

## Method and Software Model for Evaluating the Statistical Characteristics of a Speech Melody

**Boris Lobanov**

United Institute of Informatics  
Problems NAS Belarus  
lobbormef@gmail.com

**Vladimir Zhitko**

United Institute of Informatics  
Problems NAS Belarus  
zhitko.vladimir@gmail.com

### Abstract

A method for estimating the statistical characteristics of speech melody is proposed. The procedure of constructing histograms of the frequency distribution of the pitch frequency over sufficiently long intervals of speech is described. A distinctive feature of the method is that the discrete values of the pitch of speech are measured only at intervals of vowels. Two options of the pitch scales used for analysis of the melody characteristics are selected, namely: linear for speech and logarithmic for singing. A method for estimating three parameters of the histogram is proposed: register - R, range - D, asymmetry - A. Numerous examples are given showing the effectiveness of the proposed method in assessing the individuality of the melody of the speaker's speech, as well as his emotional state. A description of the prototype of the Speech Melody Meter (SMM) system is given (see also: <https://intontrainer.by>). SMM is a software implementation of the proposed method for assessing the statistical characteristics of speech melody.

**Keywords:** speech prosody, speech melody, pitch frequency, vowels detection, software model.

**DOI:** 10.28995/2075-7182-2022-21-XX-XX

## Метод и программная модель оценки статистических характеристик мелодики речи

**Б.М. Лобанов**

Объединенный институт проблем  
информатики НАН Беларуси  
lobbormef@gmail.com

**В.А. Житко**

Объединенный институт проблем  
информатики НАН Беларуси  
zhitko.vladimir@gmail.com

### Аннотация

Предлагается метод оценки статистических характеристик мелодики речи. Описана процедура построения гистограмм распределения частот встречаемости дискретных значений частоты основного тона (ЧОТ) на достаточно длительных интервалах речи. Отличительной особенностью метода является то, что дискретные значения ЧОТ речи измеряются только на интервалах гласных звуков. Выбраны два варианта шкал ЧОТ, используемые при анализе характеристик мелодики, а именно: линейная для речи и логарифмическая для пения. Предложен метод оценки трех параметров гистограммы: регистр - R, диапазон - D, асимметрия - A. Приводятся многочисленные примеры, показывающие эффективность предложенного метода при оценке индивидуальности мелодики речи диктора, а также его эмоционального состояния. Дано описание прототипа системы Speech Melody Meter (SMM), являющейся программной реализацией предложенного метода оценки статистических характеристик мелодики речи (см. также: <https://intontrainer.by>).

**Ключевые слова:** просодия речи, мелодия речи, частота основного тона, программная модель.

## 1 Введение

Мелодика — основной компонент интонации речи человека. С акустической точки зрения, мелодика — это изменения во времени частоты основного тона (ЧОТ), измеряемой в герцах (Гц) или в музыкальных интервалах. Мелодика речи занимает важное место в структуре языка. Она участвует в реализации двух основных функций интонации — семантической и экспрессивной.

Семантическая функция мелодики выполняет следующие достаточно хорошо изученные задачи:

- организует фразу, расчлняя её на синтагмы и акцентные группы и связывая её части;
- различает коммуникативные типы высказывания (вопрос, побуждение, повествование и другие).

Семантическая функция реализуется путем специфических локальных изменений ЧОТ на отдельных элементах фразы (слог, акцентная группа, синтагма). Описанию закономерностей этих изменений для различных языков посвящены многочисленные работы, обзор которых не входит в задачу данной статьи.

Экспрессивная функция мелодики выполняет следующие задачи:

- выражает персональные особенности интонирования речи путем формирования специфического мелодического рисунка, обуславливающего своеобразность, стиль, выразительность и нормативность речи;
- определяет субъективное отношение к содержанию сообщения или к собеседнику путем выражения различного вида эмоций.

Необходимо отметить также, что степень владения навыками управления экспрессией речи является, пожалуй, одним из важных показателей эмоционального интеллекта человека — способностью управлять своими эмоциями и эмоциями других людей.

