

Компьютерная лингвистика и интеллектуальные технологии:  
по материалам международной конференции «Диалог 2019»

Москва, 29 мая — 1 июня 2019 г.

## **АНАЛИЗ ПРОСОДИЧЕСКИХ ПРИЗНАКОВ ЭМОЦИОНАЛЬНОЙ ИНТОНАЦИИ С ИСПОЛЬЗОВАНИЕМ СИСТЕМЫ «INTONTRAINER» (на примере русскоязычных фраз)**

**Лобанов Б. М.** (Lobanov@newman.bas-net.by),  
**Житко В. А.** (zhitko.vladimir@gmail.com)

Объединённый институт проблем информатики  
НАН Беларуси, Минск, Беларусь

## **ANALYSIS OF PROSODIC FEATURES OF THE EMOTIONAL INTONATION USING “INTONTRAINER” SYSTEM (on the example of Russian phrases)**

**Lobanov B. M.** (Lobanov@newman.bas-net.by),  
**Zhitko V. A.** (zhitko.vladimir@gmail.com)

United Institute of Informatics Problems NAS Belarus,  
Minsk, Belarus

The main results of the update of the IntonTrainer system for the purposes of analyzing and studying the prosodic signs of emotional intonation are described. A distinctive functional feature of the updated system is the creation of an expanded set of prosodic signs of emotional intonation. The paper presents preliminary assessments of their effectiveness using the created experimental database of emotional phrases of Russian speech.

**Keywords:** Speech intonation, basic emotions, emotional intonation, melodic portrait, intonation analysis, software model

## 1. Введение

Хорошо известно, что человеческая речь передает не только смысловую, но и эмоциональную информацию. В теории эмоций многочисленные эмоциональные состояния часто отображаются в двухмерном пространстве: «приятное — не приятное», «активное — пассивное» [1]. Существует множество различных дискретных наборов эмоций. Однако, большинство исследований ограничиваются анализом просодических характеристик следующих 6-ти эмоциональных состояний.

*Нейтральность* (спокойствие, сдержанность ...) — уравновешенное состояние ума, никаких беспокойств, сомнений, волнений, забот.

*Радость* (восторг, упоение...) — положительное эмоциональное состояние, связанное с возможностью достаточно полного удовлетворения фактической потребности.

*Грусть* (тоска, уныние...) — негативное эмоциональное состояние, связанное с полученной информацией о невозможности удовлетворения важных жизненных потребностей

*Гнев* (возмущение, ярость...) — эмоциональное состояние, отрицательное по признаку, возникающее в форме аффекта и вызванное внезапным появлением серьезного препятствия.

*Страх* (испуг, тревога...) — негативное эмоциональное состояние, которое возникает, когда субъект получает информацию о реальной или воображаемой опасности.

*Удивление* (изумление, ошеломление...) — эмоциональная реакция на неожиданные обстоятельства.

Имеется также ряд эмоций, относимых довольно часто к основным, такие как: *Страдание, Отвращение, Презрение, Стыд*, а кроме того, многочисленные оттенки перечисленных выше эмоций.

До настоящего времени не имеется достаточных знаний о деталях акустических моделей, которые описывают определенные эмоции человеческого голоса. Типичные акустические характеристики, которые, как считается, вовлечены в этот процесс, включают следующее [2]–[4]:

- уровень, диапазон и форма контура основной частоты (F0);
- уровень вокальной энергии голоса;
- темп речи.

В последнее время исследованы новые важные речевые характеристики, которые содержат информацию о эмоциях, такие как частоты формант, коэффициенты линейного прогнозирования и коэффициенты мел-частотного кепстра [5]–[7]. Во многих работах особое место отводится анализу поведения мелодической кривой. В одной из последних работ, посвящённых анализу просодических характеристик эмоций [8], предложено использовать следующее описание контура высоты тона:

- Количество максимумов в контуре основного тона в вокализованном сегменте;
- Среднее значение и дисперсия пиковых значений;
- Средний наклон;
- Средний градиент между двумя точками выборки на кривой основного тона;
- Дисперсия градиентов основного тона.

