на обучении, а промежуточные границы задавались квантилями $Q_{m/K}$ (то есть $b_m = Q_{m/K}, m = 1, \dots, K-1$). После этого каждому диктору присваивались два класса: $y_L \in \{1, \dots, K\}$ – номер интервала, в который попадает I_L, и $y_U \in \{1, \dots, K\}$ – номер интервала, в который попадает I_U (в текущей реализации используется $K = 10$.

Вход модели формируется по совокупности всех записей чтения данного диктора: вычисляется вектор признаков x, включающий характеристики интенсивности и F0. Для учёта различий между группами в модель вводится групповой фактор тип диктора $g \in \{M, F, C\}$ (мужчина/женщина/ребёнок). В качестве ориентира для F0 можно использовать нормативные интервалы, полученные на материале чтения: 85-155 Гц для мужчин и 165-255 Гц для женщин [17]. Для детской речи в литературе приводятся более высокие значения, например, для чтения у 7–8-летних детей – порядка 273-297 Гц, а также оценки порядка 250-275 Гц в экспериментах на детском материале [18]. Вероятность класса j для соответствующей границы вычисляется по softmax-модели (многоклассовая логистическая регрессия) с групповыми сдвигами:

$$p\left(y^{(p,t)} = j | x, g\right) = \frac{exp\left(w_j^{(p,t)\ \top} x + b_{j,g}^{(p,t)}\right)}{\sum_{m=1}^{K} exp\left(w_m^{(p,t)\ \top} x + b_{m,g}^{(p,t)}\right)},$$

где $p \in \{F0, \int\}$ – параметр, $t \in \{L, U\}$ – нижняя/верхняя граница, x – вектор признаков диктора, $g \in$ {мужчина,женщина,ребёнок} – тип диктора, $K = 10$.

Качество модели оценивалось на отложенной выборке с разбиением по дикторам (то есть записи одного и того же диктора не попадали одновременно в обучение и тестирование); точность классификации составляет около 0.83 для предсказания классов границ диапазона.

Слоговая сегментация в предлагаемом подходе используется как единица измерения и нормирования просодических параметров, а не как утверждение о том, что каждый слог совпадает с интонационным событием. Профиль высказывания задаётся как упорядоченная последовательность слоговых позиций, для которых фиксируются измеренные значения и нормированные индексы высоты тона, интенсивности и длительности, а также производные характеристики динамики (интервал и скорость изменения F0) по начальной и конечной точкам слога. При этом известна методологическая трудность: интонационные события (тональные акценты и граничные явления) могут иметь несовпадающую с границами слога локализацию - быть внутрислоговыми, межслоговыми (смещение пика/дна на соседний слог) или распределёнными на несколько слогов. Чтобы исключить зависимость транскрипции от предположения «событие = слог», в схеме может быть рассчитан событийный слой, производный от слогового профиля: (i) слой акцентных событий и (ii) слой граничных событий:

(1) Акцентные события (A). Кандидаты в акцентные позиции задаются на основе слоговых меток разметки (позиция в предъядерной/ядерной/заядерной зоне и тип ударности включаются в состав слоговой метки), а также на основе акустических признаков выделенности. Для каждого слога вычисляется набор признаков, уже присутствующих в профиле: уровни F0 (например, Max pitch (PT), Mean pitch (PT), Start/End pitch (PT)), динамика (Interval (PT), Speed (PT)), а также показатели интенсивности и длительности в шкале 0–100.

На этой основе формируется правило выявления событий: слог помечается как акцентное событие A, если он

a) является проминантным по разметке/по совокупности акустических признаков и

b) образует локальный максимум/минимум по F0 (например, по Max pitch (PT) или по End pitch (PT)) в окрестности соседних слогов либо демонстрирует выраженную направленную динамику (по Interval (PT) и Speed (PT)).

Порог выраженности задаётся в тех же логарифмических единицах (полутоны), что и вычисление F0, и может соотноситься с порогами различимости, используемыми при обосновании ширины уровневых зон. Если слог выделен по интенсивности/длительности, допускается маркирование акцентного события без тональной спецификации.