Экспрессивная функция мелодики, в отличие от семантической, реализуется главным образом за счет специфических интегральных характеристик траектории ЧОТ на достаточно длительном временном отрезке речи. Физические свойства мелодической кривой, обуславливающие особенности экспрессивной функции в сравнении с семантической функцией, исследованы в существенно меньшей степени. В определенной мере к такого рода исследованиям можно отнести результаты работ, направленных на решение ряда практически важных задач [(M. Farru's, 2009), (M. Leena, 2008), (C. Leyns, 2010), (D. Gharavian, 2010), (Cahill, 2016), (T. Banziger, 2005), (Peppé, 2009), (A. Norton, 2009), (Celine De Looze, 2017)], таких, как оценка:

- индивидуальных особенностей речи диктора [(M. Farru's, 2009), (M. Leena, 2008), (C. Leyns, 2010);
- эмоционального состояния говорящего [(D. Gharavian, 2010), (Cahill, 2016), (T. Banziger, 2005)];
- физиологических и патологических дефектов речи [(Peppé, 2009), (A. Norton, 2009), (Celine De Looze, 2017)].

В работе (M. Farru's, 2009) индивидуальные особенности речи диктора оцениваются путем измерения «дрожания и мерцания» ЧОТ и амплитуды. В работе (M. Leena, 2008) контур F0 представлен с помощью следующих оценок: значение максимума dF0, расстояние пика F0 относительно начала, наклон и продолжительность наклона кривой ЧОТ. В работе (C. Leyns, 2010) возраст и пол диктора определялся с помощью акустического анализа четырех интонационных параметров (общий интонационный сдвиг, общий диапазон, конечный интонационный сдвиг и показатель вариации ЧОТ).

В работе (D. Gharavian, 2010) показано, что в эмоциональных предложениях средние частоты основного тона и величина этого увеличения зависят от типов групп фонем и эмоций. Работа (Cahill, 2016) посвящена анализу особенностей реализации эмоций в африканском языке Knni. В работе (T. Banziger, 2005) описываются результаты, полученные для эмоциональных выражений актера. Показано, что средний уровень и диапазон F0 на контурах сильно различаются в зависимости от степени активации изображаемой эмоции. Напротив, было сравнительно мало свидетельств появления качественно разных форм контуров для разной эмоции.

В работе (Peppé, 2009) отмечается, что для болезни Паркинсона характерна низкая высота тона, низкая скорость и малый диапазон изменений ЧОТ. Научно обоснованный метод лечения – терапия с использованием мелодической интонации (МИТ) – описывается в работе (A. Norton, 2009). Метод использует интонирование при чтении и пении для улучшения выразительности речи людей с афазией. В работе (Celine De Looze, 2017) исследовано влия-

ние когнитивных нарушений на просодические параметры речеобразования при рассеянном склерозе. Замечено, что наклон высоты тона в высказывании тем круче, чем короче высказывание.

Как следует из обзора, в цитируемых работах по анализу индивидуальных особенностей, эмоционального состояния и патологических дефектов речи используются разнородные методики и инструменты анализа ЧОТ речевого сигнала. В данной работе предпринята попытка разработки единого метода и программной модели оценки статистических характеристик мелодики речи, пригодных для решения различных прикладных задач.

Настоящая работа базируется на результатах статьи (В. Lobanov, 2021), в которой описан метод расчета статистических оценок комплекса просодических характеристик темпа речи и алгоритмы его программной реализации. В данной работе использованы те же методы предварительной обработки речевого сигнала и тот же метод детектирования гласных звуков.

## 2 Предварительная обработка и анализ речевого сигнала

Оценка статистических характеристик мелодики речи базируется на предварительном определении границ гласных звуков, на интервалах которых строится траектория движения ЧОТ анализируемого отрезка речи. Гласные звуки, в сравнении с согласными и паузами, характеризуются в среднестатистическом смысле значительной интенсивностью, высокой степенью периодичности и специфическим частотным спектром. Опора на гласные звуки позволяет, как хорошо известно, выделить наиболее значимые для восприятия участки траектории ЧОТ и в то же время избежать ряда артефактов, возникающих на интервалах согласных и пауз.