Вопросам анализа и сопоставления контуров высоты тона различных интонационных конструкций посвящена также работа [9]. С начала 2018 года на веб-сайте (см. <https://intontrainer.by>) выложена демо-версия программной системы «IntonTrainer». Она ориентирована на использование её в качестве компьютерного средства обучения интонации устной речи. 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» осуществляет сравнение и оценку интонационного сходства произнесённой и эталонной фраз. Оценка интонационного сходства производится на основе представления тонального контура в виде универсального (унифицированного) мелодического портрета (УМП) [9]. Представляет интерес соответствующая доработка и модернизация системы «IntonTrainer» для целей анализа и исследования просодических признаков эмоциональной интонации.

## 2. Основные направления и результаты модернизации системы «IntonTrainer»

В задачу модернизации существующей системы входит создание такого программного средства, которое обеспечивает анализ и визуализацию эффективного набора просодических признаков эмоциональной интонации, а также даёт возможность предварительной оценки их эффективности с использованием доступных БД эмоциональной речи. Для того чтобы модернизированная система позволяла достаточно эффективно анализировать и визуализировать эмоциональные признаки интонации, в систему добавлен ряд новых дополнительных функций.

На **рис. 1** показан вид начального окна после загрузки системы.

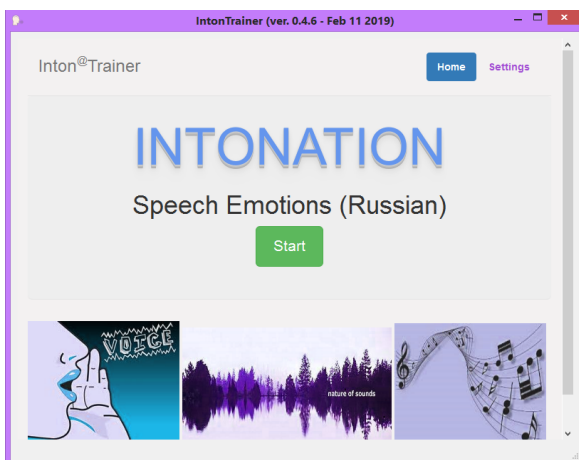


Рис. 1. Начальное окно

После нажатия кнопки «**Start**» открывается главное окно Программы, содержащее структурированный перечень эталонных фраз с указанием наименования и номера БД, имени диктора, названия эмоции и текста фразы, в которой она отражена (рис. 2).

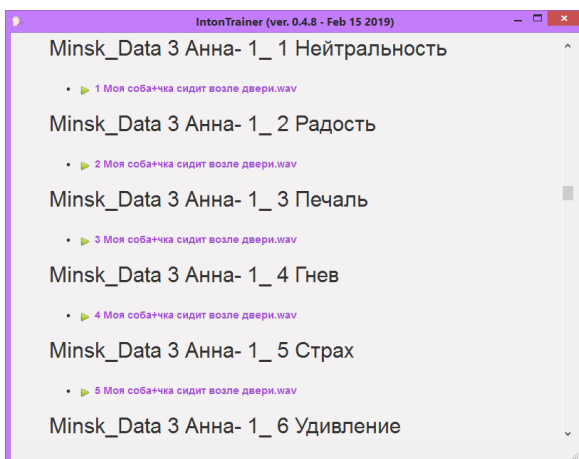


Рис. 2. Главное окно

Путём выбора с помощью курсора требуемой директории, например:

«Minsk\_Data 3 Анна-1\_1 Нейтральность 1 *Моя соба+чка сидит возле двери*»

открывается окно, в котором отображаются в графическом виде результаты интонационного анализа этой фразы (рис. 3)

