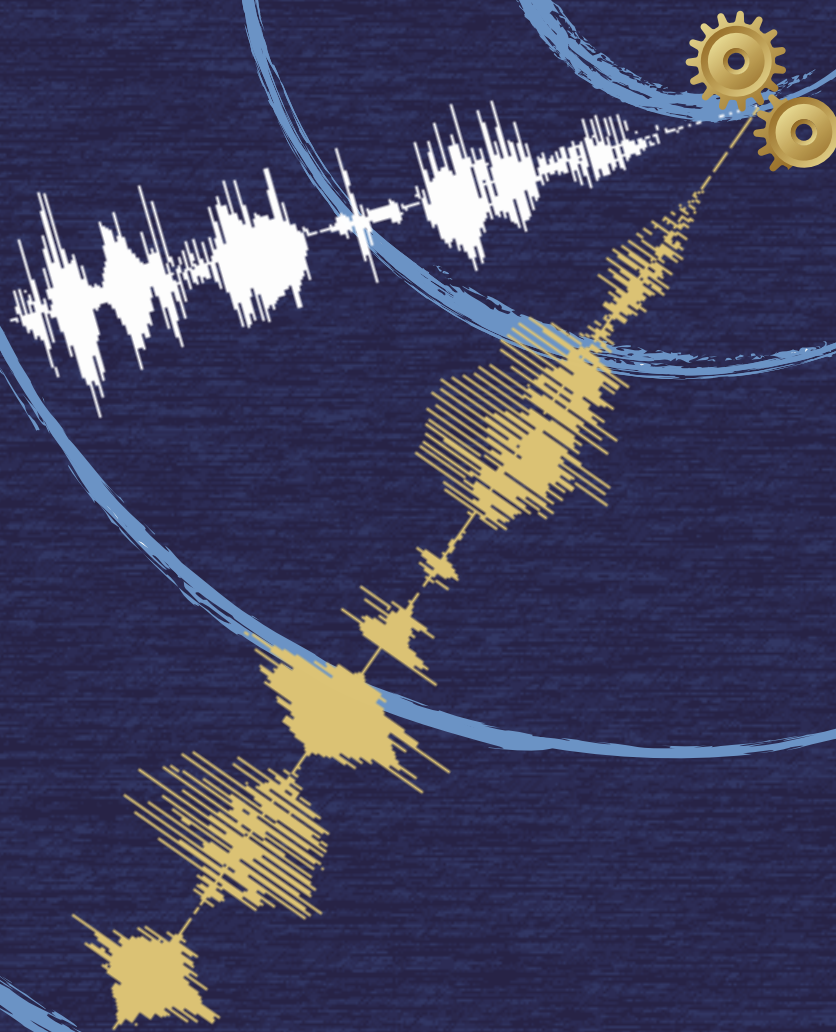


НАУЧНО-ТЕХНИЧЕСКИЙ
ЖУРНАЛ

РЕЧЕВЫЕ ТЕХНОЛОГИИ



Speech technology

1

2022



«IntonTrainer-ru» — программный комплекс для самообучения и тренинга нормативной интонации русской речи

Лобанов Б. М., главный научный сотрудник, ОИПИ НАН Беларуси, Минск, Беларусь, lobbortef@gmail.com

Житко В. А., аспирант, ОИПИ НАН Беларуси, Минск, Беларусь, zhitko.vladimir@gmail.com

В статье рассмотрены теоретико-экспериментальные основы построения систем оперативной оценки характеристик интонации русской речи. Дано описание разработанной компьютерной системы, её технические характеристики, области применения и опыт использования.

• интонация • русская речь • самообучение • тренинг

ВВЕДЕНИЕ

По словам известного лингвиста Максима Кронгауза «Интонация — самая эфемерная составляющая устной речи. На письме она передается весьма условно и это, может быть, самая загадочная область фонетики, исследования которой еще далеки от завершения». С физической точки зрения интонация — это совокупность просодических характеристик акустического речевого сигнала, таких как: мелодика, ритмика, темп, паузация и громкость речи. Из них наиболее универсальными и значимыми являются частотно-временные характеристики речевого сигнала, а именно — мелодика и текущий темп речи.

Интонация позволяет на практике и в искусстве выражать наши мысли, чувства, волевые устремления и внутреннее состояние наряду со словом, помимо слова, а иногда и вопреки ему. А. С. Макаренко, считавший умение педагога владеть своей речью одним из основных, говорил: «Я сделался настоящим мастером только тогда, когда научился говорить «иди сюда» с 15-20 оттенками, и тогда я не боялся, что кто-то ко мне не подойдет или не почувствует того, что нужно».

Интонация представляет собой важнейшее средство передачи как собственно языковой, так и социокультурной информации. В практике преподавания существует мнение, что иноязычный акцент особенно ярко

проявляется в интонации, поэтому при обучении и преподавании иностранных языков ей следует уделять особенное внимание.

Например, американские носители языка делают следующее интересное наблюдение: «Спросите у среднего американца, что они думают о российском акценте, и они говорят: "Русские звучат недружелюбно. Я чувствую, как будто они не любят меня. Одна из причин того, что русские, говорящие на английском языке, не звучат дружелюбно — это их плоский тон. Они просто не используют правильную интонацию во время разговора"».

Правильность воспроизведения интонации при говорении и адекватность её восприятия при слушании с трудом поддаётся самоконтролю учащимся (особенно при отсутствии музыкального слуха). Существующие лингафонные курсы и оборудование обеспечивают только слуховую обратную связь контроля правильности интонирования речи, что явно *недостаточно*.