Кроме задачи автоматического определения границ гласных звуков, необходимо было также выбрать наиболее эффективный алгоритм анализа ЧОТ речевого сигнала. К настоящему времени разработаны и находятся в свободном доступе (Group, 2021) четыре наиболее эффективных алгоритма: RAPT, SWIPE, REAPER и WORLD. После тщательной проверки и тестирования на большом речевом материале наилучшие результаты (минимум ошибок и сбоев) показал алгоритм RAPT [(Talkin, 1995), (Talkin, 2015)], который был выбран нами для решения поставленных задач.

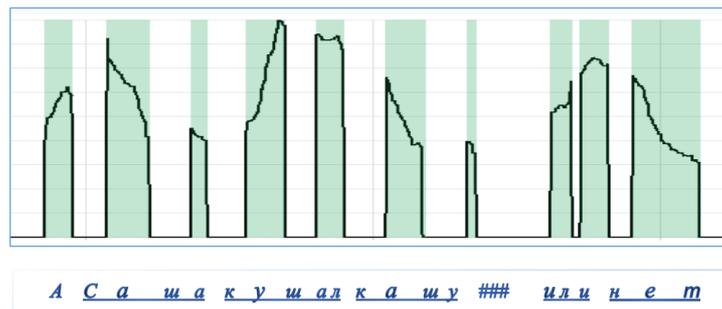


Рис. 1: Результат расчета траектория ЧОТ фразы: «А Саша кушал кашу или нет?»

На рисунке 1 приведен пример результата определения границ гласных и расчета траектории ЧОТ алгоритмом RAPT для произнесенной фразы: «А Са+ша ку+шал кашу, и+ли нет?» (знаком + указано положение словесного ударения).

## 3 Построение гистограмм статистического распределения ЧОТ речевого сигнала

Гистограмма в математической статистике — это распределение частот встречаемости дискретной случайной величины, Графически представляется в виде набора столбцов равной ширины. Высоты столбцов пропорциональны частоте их встречаемости, а количество столб-

цов задается требуемой точностью анализа в выбранном диапазоне измерения. Диапазон измерения ЧОТ и количество столбцов выбираются исходя из специфики решаемой задачи.

Так, при анализе статистических характеристик ЧОТ пения целесообразно воспользоваться логарифмической музыкальной шкалой частот в диапазоне не менее пяти октав. На рисунке 2 представлена клавиатура фортепиано с указанием названий октав и нот.



Рис. 2: Клавиатура фортепиано

Номера столбцов гистограммы, их количество и значения ЧОТ в Герцах, соответствующих клавиатуре фортепиано, представлены в таблице 1.

Контр октава	<b>1</b> 32.70	<b>2</b> 36.95	<b>3</b> 41.21	<b>4</b> 43.65	<b>5</b> 49.00	<b>6</b> 55.00	<b>7</b> 61.74
Большая октава	<b>8</b> 65.41	<b>9</b> 73.91	<b>10</b> 82.41	<b>11</b> 87.31	<b>12</b> 98.00	<b>13</b> 110.00	<b>14</b> 123.48
Малая октава	<b>15</b> 130.82	<b>16</b> 147.83	<b>17</b> 164.81	<b>18</b> 174.62	<b>19</b> 196.00	<b>20</b> 220.00	<b>21</b> 246.96
Первая октава	<b>22</b> 261.63	<b>23</b> 293.66	<b>24</b> 329.63	<b>25</b> 349.23	<b>26</b> 349.23	<b>27</b> 392.00	<b>28</b> 440.00
Вторая октава	<b>29</b> 493.88	<b>30</b> 523.25	<b>31</b> 587.32	<b>32</b> 659.26	<b>33</b> 784.00	<b>34</b> 880.00	<b>35</b> 987.75
Ноты	До	Ре	Ми	Фа	Соль	Ля	Си