(2) Граничные события (B). Границы определяются по независимым показателям членения: данным уровня marks в TextGrid и/или акустическим коррелятам границы (финальное удлинение, перераспределение интенсивности, характерное движение F0 на финальном участке), которые уже представлены в профиле через нормированные индексы длительности/интенсивности и через параметры Start/End pitch (PT), Interval (PT), Speed (PT).

Тип граничного события может задаваться через конфигурацию конечного участка:

a) через уровень конечной точки (End pitch (PT)) и

b) через знак/величину финального интервала (Interval (PT)).

В тех случаях, когда разметка границы подтверждается сегментацией/marks, но новая тональная инструкция отсутствует (малый интервал и малая скорость изменения), допускается специальная маркировка «нулевой» границы.

Предлагаемый подход реализует описание, где на первом уровне предоставлен набор точек для измерения базовых акустических параметров (F0, интенсивность, длительность), на втором уровне интонационные события (акценты и границы) моделируются как дискретные объекты, привязываемые к определённой позиции в слоговом профилю: к одиночному слогу, паре слогов или границе фразы, что обеспечивает:

1) сопоставимость количественных данных,
2) категориальное событийное представление, важное для лингвистической интерпретации и сопоставления с фонологической разметкой.

На основе нормированного представления F0 вычисляются производные характеристики, позволяющие описывать не только уровень, но и динамику внутри слога. Интервальная характеристика строится по начальной и конечной точкам слога и нормируется раздельно для подъёмов и падений, что отражается в обозначениях типа int_rise/int_fall/int_steady в столбце Interval (PT) (см. табл. 1).

При повышении величина изменения задаётся как доля конечного уровня:

$$t_{rise} = 100 \cdot \frac{PT_{end} - PT_{start}}{PT_{end}},$$

при снижении - как доля начального уровня:

$$t_{fall} = 100 \cdot \frac{PT_{start} - PT_{end}}{PT_{start}},$$

при отсутствии изменения фиксируется нулевое значение. Скорость изменения тона рассчитывается как модуль разности между конечным и начальным значениями в полутонах, отнесённый к длительности слога:

$$Speed_{ST} = \frac{|ST_{end} - ST_{start}|}{D},$$

где D выражена в секундах; в таблице скорость представлена в единицах st/s.

Абсолютные значения интенсивности фиксируются в децибелах для каждого слога (минимум, максимум, среднее, начало и конец интервала), после чего выполняется линейное преобразование в шкалу 0–100 на индивидуальных границах диапазона диктора. Для любой величины интенсивности I (например, среднего уровня или значения в конкретной точке) нормированный индекс задаётся формулой:

$$R_I = 100 \cdot \frac{I - I_{low}}{I_{high} - I_{low}},$$

где I_{low} и I_{high} – индивидуальные границы диапазона интенсивности данного диктора. В результирующем файле нормированные значения интенсивности представлены в столбцах вида Mean intensity (R), Max intensity (R), Start intensity (R), End intensity (R) и сопровождаются уровневыми интерпретациями (… description), что обеспечивает единообразие представления с контуром F0. Также сохраняются исходные значения в дБ (Min intensity, Max intensity, Start intensity, End intensity, Mean intensity).

Длительность слога D фиксируется в секундах, после чего нормируется в шкалу 0–100 относительно индивидуального диапазона длительностей данного диктора:

$$R_D = 100 \cdot \frac{D - D_{low}}{D_{high} - D_{low}},$$

Табл. 1. Проблемы современных транскрипций, решения предлагаемого подхода и сохраняющиеся риски/ограничения.
Table 1. Issues in contemporary transcription systems, the solutions offered by the proposed approach, and the remaining risks/limitations.