В статье рассмотрены теоретико-экспериментальные основы построения систем оперативной оценки характеристик интонации русской речи. Дано описание разработанного программного комплекса для самообучения и тренинга нормативной интонации русской речи, его технические характеристики, принципы работы и опыт использования.

1. ИНТОНАЦИОННАЯ МОДЕЛЬ

В 1960-х гг. Е. А. Брызгунова предложила описание интонации русского языка [1] с использованием понятия интонационной конструкции (ИК), которое вошло в академическое издание русской грамматики [2] и стало использоваться в методических пособиях по обучению русского языка как иностранного (РКИ) [3]. За основу классификации интонационных конструкций принят характер движения тона на ударном и прилегающих к нему слогах. Е. А. Брызгунова выделяет семь основных интонационных конструкций русской речи, различающих смысл звучащих предложений:

- ИК1 наблюдается при выражении завершенности в повествовательных предложениях: *Анна стоит на мосту. Наташа поет.*

Для ИК1 характерно понижение тона на ударной части.

- ИК2 реализуется в вопросе с вопросительными словами: *Кто пьет сок? Как поет Наташа?*
- При ИК2 ударная часть произносится с некоторым повышением тона.
- ИК3 характерна для вопроса без вопросительного слова: *Это Антон? Ее зовут Наташа?*

Для этой интонации характерно значительное повышение тона на ударной части.

- ИК4 — это вопросительная интонация, но с сопоставительным союзом /а /: *А она приехала? А дедушка?*



На ударной части происходит понижение, затем повышение тона, продолжающееся на безударных слогах.

- ИК5 реализуется при выражении оценки в предложениях с местоименными словами: *Какой сегодня день!*

На ударной части — повышение тона.

- ИК6, так же как и ИК5, реализуется при выражении оценки в предложениях с местоименными словами: *Какой сок вкусный!*

Повышение тона происходит на ударной части и продолжается на заударной части.

- ИК7 употребляется в предложениях со значением экспрессивного отрицания признака, действия, состояния: *Да какой он специалист!*

Повышение тона происходит на ударной части.

Очевидно, что, представленное выше описание ИК1-ИК7, также, как и более развёрнутое описание, приведенное в [2], не являются полными и строгими. Эти словесные описания не могут вполне удовлетворить разработчиков компьютерных моделей анализа и синтеза интонационных характеристик речи.

Более 30 лет назад [4] нами была предложена ПАЕ-модель описания интонации (модель Портретов Акцентных Единиц), которая с тех пор успешно использовалась во многих разработках систем анализа и синтеза речи на русском и других языках. В соответствии с ПАЕ-моделью, минимальной просодическим компонентом, из которого составляется интонация синтагмы или фразы, является Акцентная Единица (АЕ). АЕ может состоять из одного или более фонетических слов, причём одно из них должно иметь в своём составе полноударный (ядерный) слог. Каждая АЕ, в свою очередь, состоит из ядра (ударная гласная фонема), пред-ядра (все фонемы, предшествующие ударной гласной) и за-ядра (все фонемы за ударной гласной). ПАЕ-модель предполагает, что для определенного типа интонации топологические свойства мелодического контура АЕ не зависят от количественного и качественного содержания пред-ядра, ядра и за-ядра.

В работах [5-7] показано, что ПАЕ-модель обеспечивает наглядное представление интонационных конструкций русской речи (ИК1 — ИК7) — в виде набора их Универсальных Мелодических Портретов (УМП) в нормированных координатах «**Частота — Время**». Нормализация по времени осуществляется путём приведения к стандартной длине элементов АЕ: пред-ядерных, ядерных и за-ядерных участков. Этот вид нормализации устраняет различия мелодической кривой, связанные с количественным составом пред-ядерных и за-ядерных участков АЕ.

Для нормализации по частоте определяются минимальное ($F_{0\ min}$) и максимальное ($F_{0\ max}$) значения частоты основного тона — F_0 для мелодической

кривой в произнесении данного диктора. Нормализация осуществляется в соответствии с формулой:

$$F_0^N = (F_0 - F_{0\min}) / (F_{0\max} - F_{0\min}).$$

Этот вид нормализации устраняет различия мелодической кривой, связанные с индивидуальными дикторскими различиями в высоте голоса.

Таким образом, нормированное пространство для отображения УМП ИК_i может быть представлено в виде прямоугольника с координатными осями (T_N, F_{0N}) , как это представлено на схематическом рисунке 1.

При этом интервалам на оси абсцисс T_N соответствуют: $[0-1/3]$ — пред-ядро, $[1/3-2/3]$ — ядро, $[2/3-1]$ — за-ядро.

Интервалам на оси ординат F_{0N} соответствуют: $[0-1/3]$ — низкий уровень тона, $[1/3-2/3]$ — средний, $[2/3-1]$ — высокий.

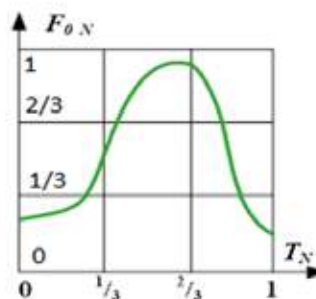


Рис. 1. Общий вид УМП

Такое представление обеспечивает основное внимание к особенностям формы кривой ЧОТ на ядре при меньшем внимании на количественные и качественные составы пред- и за-ядра.