Таблица 1: Номера столбцов и значения ЧОТ в Гц (логометрическая музыкальная шкала)

Формирование каждого из столбцов гистограммы осуществляется следующим образом. Значения ЧОТ измеряются на множестве гласных исследуемой выборки речевого сигнала в каждый момент дискретного времени – порядка 1 мсек. В эти моменты времени измеренная ЧОТ речевого сигнала сравнивается со всеми частотами таблицы 1. В ту из них, которая окажется ближе всего к измеренной ЧОТ, добавляется 1, и так далее, до конца анализируемого отрезка речевого сигнала. Результатом выполнения этих операций является построение гистограммы статистического распределения нот и, соответственно, значений ЧОТ.

Пример результатов расчета гистограмм, построенных с использованием пяти октавной логарифмической шкалы, представлены на рисунке 3 («Нашид» в исполнении Майкла Джексона и песня «Ямайка» в исполнении Робертино Лоретти). По горизонтальной оси отложены номера музыкальных нот, по вертикальной оси – относительные значения частоты их встречаемости. Гистограмма отображает относительное количество различных музыкальных нот, использованных певцом при исполнении данных песен.

На рисунке 4 показаны гистограммы «Здравной арии» из оперы Дж. Верди «Травиата» в исполнении известных итальянских певцов – тенора и сопрано. Очевидно нотное подобие представленных гистограмм. Их различие состоит лишь в сдвиге распределений нот на одну

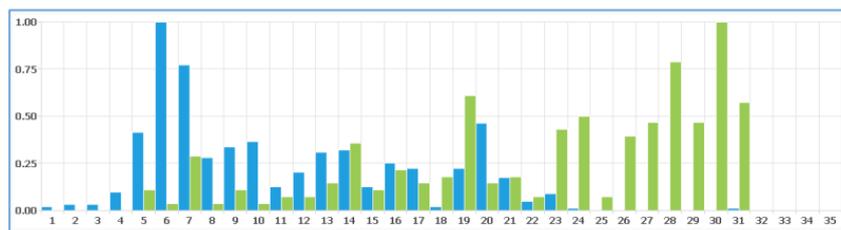


Рис. 3: Гистограммы статистических распределений нот – «Нашид» в исполнении Майкла Джексона (синие столбики) и песня «Ямайка» в исполнении Робертино Лоретти (зеленые).

октаву.

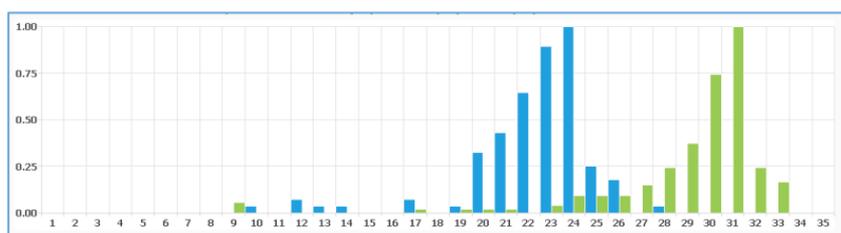


Рис. 4: Гистограмма статистического распределения нот - «Заздравная ария» в исполнении тенора (синие столбики) и сопрано (зеленые).

При анализе статистических характеристик ЧОТ речи, в отличие от пения, целесообразно воспользоваться не логарифмической, а линейной шкалой частот (Hermes and van Gestel., 1991). Рекомендуемый диапазон частот – от 60 до 400 Гц. В таблице 2 представлены номера столбцов гистограммы и значения ЧОТ в Герцах.

<b>Номер столбца</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
Частота в Гц	60	70	80	90	100	110	120
<b>Номер столбца</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>
Частота в Гц	130	140	150	160	170	180	190
<b>Номер столбца</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>
Частота в Гц	200	210	220	230	240	250	260
<b>Номер столбца</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>
Частота в Гц	270	280	290	300	310	320	330
<b>Номер столбца</b>	<b>29</b>	<b>30</b>	<b>31</b>	<b>32</b>	<b>33</b>	<b>34</b>	<b>35</b>
Частота в Гц	340	350	360	370	380	390	400

Таблица 2: Номера столбцов и значения ЧОТ в Гц (линейная речевая шкала)

Пример результатов расчета гистограмм, построенных с использованием линейной частотной шкалы (таблица 2), представлен на рисунке 5.

