

КОМПЬЮТЕРНЫЕ СИСТЕМЫ ОПЕРАТИВНОЙ ОЦЕНКИ ПАРАМЕТРОВ РЕЧЕВОЙ ИНТОНАЦИИ



Борис Лобанов,
главный научный сотрудник
Объединенного института проблем информатики НАН Беларуси, доктор технических наук



Владимир Житко,
аспирант
Объединенного института проблем информатики НАН Беларуси

Интонация – самая эфемерная составляющая устной речи. На письме она передается весьма условно. По мнению директора Института лингвистики Российского государственного гуманитарного университета Максима Кронгуза, это, может быть, самая загадочная область фонетики, исследования которой только-только начинаются. Владение этим аппаратом – почти всегда интуитивное – очень помогает человеку в коммуникации.

Интонация практически позволяет 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». С другой стороны, нарушения в этой области могут привести к серьезным смысловым изменениям, а также создать неверное впечатление о личности говорящего. 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». Отметим, что особенно важна интонация при изучении китайского и других тональных языков.

Единство взаимосвязанных компонентов устной речи – это и есть интонация. Из всех ее акустических коррелятов наиболее универсальными и значимыми являются частотно-временные характеристики речевого сигнала, а именно – мелодика и темп речи.

Мелодика речи – это текущее изменение частоты основного тона (ЧОТ) голоса, характеризующееся диапазоном, регистром и формой

кривой ЧОТ. Она выполняет следующие функции: различает коммуникативные типы высказывания (побуждение, вопрос, восклицание и др.), оформляет его в единое целое, одновременно расчленяя на ритмические группы; выражает конкретные эмоции; характеризует язык и индивидуальность говорящего, а также ситуацию общения.

Темп речи – это текущее изменение длительности ее фонетических элементов и межфразовых пауз. С его помощью выражается степень важности содержания (более важное – медленным темпом, менее – быстрым); привлекается внимание слушающего (посредством замедления или ускорения темпа); отображается эмоциональное состояние (положительные эмоции ускоряют темп, отрицательные – тормозят его).

Системы оперативной оценки параметров мелодики и темпа речи являются слабо разработанной отраслью современных речевых технологий, однако они могут быть весьма востребованными во множестве практических приложений. Авторами разработаны два программных комплекса в данной области – «IntonTrainer» и «Speech Rate Meter». В основе разработок лежат научные результаты авторов, изложенные в [1–3].

ПРОГРАММНЫЙ КОМПЛЕКС «INTONTRAINER»

Комплекс предназначен для компьютерного анализа интонации и обеспечивает визуальное отображение фразовых контуров ЧОТ и количественную оценку правильности интонирования речи на родном или иностранном языках [2]. В состав программного комплекса входят подсистемы, включающие наборы эталонных фраз, которые представляют основные интонационные модели русской, английской (британский и американский варианты), немецкой и китайской речи. В процессе обучения «IntonTrainer» сравнивает и дает численную оценку интонационного сходства произнесенной и эталонной фраз.

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

- адекватное сопоставление произносимого (с заранее не известным началом и окончанием) и эталонного сигналов при условии наличия взаимной временной деформации;
- сегментация анализируемого сигнала на участки, для которых понятие ЧОТ имеет смысл с точки зрения формирования инто-

национного 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 40 до 1000 Гц);
- интерполяция значений ЧОТ на тех участках, для которых определение является некорректным (большинство согласных звуков);
- вычисление меры подобия (сходства) интонационных контуров произносимой и эталонной фраз в условиях их различной длительности и различных диапазонов ЧОТ.

Оценка интонационного сходства производится на основе представления тонального контура в виде универсального (унифицированного) мелодического портрета (УМП) [3], а именно в виде кривой на плоскости с нормированными координатными осями: время – TN и ЧОТ – F0. При этом дискретным интервалам на оси абсцисс соответствуют: [0–1/3] – пред-ядру, [1/3–2/3] – ядру, [2/3–1] – за-ядру. Интервалам на оси ординат соответствуют: [0–1/3] – низкому тону, [1/3–2/3] – среднему, [2/3–1] – высокому. При таком представлении основное внимание уделяется особенностям формы кривой ЧОТ на ядре при меньшем внимании на количественные и качественные составы пред- и за-ядра. В качестве критериев сходства интонации произнесенной фразы с эталоном выступают степень их близости по диапазону изменения частоты основного тона (F0) и по форме кривой, отображаемой в виде УМП.