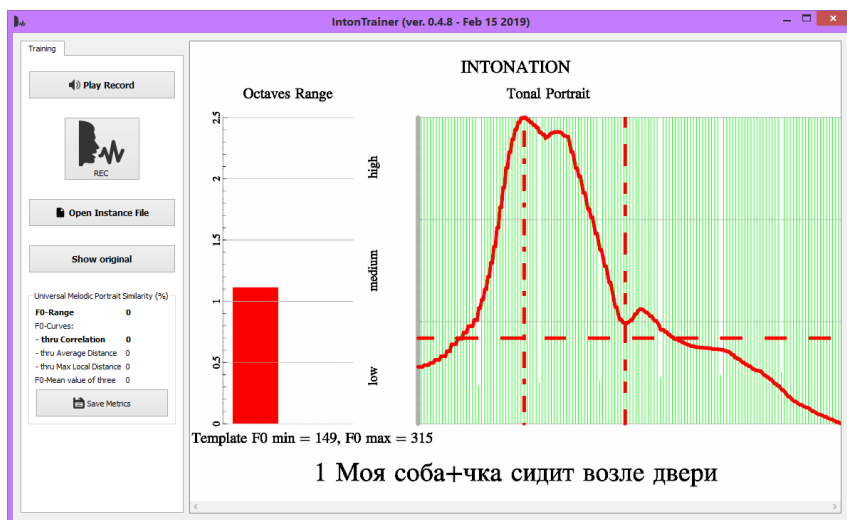



Рис. 3. Окно отображения кривой НМП: диктор «Анна-1» (Нейтральность)

Непрерывная кривая отображает траекторию изменения ЧОТ на голосовых участках фразы — нормированный мелодический портрет (НМП). Построение кривой НМП в отличие от универсального мелодического портрета УМП не требует «ручной» разметки фразы на участки пред-ядра, ядра и за-ядра. При выборе в разделе **Setting** режима **Auto Marking** сегментация сигнала на голосовые регионы осуществляется автоматически на основе информации о наличии периодичности в сигнале (голоса) при одновременном присутствии достаточно высокой амплитуды сигнала —  $A_0(t)$ .

Горизонтальная штриховая линия на кривой НМП показывает среднее значение нормированной ЧОТ. Две вертикальные линии характеризуют положение центра кривой и её ширину (размытость) на нормированной временной оси. Высота столбика (слева от НМП) показывает диапазон изменения ЧОТ в октавах.

В левой части на рис. 3 показаны кнопки управления, с помощью которых доступно осуществление следующих функций:

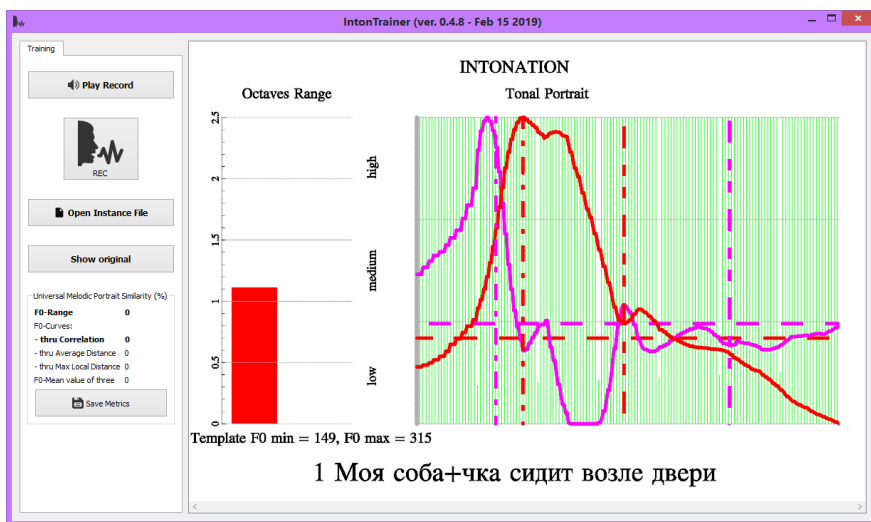
- «Play Record» — прослушивание перечня эталонных фраз.
- «Rec» — оперативная запись фраз пользователя через микрофон,
- «Open Instance File» — вызов тестовых фраз из папки «TEST»,
- «Show original» — просмотр исходных сигналов,
- «Save Metrics» — сохранение данных об измеренных просодических признаках.

При нажатии кнопки «Save Metrics» появляется дополнительный значок  и открывается страница в EXCEL, на которой записывается полный набор из 10-ти просодических признаков эталонной фразы (см. табл. 1). Полученные данные сохраняются в той же папке, где хранится исследуемая эталонная фраза.