Проблема	Решение	Риски / ограничения
Субъективность и невысокая согласованность экспертов при выборе меток (особенно в пограничных случаях)	Первичный слой - числовые измерения; категориальный слой выводится по формальным пороговым правилам; результат фиксируется в таблицах, что облегчает проверку и повторяемость	Зависимость от качества исходной сегментации
Смешение уровней описания: неясно, где фонетические реализации, а где фонологические единицы	Разделение на два слоя: (а) градиентные параметры (F0/интенсивность/длительность и производные), (б) категориальные уровни, получаемые как дискретизация	Уровни в системе не равны фонологическим единицам; для интерпретации структура/значение требуется отдельный аналитический шаг
Ограничение описания только F0 в ряде транскрипций	Единая процедура извлечения и нормирования не только F0, но и интенсивности и длительности, плюс производные динамические показатели	Рост параметров увеличивает объём данных и требования к контролю качества
Категориальные категории там, где различия могут быть градиентны	Введение шкалы 0–100 и дискретизации, согласованной с минимально различимыми шагами (перцептивное обоснование шага/порога); возможность анализировать эффекты как непрерывные	Выбор числа уровней и порогов влияет на итоговую категориальность; при неудачных порогах возможно искусственное создание/сглаживание различий
Слабая сопоставимость дикторов, которая проявляется в (i) несопоставимости абсолютных величин между дикторами; (ii) несопоставимости формы без нормирования	Одновременное хранение абсолютных значений и нормированных (0–100 по диапазону диктора)	Нормирование чувствительно к тому, как оценивается индивидуальный диапазон
Неявность неопределённости: спорные случаи могут быть обозначены одним символом	Формализованный вывод категорий из чисел позволяет явно фиксировать пограничные зоны (по значениям около порогов) и хранить первичные данные для последующей проверки	В спорных случаях всё равно требуется правило: считать ли близкие к порогу значения разными уровнями или вводить специальную заранее описанную маркировку неопределённости
Проблема выравнивания событий во времени: события не всегда совпадают с границами слога/слова	Слоговой профиль как сетка для измерений + отдельный событийный слой (акцентные A и граничные B), который можно вычислять как производный от профиля	Слоговая сетка остаётся базовой; может потребоваться уточнение правил детекции событий
Масштабируемость	Автоматизируемые измерения и табличный вывод	Автоматизация не снимает необходимости контроля качества; при расширении набора параметров возрастает сложность верификации

где D_{low} и D_{high} - индивидуальные границы распределения длительностей. Нормированные значения длительности представлены в таблице в столбцах Duration и сопровождаются уровневыми описаниями (Duration description), что позволяет интерпретировать длительность как позиционно распределённый ресурс на той же шкале, что и F0/интенсивность. Таким образом, на уровне числового профиля для каждой слоговой позиции одновременно доступны исходные значения параметров (Гц/дБ/с) и их сопоставимые индексы 0-100, рассчитанные индивидуально по диктору.

Для получения категориальной записи нормированных значений используется пороговая система, основанная на разделении общей шкалы (0-100) на N дискретных уровней. Данная система применяется к нормированным индексам F0, интенсивности и длительности, переводя количественные измерения в ряд уровневых меток. Выбор N задаёт ширину зоны в процентах $\Delta_{\%} = 100/N$; при этом важно, чтобы такое разбиение не приводило к фиксации различий, которые не являются устойчиво различимыми в восприятии. Величина N может подбираться экспериментально, для русского языка нами было выбрано $N = 6$, поэтому нормированное значение $R \in [0; 100]$ соотносится с одной из зон:

$$\textit{ЭН:}\ R \le 16.7$$
$$\textit{Н:}\ 16.7 < R \le 33.3$$
$$\textit{СН:}\ 33.3 < R \le 50.0$$
$$\textit{СВ:}\ 50.0 < R \le 66.7$$
$$\textit{В:}\ 66.7 < R \le 83.3$$
$$\textit{ЭВ:}\ R > 83.3$$