Система «IntonTrainer» наглядно отображает на экране УМП фразы и критерии их сходства. Красный столбец слева показывает диапа-

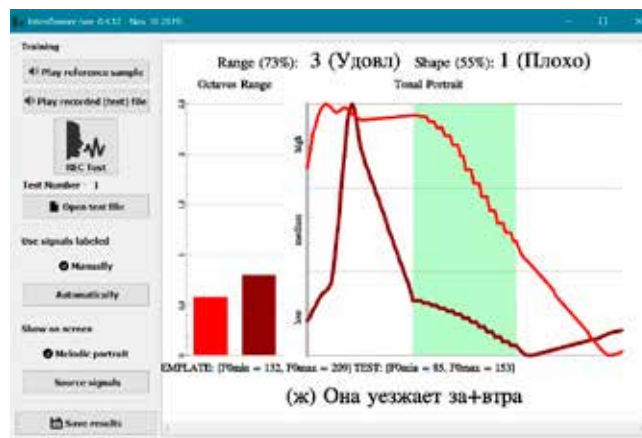


Рис. 1. Результат сравнения интонации фразы «Она уезжает завтра» (эталонной и произнесенной)



Рис. 2. Стартовое окно программы



Рис. 3. Модуль запуска прикладных задач (Launcher)

зон изменения F0 эталонной фразы «Она уезжает завтра» (рис. 1), а коричневый – произнесенной, справа – их линейный график УМП, внизу – минимальное и максимальное значения F0 эталонной (Template) и произнесенной (User) фраз. Оценка интонационного сходства синтезированной и естественной речи осуществляется по двум основным критериям: степени сходства диапазонов изменения кривых F0 и УМП. Вверху над графиками приведены значения для мер сходства по диапазону – Range 57% и по форме кривой мелодического портрета – Share 64%.

Версия программного комплекса, представленная на сайте, реализована для нескольких языков и с учетом решения различного рода прикладных задач, в том числе, для анализа эмоциональной речи и пения. (рис. 2, 3).

Программный комплекс «IntonTrainer» является открытой системой. Допускается модификация используемых настроек и аудиоматериалов. Эталонные базы данных – PATTERNS – могут быть дополнены или модифицированы в соответствии с поставленной задачей либо сформированы заново для работы с новыми языковыми приложениями. Подробное описание работы с программой приведено в текстовом файле «User Guide» на веб-сайте <https://intontrainer.by>.

Программный комплекс «IntonTrainer» рекомендуется использовать:

- как средство визуализации интонации для первичного ознакомления с основными интонационными конструкциями речи, используемыми в различных языках;
- для самостоятельного тренинга и самоконтроля при изучении иностранных языков и при

совершенствовании интонационных навыков родной речи и пения;

- в научных исследованиях при изучении индивидуальных, эмоциональных и стилистических особенностей реализации интонации, для сравнительной оценки интонации речи и пения в норме и патологии, интонационного качества синтезированной речи.

ПРОГРАММНЫЙ КОМПЛЕКС «SPEECH RATE METER» (SRM)

Комплекс предназначен для численной оценки и экранного отображения просодических параметров темпа устной речи. Под темпом речи понимается скорость ее производства и обычно она измеряется числом слов, произносимых в 1 минуту (в англоязычной литературе обозначается как: wpm [words per minute]).

Темп речи мы предлагаем описывать как комплексный просодический компонент интонации, параметрами которого являются:

- *полный темп, или скорость чтения текста* (подсчитывается как количество слов в минуту за фактическое время произнесения с учетом длительности межфразовых пауз);
- *артикуляторный темп, или скорость речедвижений* (количество слов в минуту за суммарное время артикуляции слов, без учета длительности межфразовых пауз);
- *показатель паузации* (среднеквадратичное значение длительности паузальных участков – межсинтагменных или межфразовых пауз).