**Таблица 1.** Результаты вычисления просодических признаков: Анна-1 (Нейтральность)

№	Названия просодических признаков	Names of Prosodic Features	Results
1	Диапазон изменения ЧОТ [F0max/F0min]	Pitch Range F0	2,11
2	Регистр ЧОТ [(F0max + F0min)/2]	Register F0 [Hz]	232,00
3	Среднее значение кривой НМП	Mean Value of the curve NMP	27,97
4	Положение центра кривой НМП	Center of the curve NMP	36,99
5	Эффективная ширина кривой НМП	Width of the curve NMP	23,89
6	Среднее значение кривой d/dt (НМП)	Mean Value of the Derivative curve NMP	32,68
7	Положение центра кривой d/dt (НМП)	Center rangeer of the Derivative curve NMP	46,26
8	Ширина кривой d/dt(НМП)	Width of the Derivative curve NMP	55,24
9	Средний уровень звонких звуков	Voiced Sounds Level	0,15
10	Суммарная длительность звуков	Voiced Sounds Duration	239,00

В таблице 1 кроме данных о параметрах исходной кривой НМП, представлены также данные о величине её производной по времени — d/dt (НМП). Сравнительный вид этих кривых представлен на **рис. 4**. Дополнительный анализ параметров производной от НМП оказывается полезным для учёта динамических характеристик движения ЧОТ, характерных для некоторых видов эмоций.



**Рис. 4.** Пример отображения кривых НМП и d/dt (НМП): Диктор «Анна-1» (Нейтральность)

При нажатии кнопки «Open Instance File» осуществляется вызов тестовых фраз из папки «TEST», в которой для сравнительного отображения различных эмоций могут быть помещены те же аудио файлы, что и в папке PATTERNS. На рис. 5 представлен пример сравнения фразы «Моя собачка сидит возле двери» с эмоцией «Нейтральность» (светлая линия) с той же фразой с эмоцией «Гнев» (тёмная линия). Как видно из рисунка, отличие этих эмоций наблюдается в форме кривых НМП и в их параметрах, таких как средние значения, центры кривых и их ширины.

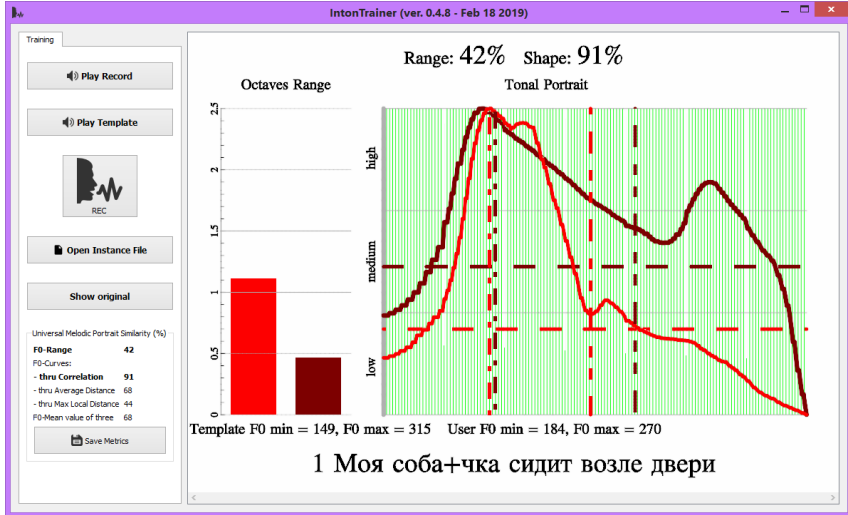


Рис. 5. Пример сравнения кривых НМП:  
Диктор «Анна-1» (Гнев/Нейтральность)

Детальную информацию о результатах сравнения просодических признаков этих фраз можно получить при нажатии кнопки «Save Metrics» (см. таблицы 2, 3, 4).

В таблице 2 приведен пример результатов вычисления численных значений просодических признаков фразы, произнесённой с выражением эмоции «Гнев».