В качестве проверочного условия используется принцип перцептивной достаточности: ширина зоны, выраженная в полутонах, должна быть не меньше порога различимости уровня F0 (JND_{st}). Для перехода от шкалы 0–100 к полутонам используется «вес полутона» $w_{st} = 100/R_{st}$, где R_{st} - индивидуальный диапазон $F0$ диктора в полутонах. Тогда ширина одной зоны в полутонах равна $\Delta_{st} = R_{st}/N$. Перцептивное условие задаётся неравенством $\Delta_{st} \ge k \cdot JND_{st}$, где JND_{st} - порог различимости уровня $F0$ (в полутонах), а $k \ge 1$ - множитель, задающий отношение ширины зоны к порогу различимости. В эквивалентной форме:

$$R_{st} \ge N \cdot k \cdot JND_{st}.$$

Порог различимости (минимально различимый шаг) для частот порядка 100 Гц составляет примерно 1–5 Гц, что соответствует приблизительно 0.17–0.84 полутона при пересчёте в логарифмические единицы [19; 20]. Если при выбранном N условие $\Delta_{st} \ge k \cdot JND_{st}$ не выполняется для конкретного диктора (или подкорпуса), число зон уменьшается либо объединяются крайние зоны, что обеспечивает соотнесение категориальной записи с перцептивно различимыми шагами по высоте.

Пороговая система применяется к нормированным индексам высоты тона, интенсивности и длительности; в результирующей таблице она представлена как уровневые подписи вида «Н: …», «СВ: …» и т. п. в столбцах описаний (… Description/… description), что обеспечивает интерпретацию транскрипции как последовательности уровней по слоговым позициям при сохранении исходных числовых значений и нормированных индексов (см. рис. 2).

Те же операции нормирования и уровневой интерпретации применяются к интенсивности и длительности - для них строятся аналогичные слоговые траектории и формируется транскрипционная запись на едином инвентаре уровней, что позволяет описывать различия как совокупность согласованных параметров (уровень и динамика F0, интенсивность, длительность), при сохранении единого способа представления данных: таблица измерений и нормированных индексов – как основание для статистического анализа, графики усреднённых траекторий – как способ сопоставления контуров по позициям, последовательность уровневых меток – как транскрипция, пригодная для дальнейшей категоризации и соотнесения с фонологическими метками (см. рис. 3, 4).

Таким образом, представленная схема формирует воспроизводимый формат описания, в котором числовой слой измерений и производных показателей служит основанием для последующей категориальной интерпретации. Вместе с тем очевидно, что предлагаемый подход не устраняет всех трудностей, характерных для современных транскрипций: часть проблем переносится на уровень предварительной разметки, выбора порогов, оценки индивидуальных диапазонов и процедур контроля качества. Поэтому необходимо отметить, какие именно ограничения учитываются на текущем этапе и какие остаются зонами риска, требующими дальнейшей доработки (см. табл. 1).