Для успешного расчета статистических характеристик темпа речи необходимо осуществить предварительную обработку речевого сигнала, состоящую из следующих этапов:

- *обработка с помощью фильтра, частотная характеристика которого согласована со среднестатистическими спектрами гласных и согласных звуков;*
- *расчет кривой интенсивности звуков речи $E(n)$ при сглаживании возможных колебаний кривой интенсивности с частоты основного тона – F0 и сохранении динамики изменения интенсивности звуков речи;*
- *вычисление сглаженной кривой интенсивности $E(m)$ путем дополнительной обработки кривой $E(n)$, полученной на предшествующем этапе;*

- *определение границ сегментов гласных звуков путем сравнения двух кривых – $E(n)$ и $E(m)$, точки пересечения которых показывают границы гласных звуков;*
- *вычисление статистических характеристик длительности сегментов, необходимых для оценки параметров темпа речи. Необходимым и достаточно эффективным оказался следующий набор характеристик:*
 1. N_s – количество сегментов согласных и пауз;
 2. T_s – длительность анализируемого отрезка речи;
 3. T_v – суммарная длительность гласных сегментов;
 4. T_{ca} – среднеквадратичная длительность согласных и пауз;
 5. T_{cm} – среднемедианная длительность согласных и пауз.

Полученные значения используются затем для расчета комплекса численных оценок параметров темпа анализируемой речи, а именно:

- R_s (wpm) – *полный темп – Speech Rate;*
 - R_a (wpm) – *артикуляторный темп – Articulation Rate;*
 - T_p (sec) – *показатель паузации – Phrase Pauses.*
- Сразу после запуска программы (рис. 4) пользователю предоставляются следующие возможности:
- *запись через микрофон достаточно длительного отрезка анализируемой речи (желательно не менее 10 сек);*
 - *вызов одного из заранее подготовленных тестовых аудиофайлов, хранящихся в папке «data-tests».*

После того как речевой сигнал введен, рассчитываются параметры темпа и открывается

главное информационное окно SRM с результатами (рис. 5).

В верхней части окна отображаются численные значения темповых параметров: Speech Rate (wpm) и Articulation Rate (wpm), а в нижней – Phrase Pauses (sec), а также длительность анализируемой речи – Speech Duration (sec). В средней части окна в виде двух подвижных дуг синего и зеленого цветов отображаются Speech Rate и Articulation Rate. Их длина нормирована относительно задаваемых пользователем минимального (R_{min}) и максимального (R_{max}) значений параметров темпа. Выбор этих значений определяется конкретными условиями применения SRM, при которых темп речи будет считаться медленным (Slow), средним (Average) или быстрым (Fast), что показано («wpm») на концах полуокружности.

Пользователю, исходя из его собственных научных или практических интересов, предоставляется возможность записи неограниченного количества речевых аудиофайлов от различных источников. Содержимое папки «tests» можно пополнять и озвучивать с помощью стандартных средств.

Подробное описание работы с программой приведено в текстовом файле «User Guide» на веб-сайте <https://intontrainer.by>.

Программный комплекс «Speech Rate Meter» рекомендуется использовать:

- *при подготовке выступлений и презентаций для самоконтроля и обучения желаемому темпу речи;*
- *для совершенствования навыков свободного чтения на родном и иностранном языках;*