2. СТРУКТУРНАЯ СХЕМА ПРОГРАММНОГО КОМПЛЕКСА «INTONTRAINER-RU»

В состав программного комплекса (ПК) входят подсистемы, включающие наборы эталонных фраз, которые представляют основные интонационные модели русской речи. В ходе процесса обучения ПК «IntonTrainer-Ru» осуществляет сравнение произнесенной и эталонной фраз и дает численную оценку их интонационного сходства.

Для создания данной системы потребовалось решить ряд принципиальных проблем, к которым относятся:

- адекватное сопоставление сигнала произносимой фразы (с заранее неизвестным началом и окончанием) с эталонной фразой при условии их взаимной временной деформации;
- сегментация анализируемого сигнала на участки, для которых понятие ЧОТ имеет смысл с точки зрения формирования интонационного контура фразы в целом: участки гласных и большинства сонорных согласных;
- высокоточное вычисление частоты основного тона, произносимого и эталонного сигналов без подстройки для мужских и женских голосов в широком диапазоне (от 50 до 500 Гц);



- интерполяция значений ЧОТ на тех участках, для которых определение ЧОТ является некорректным (большинство согласных звуков);
- представление фраз в виде УМП ИК русской речи и вычисление численного сходства УМП произносимой и эталонной фраз.

На рисунке 2 приведена структурная схема компьютерной обучающей системы.



Рис. 2. Структурная схема компьютерной системы обучения «IntonTrainer»

В БД эталонов хранятся образцы фраз различных ИК и примеров их употребления. Для каждой из эталонных фраз проводится предварительная просодическая разметка, включающая указание фразовых (синтагматических) границ, а также положение ядерных, пред-ядерных и за-ядерных участков во фразе. Учащийся, в соответствии с предлагаемой методикой обучения интонации, выбирает нужные фразы, прослушивает их и затем произносит их. При этом, в соответствии с рисунком 2, осуществляется запись речевого сигнала в буфер и затем производится обработка эталонного и произнесённого сигналов.

Рассчитываются спектры эталонного и произнесённого сигналов, затем с использованием метода динамического программирования (ДП) осуществляется их сопоставление, перенос просодических меток и разметка произносимых фраз. Далее производится расчёт частоты основного тона F_0 обоих сигналов, осуществляются операции медианного сглаживания, интерполяции траектории F_0 на неголосовых участках

и осуществляется её нормировка относительно минимального и максимального значений. Затем рассчитываются и сравниваются УМП эталонного и произнесённого сигналов.

На конечном этапе осуществляется экранное отображение результатов анализа, позволяющее получить визуальную, слуховую и численную оценку интонационного сходства произносимой фразы с эталонной.

3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ И ИНТЕРФЕЙС ПК «INTONTRAINER–RU»

3.1. Начальный режим просмотра эталонных фраз

Начальное окно программного комплекса (ПК), открывающееся после запуска программы, показано на рисунке 3.

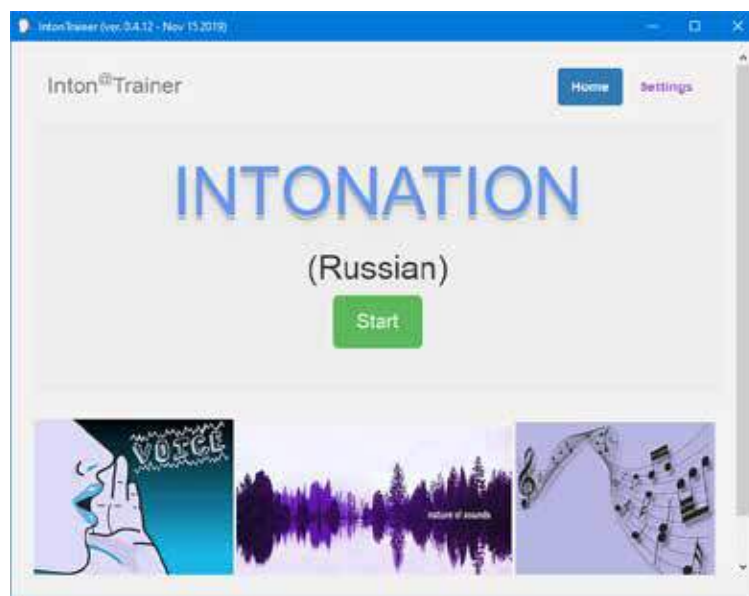


Рис. 3. Начальное окно ПК

При нажатии кнопки «**Start**» (Рис. 3) открывается Информационное окно ПК (Рис. 4), содержащее структурированный перечень текстов эталонных аудиофраз, произнесенных женским и мужским голосами.

Прокручивая страницу информационного окна (Рис. 4), пользователь получает возможность слухового и визуального ознакомления с основными интонационными конструкциями русской речи путём выбора с помощью курсора требуемой директории.

При выборе, например: ИК1 «(ж) Он гуляет в парке с собакой», открывается Главное Окно ПК (см. рис.5).