Таблица 2. Результаты вычисления  
просодических признаков: Анна (Гнев)

№	Названия просодических признаков	Names of Prosodic Features	Results
1	Диапазон изменения ЧОТ [F0max/F0min]	Pitch Range F0	1,45
2	Регистр ЧОТ [(F0max + F0min)/2]	Register F0 [Hz]	228,00
3	Среднее значение кривой НМП	Mean Value of the curve NMP	54,47

№	Названия просодических признаков	Names of Prosodic Features	Results
4	Положение центра кривой НМП	Center of the curve NMP	46,00
5	Эффективная ширина кривой НМП	Width of the curve NMP	37,40
6	Среднее значение кривой d/dt (НМП)	Mean Value of the Derivative curve NMP	49,28
7	Положение центра кривой d/dt (НМП)	Center of the Derivative curve NMP	44,31
8	Ширина кривой d/dt (НМП)	Width of the Derivative curve NMP	50,73
9	Средний уровень звонких звуков	Voiced Sounds Level	0,19
10	Суммарная длительность звуков	Voiced Sounds Duration	226,00

В таблице 3 приведен пример результатов вычисления по данным, приведенным в таблицах 1 и 2, относительных значений для просодических признаков пары фраз с эмоциями «Гнев/Нейтральность», выраженных в децибелах. Использование относительных величин позволяет осуществлять сравнение пары фраз с различными эмоциями, используя просодические признаки различной природы с оценкой в различных единицах измерения.

**Таблица 3.** Численные значения относительных признаков: Анна (Гнев/Нейтральность)

№	Названия просодических признаков	Names of Prosodic Features	Relations [dB]
1	Диапазон изменения ЧОТ [F0max/F0min]	Pitch Range F0	-0,67
2	Регистр ЧОТ [(F0max + F0min)/2]	Register F0 [Hz]	0,92
3	Среднее значение кривой НМП	Mean Value of the curve NMP	-1,59
4	Положение центра кривой НМП	Center of the curve NMP	-0,09
5	Эффективная ширина кривой НМП	Width of the curve NMP	2,38
6	Среднее значение производной кривой НМП	Mean Value of the Derivative curve NMP	0,65
7	Положение центра производной кривой НМП	Center of the Derivative curve NMP	1,41
8	Эффективная ширина производной кривой НМП	Width of the Derivative curve NMP	1,55
9	Средний уровень звонких звуков	Mean Value of Voiced Sounds Level	-0,07
10	Общая длительность звуков фразы	Total duration of phrase sounds	-0,26

В таблице 4 приведен также пример результатов вычисления численных мер сходства и расстояний между 2-мя предъявленными реализациями эмоциональных фраз (в данном случае пара: (Нейтральность — Гнев). Описание способов вычисления и соответствующие им формулы приведены в [9].



**Таблица 4.** Результаты вычисления мер сходства и расстояний: Анна (*Гнев/Нейтральность*)

№	Способ сравнения сходства	Type of the proximity	Proximity	Distance
1	Коэффициент взаимной корреляции	Cross correlation coefficient	91	9
2	Интегральное сравнение кривых НМП	Integral comparison of NMP curves	68	32
3	Локальное сравнение кривых НМП	Local comparison of NMP curves	44	56
4	Среднее значение 3-х способов сравнения	Average of the three above proximities	68	32
5	Сравнение диапазонов изменения ЧОТ	Comparison of pitch ranges	42	58

### 3. Экспериментальная оценка разработанного набора просодических признаков

Для экспериментальной оценки эффективности разработанного набора просодических признаков эмоциональной интонации на примере русскоязычных фраз совершенно необходима соответствующая БД, подобная, например, англоязычной БД эмоциональных фраз [10], которая доступна для бесплатного скачивания. Существуют по крайней мере одна работа [11], в которой указывается на существование русскоязычной БД эмоциональных фраз, однако условия доступа к ней нам не известны. По этой причине для целей экспериментальной оценки разработанного набора просодических признаков мы решили создать собственную экспериментальную БД небольшого объема по следующей методике.