На рисунке 5 по горизонтальной оси отложены номера отсчетов речевой шкалы, по вертикальной оси – относительные значения частоты встречаемости ЧОТ, соответствующей номеру столбика шкалы. Зная номер столбца, легко подсчитать соответствующее ему значение в Герцах: ЧОТ [Гц] = [(номер столбца) × 10] + 50.

Ниже будут рассмотрены многочисленные примеры гистограмм с использованием линейной шкалы (таблица 2), построенных для речевых сигналов различного текстового и

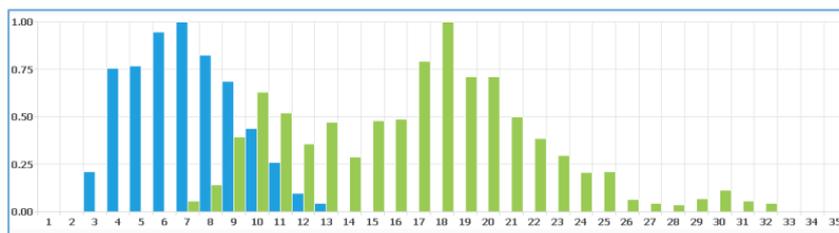


Рис. 5: Гистограмма статистических распределений ЧОТ английской речи двух профессиональных дикторов мужского пола – нейтральной (синие) и эмоциональной (зеленые).

эмоционального содержания в исполнении различных дикторов – мужчин и женщин.

#### 4 Иллюстрация проявления экспрессивной функции мелодики речи на примерах гистограмм ЧОТ

Гистограммы ЧОТ речевого сигнала могут служить в качестве полезных признаков проявления экспрессивной функции мелодики речи различного рода: дикторской индивидуальности (гендерной и внутри-гендерной), эмоционального состояния говорящего, актерских навыков, стиля речи и др. В качестве примеров рассмотрим серию гистограммы, рассчитанных для реальных реализаций речевого сигнала (см. рисунки 6 – 12).

Рисунок 6 иллюстрирует гендерные различия. На рисунке приведены гистограммы статистических распределений ЧОТ для 2-х отрезков речевого сигнала с одинаковым текстовым содержанием (около 200 слов), произнесённых дикторами мужчиной и женщиной. Отличие гистограмм очевидно. Гистограммы различаются формой распределений, центрами концентрации, шириной диапазонов и др.

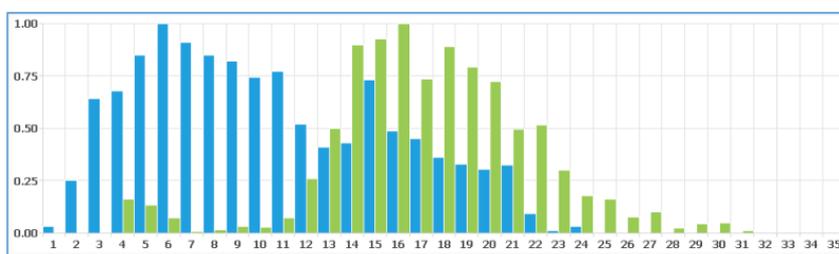


Рис. 6: Гистограммы статистических распределений ЧОТ для двух отрезков речи, произнесённых мужчиной (синие столбики) и женщиной (зеленые).