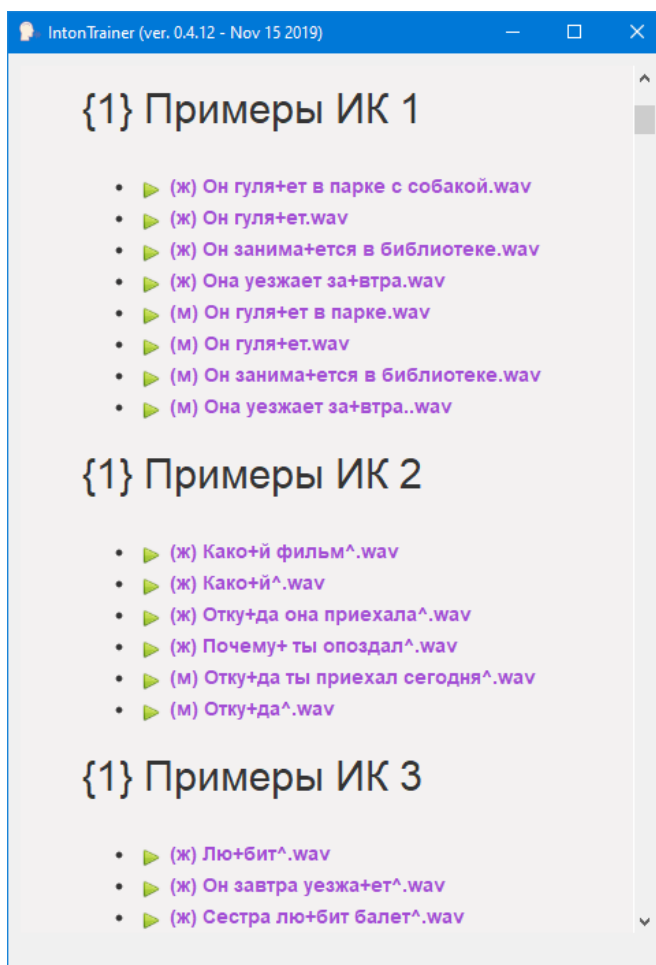


Рис. 4. Информационное окно

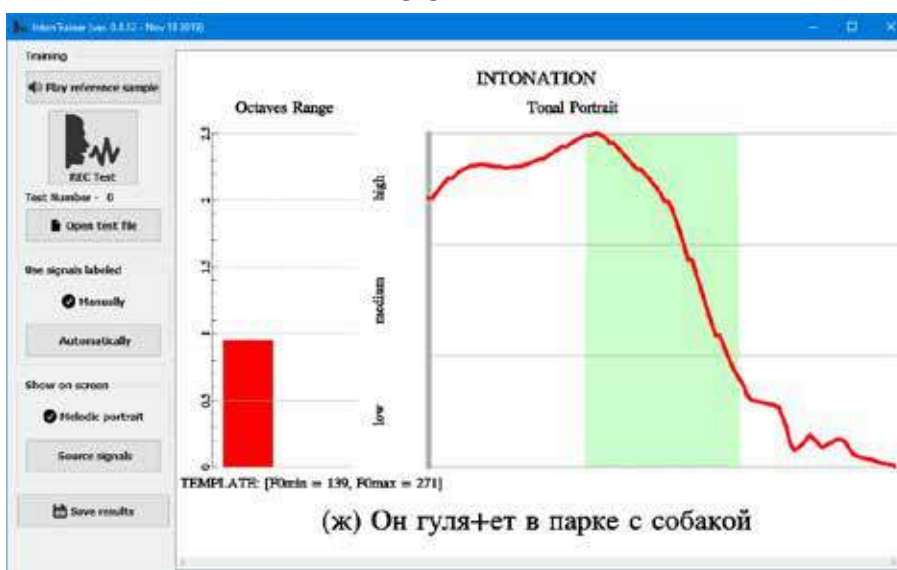


Рис. 5. Главное окно ПК

В левой части окна сосредоточены органы управления, позволяющие осуществлять прослушивание эталонного речевого сигнала — Play reference sample, запись учащимся тестовой фразы — **REC Test**, вызов заранее подготовленных тестовых фраз — Open test file, а также задавать различные режимы графического отображения анализируемых сигналов и результатов интонационного анализа.

В правой части окна в графическом виде представлены результаты анализа эталонной фразы:

- График отображения изменений ЧОТ во фразе в виде УМП — Tonal Portrait, состоящего из трёх нормированных во времени участков: пред-ядра, ядра и за-ядра. Ядро помечено зеленым цветом.
- Красный столбец (слева от мелодической кривой) отображает диапазон изменения частоты основного тона во фразе — Octaves Range, рассчитываемый по формуле:

$$\text{Octave's Range} = (F0_{\max}/F0_{\min}) - 1.$$

Внизу под графиками для выбранной эталонной фразы приведены минимальное и максимальное значения F0 в Герцах — **TEMPLATE: [F0min = 139, F0max = 271]**, а также текст фразы, в которой ядерный гласный обозначен знаком «+».

На рисунке 6 представлены копии с экрана эталонных примеров для каждой из ИК1 — ИК8, произнесенных квалифицированным преподавателем русского языка для иностранцев. Различия в форме рассчитанных кривых УМП вполне очевидны.