Созданы специальные текстовые сценарии, провоцирующие диктора на выражение одной из 6-ти видов эмоций (*Нейтральность, Радость, Грусть, Гнев, Страх, Удивление*) при произнесении фразы «Моя **собачка** сидит возле двери» с фразовым акцентом на втором слове. Выбранная фраза является вольным переводом с английского фразы, используемой для тестирования в англоязычной БД эмоциональных фраз [10].

**Таблица 5.** Тексты сценариев, провоцирующих различные эмоции

№	Эмоция	1-й образец фразы	2-й образец фразы
1	Нейтраль	<i>Наконец я на даче. Кажется, я вижу её. Моя <b>собачка</b> сидит возле двери.</i>	<i>Да, действительно, это она: Моя <b>собачка</b> сидит возле двери.</i>
2	Радость	<i>Ура! Какое счастье! Она вернулась! Моя <b>собачка</b> сидит возле двери!</i>	<i>Маша, посмотри! Это же наш Шарик! Моя <b>собачка</b> сидит возле двери.</i>

№	Эмоция	1-й образец фразы	2-й образец фразы
3	Печаль	<i>Какая жалость... Её не взяли с собой... Моя <b>собачка</b> сидит возле двери...</i>	<i>Что же теперь делать... Бедная, бедная... Моя <b>собачка</b> сидит возле двери...</i>
4	Гнев	<i>Эй! Кто её выпустил!? Посмотри! Моя <b>собачка</b> сидит возле двери!!!</i>	<i>Ты что, не слышишь?! Моя <b>собачка</b> сидит возле двери!!!</i>
5	Страх	<i>Боже мой! Волки! Я боюсь! Она же щенок! Моя <b>собачка</b> сидит возле двери!</i>	<i>Волки уже близко! Что же делать?!! Моя <b>собачка</b> сидит возле двери!</i>
6	Удивлен	<i>Да, что ты говоришь? Это она? Моя <b>собачка</b> сидит возле двери!?</i>	<i>Неужели, правда? Неужели, она? Моя <b>собачка</b> сидит возле двери!?</i>

В качестве дикторов были привлечены 5 мужчин и 5 женщин из числа студентов театрального института и преподавателей русского языка для иностранцев. Перед ними была поставлена задача прочесть представленные в **таблице 5** тексты с максимально полной имитацией эмоциональных состояний, подсказываемых контекстом. В результате получены аудио записи фразы «*Моя собачка сидит возле двери*» по 2 варианта для каждой из 6 эмоций. Затем все записи в случайном порядке были представлены 3-м аудиторам, в задачу которых входило распознавание одной из 6 эмоций в предъявляемых образцах.

По результатам аудирования были отобраны наилучшие результаты: с женским голосом — «Анна» и с мужским — «Борис», которые использовались при проведении экспресс-исследования эффективности анализа просодических признаков эмоциональной интонации модернизированной системой «IntonTrainer». Однако, даже только для двух дикторов, полный анализ полученных данных и их графическое представление в данной работе представляется не реальным. Как видно из **таблиц 1–4**, разработанная система анализа просодических признаков генерирует для каждой пары эмоций 35 количественных показателей. Анализ же всевозможных пар становится возможным только с привлечением специальных программ обработки больших данных, например, с использованием нейросетевых алгоритмов.

Тем ни менее, для достаточно наглядной оценки качества предложенного набора просодических признаков эмоций мы посчитали возможным ограничиться здесь нижеследующим набором. На **рисунках 6–10** представлены графические диаграммы нормированных значений различных пар признаков, рассчитанных и усреднённых по двум реализациям: «Борис-1» и «Борис-2».

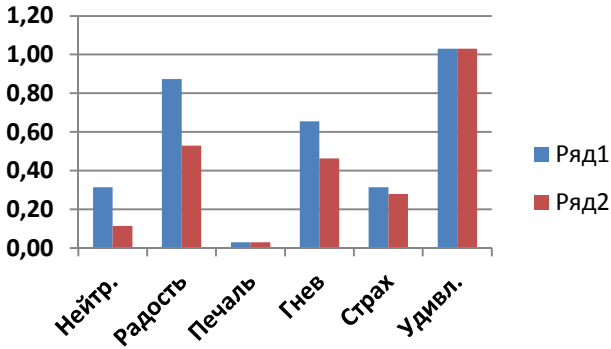


Рис. 6. Диапазон изменения ЧОТ (ряд 1), Регистр ЧОТ (ряд 2)