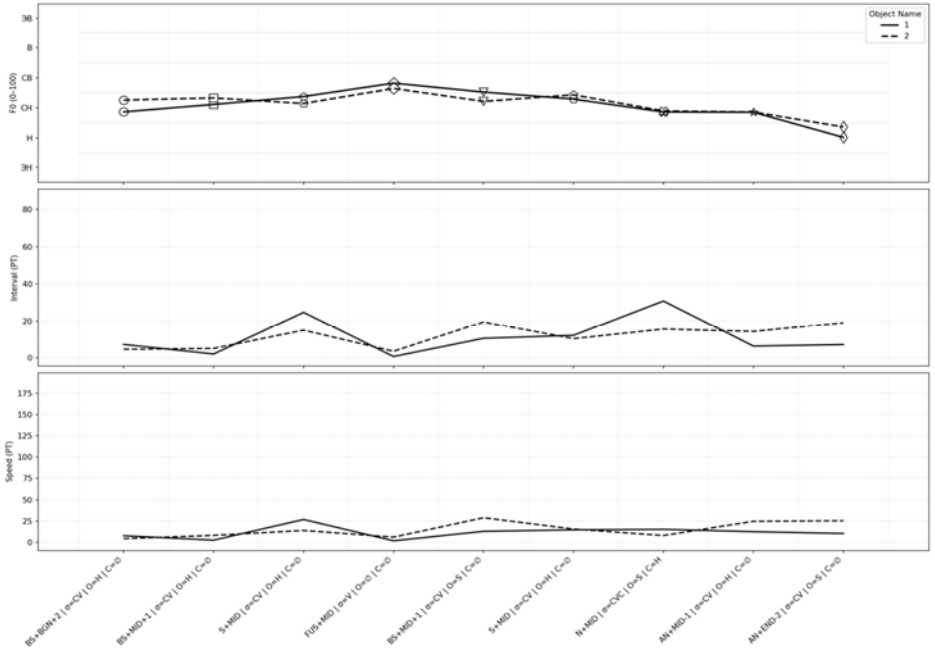


Рис. 2. Слоговые траектории F0 для двух вариантов: нормированный уровень (0–100) с пороговыми зонами (ЭН–ЭВ), интервальная характеристика (Interval) и скорость изменения (Speed).
Fig. 2. Syllable-level F0 trajectories for two variants: the normalized level (0–100) with threshold zones (ЭН–ЭВ), the interval measure (Interval), and the rate of change (Speed).

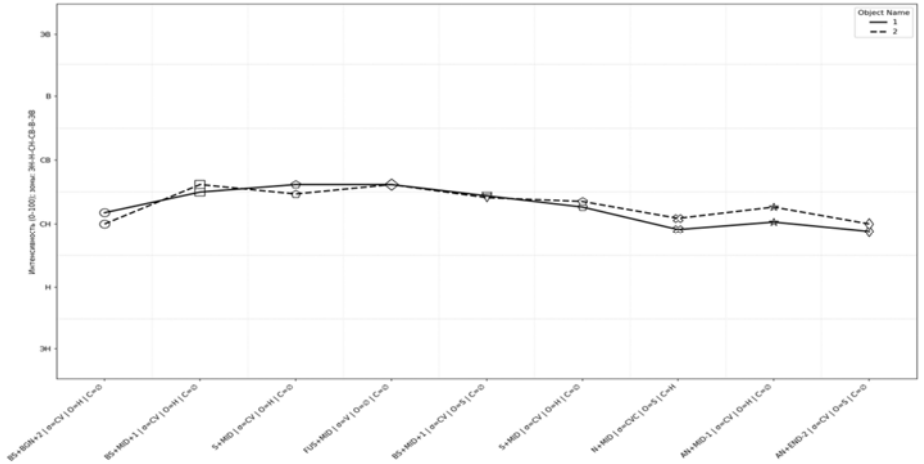


Рис. 3. Слоговые траектории интенсивности: нормированный уровень (0-100) и уровневые метки по пороговой дискретизации.
Fig. 3. Syllable-level intensity trajectories: the normalized level (0-100) and level labels based on threshold discretization.

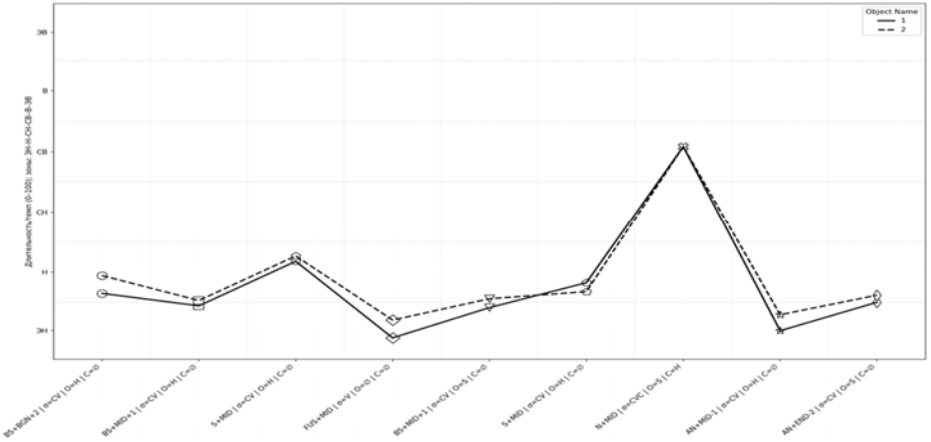


Рис. 4. Слоговые траектории длительности: нормированный уровень (0–100) и уровневые метки по пороговой дискретизации.
Fig. 4. Syllable-level duration trajectories: the normalized level (0–100) and level labels based on threshold discretization.