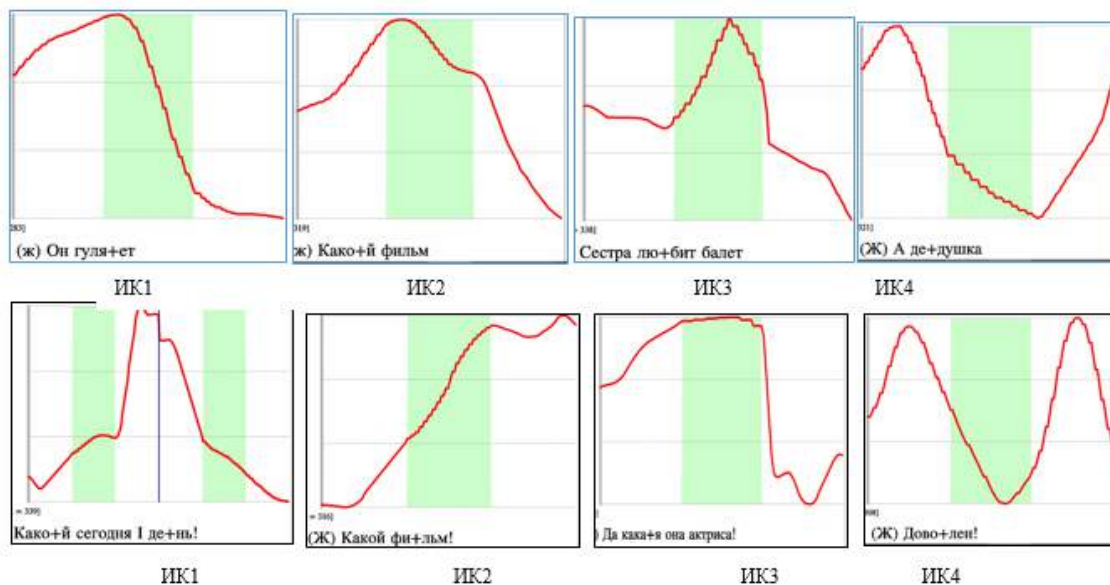


Рис. 6. Вид кривых УМП, рассчитанных для каждой из ИК1 — ИК8



3.2. Рабочий режим сравнения произносимых и эталонных фраз

Основное назначение ПК «IntonTrainer» — предоставление пользователю возможности индивидуального интонационного тренинга при изучении русского как иностранного (РКИ), или для совершенствования уже имеющихся устно-речевых интонационных навыков русскоязычного пользователя.

После того, как пользователь прослушал и просмотрел УМП выбранной фразы, он может воспользоваться внешним или встроенным микрофоном для ввода речевого сигнала своим голосом. При этом он должен нажать кнопку «REC Test» в левой секции окна (Рис. 5), дождаться короткого «бип-сигнала» и произнести в микрофон фразу, текст которой указан в нижней части окна на рисунке 5.

После записи введенного речевого сигнала пользователь услышит 2-й «бип-сигнал», а изображение в окне (Рис. 5) заменится на изображение, показанное на рисунке 7.

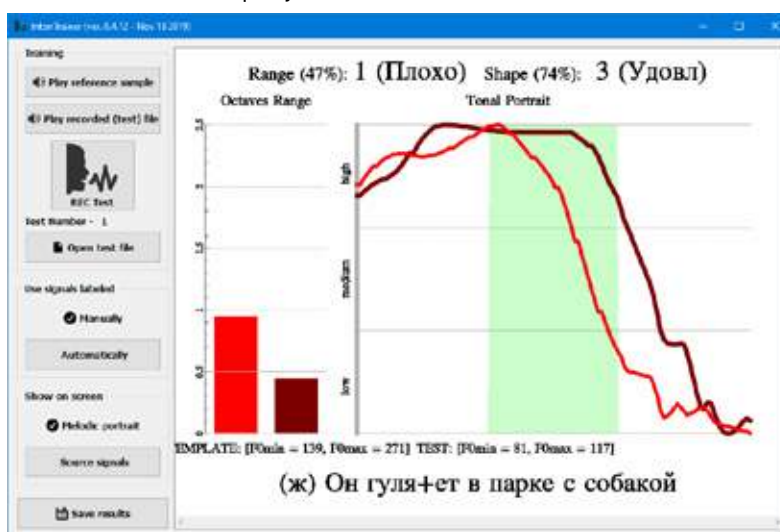


Рис. 7. Главное окно ПК
(результат сопоставления УМП эталонной и произнесенной фразы с ИК1)

На рисунке 7 представлены результаты сопоставления УМП эталонной фразы «(ж) Он гуля+ет в парке с собакой» (красные линия и столбик слева) и произнесённой учащимся той же фразы (коричневые линия и столбик слева). Оценка интонационного сходства фраз осуществляется по двум основным критериям: степени сходства диапазонов изменения F0 — Range, и подобия кривых УМП — Shape.

В верхней части окна показаны результаты сравнения эталонной и произнесённой фраз: Range (47 %) — близость по диапазону изменения F0 (t), и Shape (74 %) — близость по форме сравниваемых кривых УМП. Рядом с процентными оценками близости проставлены оценки по пятибалльной школьной системе.

Учащийся, не удовлетворившись результатом теста, имеет возможность многократно повторить фразу, добиваясь желательного освоения навыка правильного интонирования каждой из выбранных ИК русской речи. На рисунке 8 показан улучшенный в сравнении с рисунке 7 результат интонирования той же фразы, достигнутый учащимся.

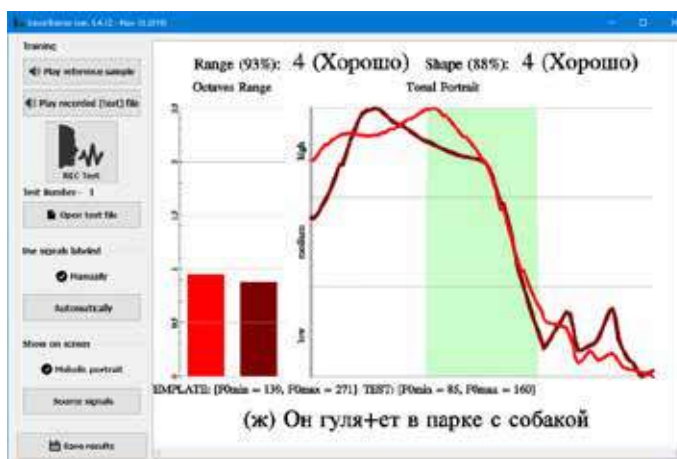


Рис. 8. Главное окно ПК

(результат повторного сопоставления УМП эталонной и произнесённой фразы с ИК1)

На рисунке 9 представлены УМП достаточно успешного сопоставления двух фраз с различными ИК.