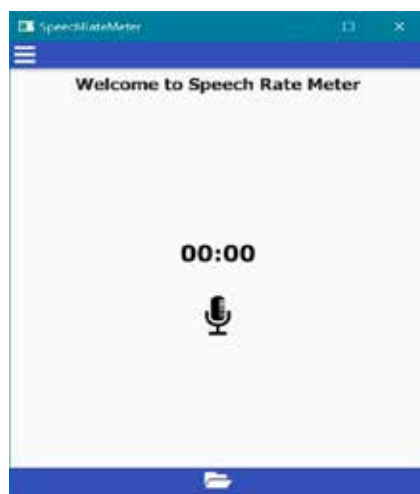


Рис. 4. Стартовое окно SRM, открываемое после запуска программы

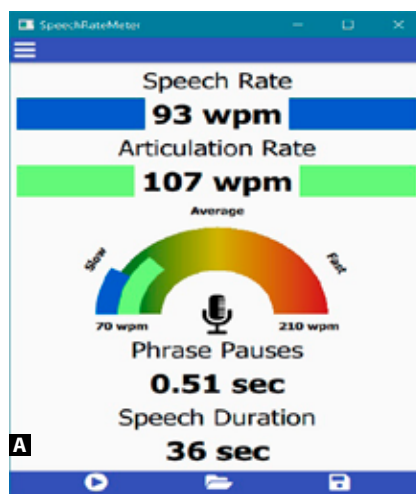


Рис. 5. Главное информационное окно SRM: А) медленный темп, Б) быстрый темп

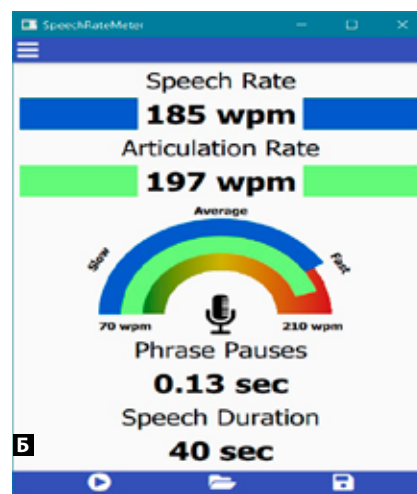




Рис. 6. Все страны на карте мира



Рис. 7. Страны Европы



Рис. 8. Страны Северной и Центральной Америки



Рис. 9. Страны Африки, Азии, Австралии

- в качестве средства самоконтроля для устранения дефектов речи, связанных с различными физиологическими нарушениями и болезнями;
- как средство мониторинга темпа речи в реальных условиях работы профессиональных операторов колл-центров, дикторов радио и телевидения.

Версии описанных компьютерных систем «IntonTrainer» и «Speech Rate Meter» для OS Windows размещены на сайте <https://intontrainer.by>. Обе программы общедоступны для бесплатной загрузки и тестирования. С использованием <https://datastudio.google.com/navigation/reporting> проведен анализ активности обращений пользователей к нашему сайту. К настоящему времени (начало мая 2021 г.) зафиксировано 3405 обращений к сайту от пользователей из 91 страны. На экранных копиях (рис. 6–9) наглядно представлена география обращений пользователей из населенных пунктов различных стран, полученная с помощью данных сайта <https://www.google.com/maps/d/u/0/viewer?mid=18QwNw88mRMaRnPC-EieTd7EhOzhxGZe>. Эти данные достаточно убедительно свидетельствуют о полезности разработанных программных продуктов.

Программные системы «IntonTrainer» и «Speech Rate Meter» для OS Windows, размещенные на сайте <https://intontrainer.by>, выпущены под разрешительной лицензией MIT, что позволяет программистам использовать лицензируемый код в закрытом программном обеспечении при условии, что текст лицензии предоставляется вместе с ним. Закончена разработка мобильной версии системы «Speech Rate Meter» в OS Android, которая доступна в Google Play (<https://play.google.com/store/apps/details?id=by.intoncore.SpeechRateMeter>). ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Лобанов Б. М. Компьютерный синтез и клонирование речи / Б. М. Лобанов, Л. И. Цирульник. – Минск, 2008.
2. Лобанов Б. М. Компьютерная система обучения интонационным конструкциям русской речи / Б. М. Лобанов, В. А. Житко, А. А. Харламов // Компьютерная лингвистика и интеллектуальные технологии: материалы Междунар. Конф. «Диалог-2017», Москва, 31 мая – 3 июня 2017 г.). Вып. 16 (23). В 2 т. – М., 2017. С. 287–302.
3. Лобанов Б. М. Универсальные мелодические портреты интонационных конструкций русской речи / Б. М. Лобанов, Т. И. Окрут // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва – Бекасово, 4–8 июня 2014 г.). Вып. 13 (20). – М., 2014. С. 330–339.