6. Заключение

Сопоставительный анализ просодии требует такого способа описания, который сохраняет измеримую структуру сигнала и в то же время допускает интерпретацию в дискретных терминах. Рассмотренные системы просодической транскрипции и процедуры нормализации показывают, что универсального решения не существует: одни подходы повышают интерпретируемость ценой неполной воспроизводимости, другие обеспечивают формальную упорядоченность, но теряют часть параметров просодии или затрудняют сопоставление. В связи с этим обоснованной является постановка задачи описания как единой процедуры получения сопоставимых измерений и как нотации, пригодной для статистического анализа.

Предложенная система решает данную задачу через совокупность требований к данным и к записи результатов: формализованное извлечение и фиксация параметров, сохранение абсолютных характеристик при наличии относительной шкалы сравнения, применение одного способа приведения к сопоставимому виду для нескольких просодических переменных, наличие числового и категориального представлений, однозначные правила категоризации, устойчивость к аномальным данным. Согласно данным требованиям, транскрипция превращается в воспроизводимую процедуру, обеспечивающую сопоставимость описаний.

Таким образом, предложенная транскрипция обеспечивает представление просодии в виде (а) числовых измерений и производных показателей и (б) дискретных уровневых меток, получаемых по заданным правилам, что создаёт основу для сопоставления говорящих, речевых ситуаций и языковых материалов без утраты абсолютных характеристик и без сведения просодии к одному тональному параметру.

Перспективы развития подхода связаны с проверкой устойчивости выбранных порогов и правил категоризации на расширенных данных, а также с оценкой того, как уровневые классы и регистровые коды соотносятся с перцептивными категориями и фонологическими противопоставлениями конкретного языка.

Список литературы / References

[1]. Jun S. A. The ToBI transcription system conventions strengths and challenges. Prosodic Theory and Practice. Barnes J., Shattuck Hufnagel S. (eds.). Cambridge, MA: The MIT Press, 2022, 151-181. DOI: 10.7551/mitpress/10413.003.0007.

[2]. Cole J., Shattuck Hufnagel S. New methods for prosodic transcription capturing variability as a source of information. Laboratory Phonology, 2016, vol. 7, no. 1, pp 1-29, art. 8. DOI: 10.5334/labphon.29.

[3]. Nolan F. The rise and fall of the British school of intonation analysis. Prosodic Theory and Practice. Barnes J., Shattuck Hufnagel S. (eds.). Cambridge, MA: The MIT Press, 2022, pp. 319-350. DOI: 10.7551/mitpress/10413.003.0012.

[4]. Crystal D. Intonation and grammar in British English. By M. A. K. Halliday. (Book review). Language, 1969, vol. 45, no. 2, pp. 378-393.

[5]. Breen M., Dilley L. C., Kraemer J., Gibson E. Inter Transcriber reliability for two systems of prosodic annotation ToBI and RaP. Corpus Linguistics and Linguistic Theory, 2012, vol. 8, no. 2, pp. 277-312.

[6]. Kodzasov S. A new descriptive system for Russian intonation. Proceedings of the 14th International Congress of Phonetic Sciences *(ICPhS 14)*, 1999, pp. 913-916.

[7]. Кодзасов С. В. Проект просодической транскрипции для русского языка. Бюллетень фонетического фонда русского языка, 1989, 2.