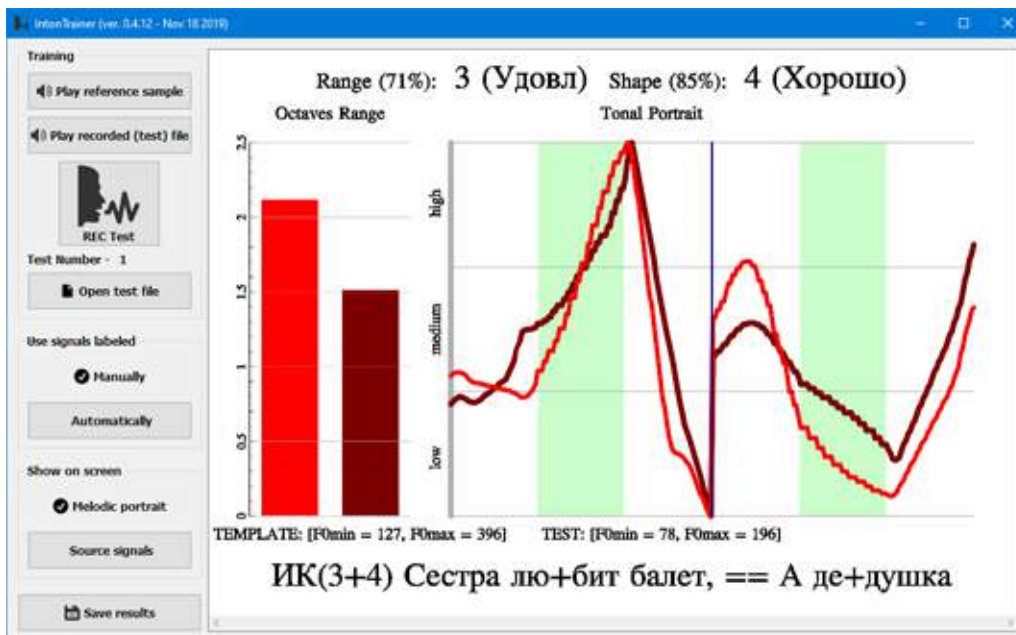


Рис. 9. Главное окно ПК

(результат сопоставления эталонных и произнесенных двух фраз — ИК3 и ИК4)



3.3. Состав БД русскоязычных эталонных фраз

Структурированный перечень текстов эталонных фраз, произнесенных женским и мужским голосами, хранится в папке «PATTERNS». Папка содержит:

- Отдельные примеры основных интонационных конструкций:

ИК1 — 8 примеров, ИК2 — 6, ИК3 — 6, ИК4 — 4, ИК5 — 5, ИК6 — 7, ИК7 — 4, ИК8 — 4;

- Сопоставительные пары ИК1— 8-29 примеров;
- Употребления каждой из ИК в различных речевых ситуациях — 188 примеров, в том числе:

ИК1 — 16 (4 ситуации по 4 примера в каждой ситуации);

ИК2 — 48 (12 ситуации по 4 примера в каждой);

ИК3 — 36 (9 ситуаций по 4 примера в каждой);

ИК4 — 36 (9 ситуаций по 4 примера в каждой);

ИК5 — 12 (3 ситуации по 4 примера в каждой);

ИК6 — 20 (5 ситуаций по 4 примера в каждой);

ИК7 — 8 (2 ситуации по 4 примера в каждой);

ИК8 — 12 (3 ситуации по 4 примера в каждой ситуации);

- Отрывок из прозы А. П. Чехова;
- Стихотворение Александра Блока.

При создании эталонных фраз для основных аудиопримеров ИК1 — ИК7 использовалась интернет-версия методического пособия по обучению русского языка как иностранного [3].

Программный комплекс «IntonTrainer-Ru» является открытой системой. Допускается модификация используемых настроек и аудиоматериалов. Эталонные базы данных — PATTERNS — могут быть дополнены или модифицированы в соответствии с поставленной задачей, либо созданы заново для работы с новыми языковыми приложениями.

ЗАКЛЮЧЕНИЕ

Описанный в статье программный комплекс «IntonTrainer-Ru» размещен на сайте <https://intontrainer.by>. Подробное описание работы

с программой приведено на сайте в PDF файле «User Guide». ПК «IntonTrainer» реализован в виде отдельного приложения под ОС Windows (7, 8, 10) и Linux (Debian, Ubuntu). В реализации используется библиотека Qt в качестве отладочно-пользовательского интерфейса и ядра приложения, а также библиотека MathGL для вывода графиков. Для реализации основного пользовательского интерфейса используется приложение, написанное на HTML/JS. Алгоритмы расчёта и обработки данных написаны на языке C с использованием библиотек GNU Scientific Library, Speech Signal Processing Toolkit (SPTK) и OpenAL.

ПК «IntonTrainer-Ru» общедоступен для бесплатной загрузки и использования. Кроме него на сайте на тех же условиях размещены обучающие ПК «IntonTrainer-Chin» (китайский), «IntonTrainer-AmEng» (Английский), «IntonTrainer-Ger» (немецкий), «IntonTrainer-Song» (пение), а также «Speech Rate Meter», «Speech Melody Meter», «Singer Voice Tester» — измерители характеристик темпа, мелодии речи и пения.