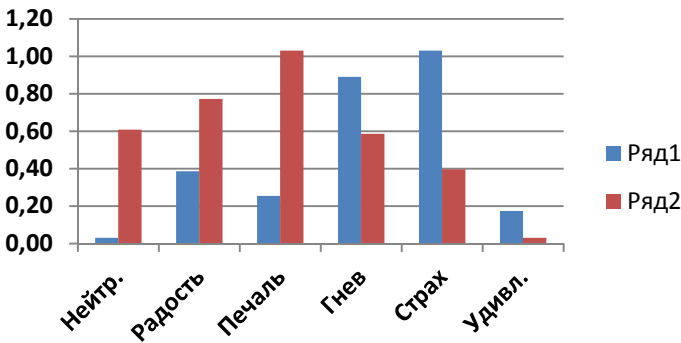


Рис. 7. Среднее значение кривой НМП (ряд 1) и её производной (ряд 2)

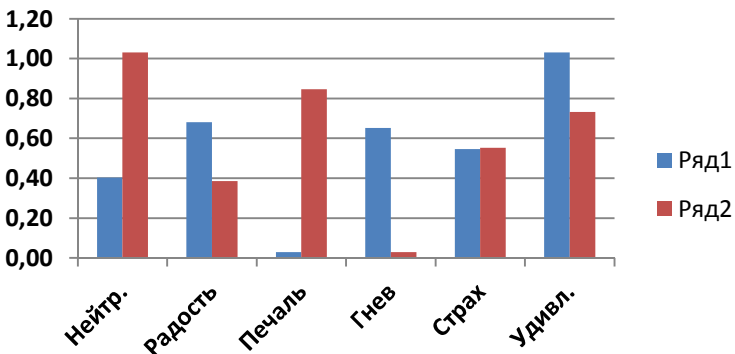


Рис. 8. Эффективная ширина кривой НМП (ряд 1) и её производной (ряд 2)

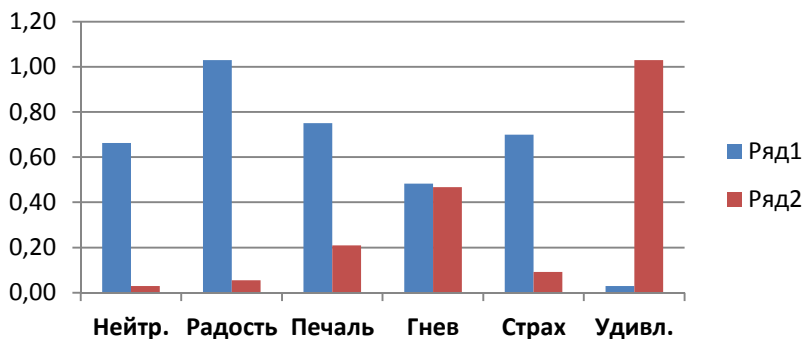


Рис. 9. Положение центра кривой НМП (ряд 1) и её производной (ряд 2)

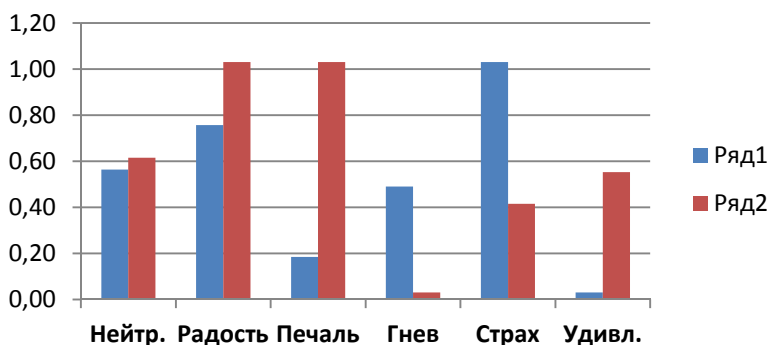


Рис. 10. Средний уровень звонких звуков МП (ряд 1) и их длительность (ряд 2)