[8]. Biron T., Barboy M., Ben Artzy E., Golubchik A., Marmor Y., Szekely S., Winter Y., Harel D. Non verbal information in spontaneous speech towards a new framework of analysis. arXiv, 2024. DOI: 10.48550/arXiv.2403.03522, v2.

[9]. Post B., Delais Roussarie E. Transcribing intonational variation at different levels of analysis. Speech Prosody 2006: Proceedings of the 3rd International Conference on Speech Prosody (Dresden, Germany, May 2–5, 2006), 2006. DOI: 10.21437/SpeechProsody.2006-182.

[10]. Hirst D. Intonation in British English. Intonation Systems. Hirst D., Di Cristo A. (eds.). Cambridge: Cambridge University Press, 1998, pp. 56-77.

[11]. Hirst D., Di Cristo A. A survey of intonation systems. Intonation Systems. Hirst D., Di Cristo A. (eds.). Cambridge: Cambridge University Press, 1998, 1–44.

[12]. Taylor P.A. The tilt intonation model. Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 1998), 1998, pp. 1383-1386.

[13]. Reichel U.D. Linking bottom up intonation stylization to discourse structure. Computer Speech and Language, 2014, vol. 28, no. 6, pp. 1340-1365. DOI: 10.1016/j.csl.2014.03.005.

[14]. Kochanski G., Shih C. Prosody modeling with soft templates. Speech Communication, 2003, vol. 39, no. 3-4, pp. 311-352. DOI: 10.1016/S0167-6393(02)00047-X.

[15]. Thangthai A., Thatphithakkul N., Wutiwiwatchai C., Rugchatjaroen A., Saychum S. T Tilt a modified Tilt model for F0 analysis and synthesis in tonal languages. Proceedings of Interspeech 2008, 2008, pp. 2270-2273. DOI: 10.21437/Interspeech.2008-557.

[16]. Ladd D.R. The trouble with ToBI. Prosodic Theory and Practice. Barnes J., Shattuck Hufnagel S. (eds.). Cambridge, MA: The MIT Press, 2022, pp. 247-257. DOI: 10.7551/mitpress/10413.003.0009.

[17]. Fitch J.L., Holbrook A. Modal vocal fundamental frequency of young adults. Arch Otolaryngol., 1970, vol. 92, no. 4, pp. 379-382. DOI: 10.1001/archotol.1970.04310040067012.

[18]. Keating P., Buhr R. Fundamental frequency in the speech of infants and children. J Acoust Soc Am., 1978, vol. 63, no. 2, pp. 567-571. DOI: 10.1121/1.381755.

[19]. Edlund J., Heldner M. Automatic estimation of pitch range and speaker relative pitch. Speaker Classification II: Selected Papers. Müller C. (ed.). Berlin; Heidelberg: Springer, Lecture Notes in Computer Science, 4441, 2007, pp. 229-242.

[20]. 't Hart J., Collier R., Cohen A. A Perceptual Study of Intonation: An Experimental Phonetic Approach to Speech Melody. Cambridge: Cambridge University Press, 1990. 232 p.

Информация об авторах / Information about authors

Оксана Владимировна ГОНЧАРОВА – кандидат филологических наук, доцент, старший научный сотрудник лаборатории Лингвистических платформ НИИ «Институт системного программирования им. В. П. Иванникова РАН» с 2024 года, доцент кафедры русского языка и методики его преподавания, Российский университет дружбы народов им. П. Лумумбы. Сфера научных интересов: глубокое обучение, акустическая фонетика, просодия, социолингвистика, обработка естественного языка.

Oxana Vladimirovna GONCHAROVA – Cand. Sci. (Philology), Assoc. Prof., Senior Researcher at the Laboratory of Linguistic Platforms of the Ivannikov Institute for System Programming of the Russian Academy of Sciences since 2024. Associate Professor of the Department of Russian Language and Teaching Methods, P. Lumumba Peoples' Friendship University of Russia. Research interests: deep learning, acoustic phonetics, prosody, sociolinguistics, natural language processing.

.