ПК «**IntonTrainer**» рекомендуется к использованию в следующих достаточно широко востребованных приложениях:

- в лингвистическом образовании, как средство визуализации интонации;

(Первичное ознакомление и изучение основных интонационных конструкций (ИК) русской речи, их попарного сопоставления, особенностей употребления, а также их реализации в диалоге, прозе и стихах.)

- для самостоятельного обучения и самоконтроля, как средство интонационного тренинга;

(Персональный тренинг ИК русской речи: изучение русского как иностранного (РКИ), совершенствование устно-речевых интонационных навыков в ряде профессий, таких как операторы колл-центров, дикторы радио, ТВ и др.)

- в научно-практических исследованиях, как средство сравнения интонации от разных источников.

(Исследование индивидуальных, эмоциональных и стилистических особенностей реализации интонации. Сравнительная оценка интонации речи в норме и патологии. Оценка интонационного качества синтезированной речи.)

Программные комплексы, размещённые на сайте, выпущены под разрешительной лицензией MIT, что позволяет программистам использовать лицензируемый код в закрытом программном обеспечении при условии, что текст лицензии предоставляется вместе с ним.

К настоящему времени закончена также разработка мобильных версий систем «Speech Rate Meter», «Speech Melody Meter», «Singer Voice Tester» в OS Android, которые доступны в Google Play.

С использованием возможностей сайта <https://datastudio.google.com/navigation/reporting> проведен анализ активности обращений пользователей к нашему сайту. К настоящему времени (с начала 2020 г.) зафиксировано 5600 обращений к сайту от пользователей 770 населенных пунктов из 113 стран. На рисунке 10 представлено



распределение числа обращений к сайту для первых 25 стран (без Беларуси).

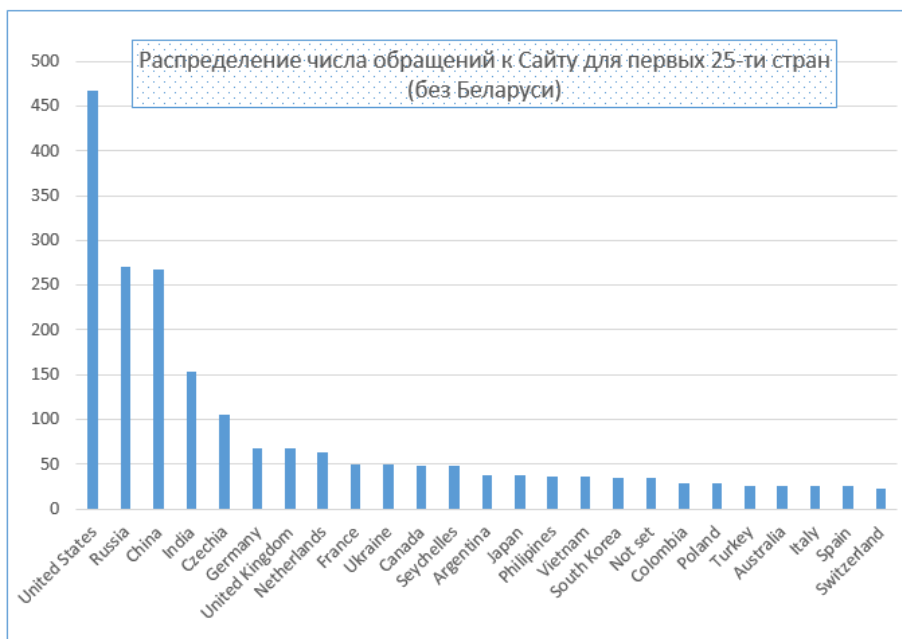


Рис. 10. Распределение числа обращений к сайту <https://intontrainer.by>

На экранных копиях (Рис. 11-14) наглядно представлена география обращений пользователей различных стран к сайту.