Как видно из **рисунков 6–10**, полученные пары значений 10-ти просодических признаков сравнительно слабо коррелированы и характеризуются значительными различиями для каждого из 6-ти видов эмоций. Отметим некоторые очевидные результаты сопоставления просодических признаков эмоций. Из **рис. 6** видно, что эмоции «Радость», «Гнев» и «Удивление» характеризуются расширенным диапазоном изменения и высоким регистром ЧОТ. В то же время наименьшие значения этих признаков характерны для эмоций: «Печаль», «Нейтральность» и «Страх», что вполне соответствует нашим интуитивным представлениям. Менее очевидные результаты отражены на **рис. 7–10**. Их статистическая достоверность может быть проверена лишь после проведения более масштабных экспериментов.

## 4. Заключение

В настоящей работе описаны основные результаты модернизации системы «IntonTrainer» для целей анализа и исследования просодических признаков эмоциональной интонации. Отличительной функциональной особенностью обновлённой версии системы является реализация возможности расчёта численных значений расширенного набора просодических признаков, а также их сохранения в формате EXCEL таблиц. Модернизированная система установлена на сайте <https://intontrainer.by> под именем «*Russian Emotions Inton-Trainer*» и доступна для бесплатного скачивания. На сайте помещена также подробная инструкция для пользователя.

В задачу модернизации системы не входило создание действующей модели распознавания речевых эмоций. Конечная цель доработки ограничивалась созданием такого программного средства, которое бы обеспечивало анализ и визуализацию расширенного набора просодических признаков эмоциональной интонации, и которое могло бы быть использовано как новое инструментальное средство для фонетических исследований речи. Не исключены, мы полагаем, и некоторые прикладные аспекты применения системы, например, в задачах обучения требуемой эмоциональной интонации актёров, а также людей различных профессий, стремящихся к повышению своего, так называемого, «эмоционального интеллекта (EQ)». Этот новый термин появился сравнительно недавно и уже обсуждаются достоинства его применения в дополнение к IQ при оценке не только человека, но и системы искусственного интеллекта.

## Литература

1. Scherer, K. R., Schorr, A., Johnstone, T. (2001) *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press, New York and Oxford.
2. Banse, R., Scherer, K. R. (1996) Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70 (3), 614–636.
3. Banse, R., Scherer, K. R. (1996) Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70 (3), pp. 614–636.
4. Abelin, A., Allwood, J. (2000) Cross-linguistic interpretation of emotional prosody. In: *Proceedings of the ISCA Workshop on Speech and Emotion*.
5. D. Ververidis, C. Kotropoulos, and I. Pitas (2004) “Automatic emotional speech classification”, in *Proc. 2004 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol. 1, pp. 593–596, Montreal, May 2004.
6. Xiao, Z., E. Dellandrea, Dou W., Chen L. (2005) “Features extraction and selection for emotional speech classification”. *2005 IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 411–416, Sept 2005.
7. T.-L. Pao, Y.-T. Chen, J.-H. Yeh, P.-J. Li (2006) “Mandarin emotional speech recognition based on SVM and NN”, *Proceedings of the 18th International Conference on Pattern Recognition (ICPR’06)*, vol. 1, pp. 1096–1100, September 2006.

8. *C. Rinaldi, R. Tedesco, M. Matteucci, and A. Trivilini* (2014) Extracting Emotions and Communication Styles from Prosody. Springer-Verlag Berlin Heidelberg, pp. 21–42.
9. *Lobanov, B. A.* (2018) Prototype of the Software System for Study, Training and Analysis of Speech Intonation / B. Lobanov, V. Zhitko, V. Zahariev // *Speech and Computer: 20th International Conference, SPECOM 2018, Leipzig, Germany, September 18–22, 2018, Proceedings* / — Springer, 2018. — P. 337–346.
10. *Livingstone SR, Russo FA* (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE* 13(5): e0196391. <https://doi.org/10.1371/journal.pone.0196391>.
11. *Veronika Makarova and Valery A. Petrushin* (2002) RUSLANA: a database of Russian emotional utterances / 7th International Conference on Spoken Language Processing / Denver, Colorado, USA, September 16&20, 2002.