Рис. 11. Все страны на карте мира

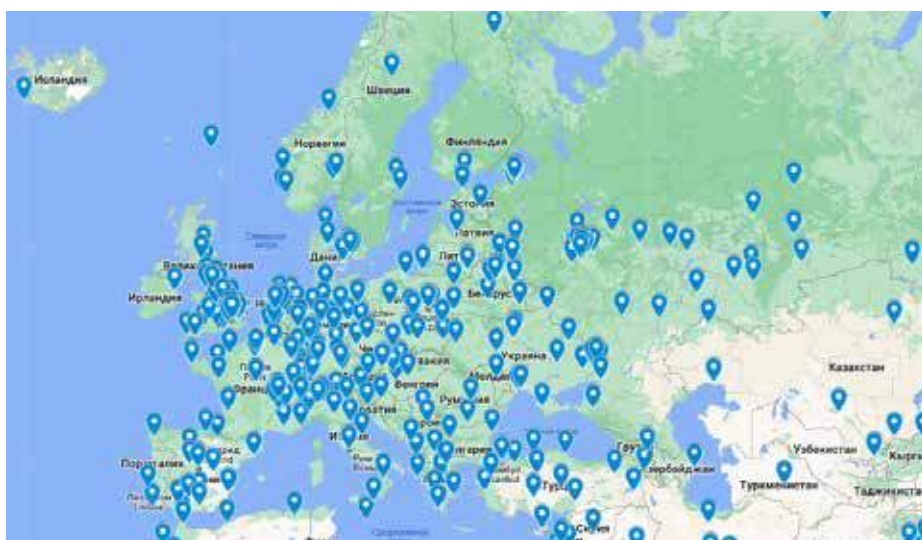


Рис. 12. Страны Европы

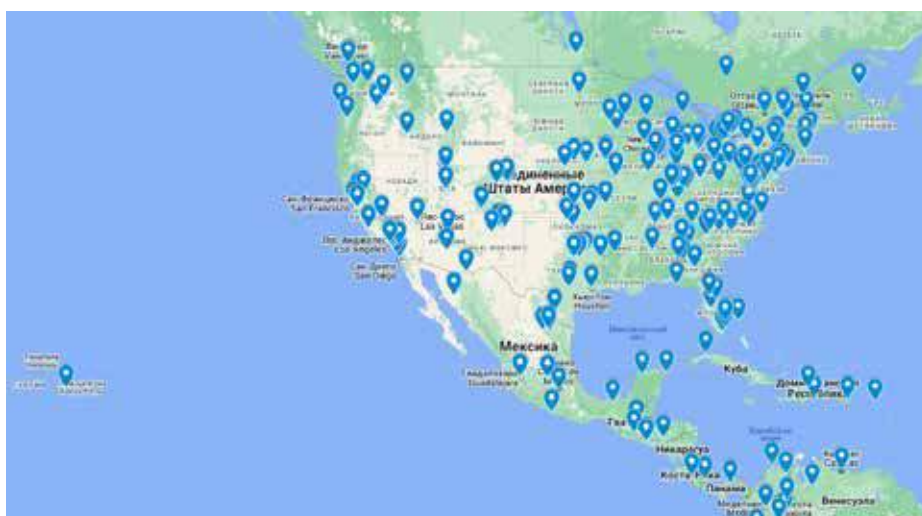


Рис. 13. Страны Северной и Центральной Америки



Рис. 14. Страны Африки, Азии, Австралии

Реальные данные, представленные на рисунках 10-14 достаточно убедительно свидетельствуют о востребованности разработанных программных продуктов, а также попутно наглядно показывают степень интенсивности «речевых» исследований и разработок в отдельных регионах и странах мира.

Литература

1. Брызгунова Е. А. Звуки и интонация русской речи/ — Наука, М., — 1968.
2. Брызгунова Е. А. Интонация // Русская грамматика/ — Наука, М., — 1980.
3. Одинцова И. В. Звуки. Ритмика. Интонация. М.: Флинта-Наука, — 2011.
4. Lobanov B. The Phonemophon text-to-speech system // International Congress of Phonetic Sciences: proc. of the 11-th seccion ICPhS'87, — Tallin, — 1987 — P. 120-124.
5. Lobanov B. Language- and Speaker Specific Implementation of Intonation Contours in Multilingual TTS Synthesis/Lobanov B., Tsirolnik L., Zhadinets D., Karnevskaia E. // Speech Prosody: proceedings of the 3rd International conference. Dresden, Germany — Vol. 2. — 2006 — P. 553-556.
6. Лобанов Б. М. Компьютерный синтез и клонирование речи/Б. М. Лобанов, Л. И. Цирульник // Белорусская Наука, Минск. — 2008
7. Lobanov B. Universal Melodic Portraits of Intonation Patterns in Russian Speech/Lobanov B., Okrut T. // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2014», — Bekasovo, — P. 330-339ниже:

References

1. Bryzgunova E. A. Zvuki i intonaciya russkoj rechi / — Nauka, M., — 1968.
2. Bryzgunova E. A. Intonaciya // Russkaya grammatika / — Nauka, M., — 1980.
3. Odincova I. V. Zvuki. Ritmika. Intonaciya. M.: Flinta-Nauka, — 2011.
4. Lobanov B. The Phonemophon text-to-speech system // International Congress of Phonetic Sciences: proc. of the 11-th seccion ICPhS'87, — Tallin, — 1987 — P. 120-124.

Лобанов Б.М., Житко В.А.

«IntonTrainer-ru» – программный комплекс для самообучения и тренинга нормативной интонации русской речи

5. Lobanov B. Language- and Speaker Specific Implementation of Intonation Contours in Multilingual TTS Synthesis/Lobanov B., Tsirulnik L., Zhadinets D., Karnevskaia E. // *Speech Prosody: proceedings of the 3rd International conference. Dresden, Germany* — Vol. 2. — 2006 — P. 553-556.
6. Lobanov B. M. Komp'yuternyj sintez i klonirovanie rechi/B. M. Lobanov, L. I. Cirul'nik // *Belorusskaya Nauka, Minsk*. — 2008
7. Lobanov B. Universal Melodic Portraits of Intonation Patterns in Russian Speech/Lobanov B., Okrut T. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2014»*, — Bekasovo, — P. 330-339. Bekasovo, — P. 330-339.

«INTONTRAINER-RU» — SOFTWARE COMPLEX FOR SELF-EDUCATION AND TRAINING OF NORMATIVE INTONATION OF THE RUSSIAN SPEECH

Lobanov B. M., *Chief Researcher, UIIP NAS of Belarus, Minsk, Belarus, lobbormef@gmail.com*

Zhitko V. A., *PhD student, UIIP NAS of Belarus, Minsk, Belarus, zhitko.vladimir@gmail.com*

The article considers the theoretical and experimental foundations for constructing systems for the operational assessment of the intonation characteristics of Russian speech. The description of the developed computer system, its technical characteristics, areas of application and experience of use are given.

• *Intonation • Russian speech • self-study • training*