Рисунки 7 и 8 иллюстрируют наличие внутри гендерных различий. На рисунке 7 приведены гистограммы ЧОТ двух дикторов-мужчин (актеры И. Змеев и А. Канцевич), а на рисунке 8 – двух дикторов-женщин (актриса Ю. Тархова и специалист по фонетике русского языка Г. Зимовина). В их исполнении озвучен отрывок в 200 слов из рассказа А.П. Чехова «Дама с собачкой». На рисунке показаны гистограммы ЧОТ мужчин И. Змеева и А. Канцевича, а на рисунке 7 – гистограммы ЧОТ женщин - Ю.Тарховой и Г.Зимовиной. Визуальное сравнение полученных гистограмм ясно свидетельствует о наличии существенных персональных особенностей вида гистограмм ЧОТ рассмотренных пар дикторов.

Важно отметить, что визуальный «портрет» гистограмм ЧОТ характеризует манеру чтения не только отдельных фрагментов, но и на протяжении всего произведения. На рисунке 9 представлены гистограммы двух фрагментов речевого сигнала (диктор А.Канцевич - мужчина) длительностью порядка одной минуты. Один из них (помечено синим) взят из начала рассказа «Дама с собачкой», а другой – из его середины.

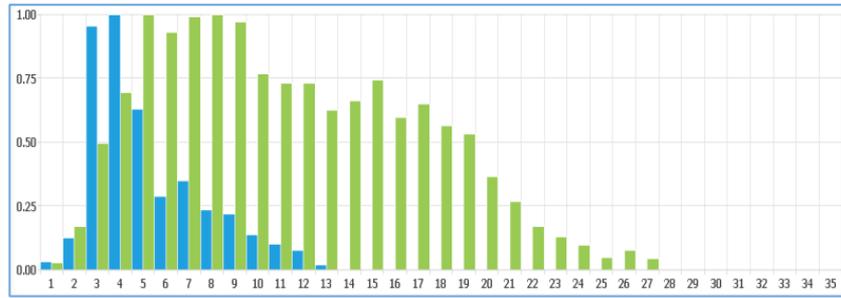


Рис. 7: Гистограммы ЧОТ мужчин: И. Змеев (синие) и А. Канцевич (зеленые)

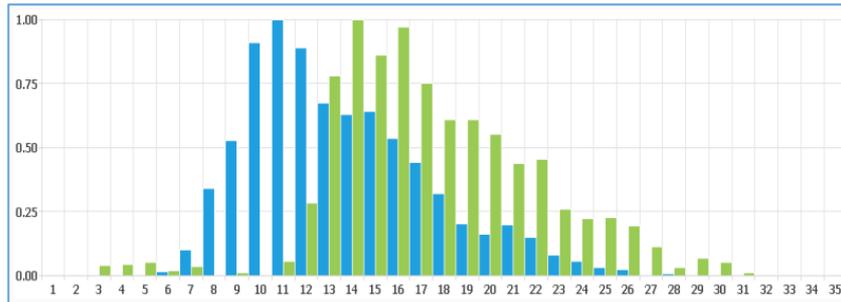


Рис. 8: Гистограммы ЧОТ женщин: Ю. Тархова (синие) и Г. Зимовина (зеленые)

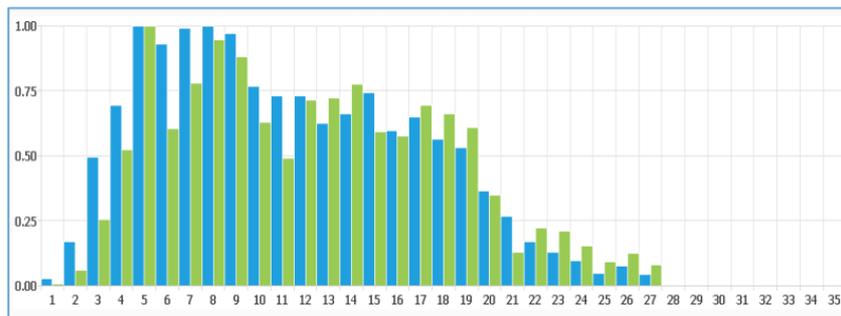


Рис. 9: Гистограммы ЧОТ мужского голоса: диктор А. Канцевич – в начале рассказа «Дама с собачкой» (синие) и в середине (зеленые)

Аналогичная картина, представленная на рисунке 10, получена для двух отрезков речевого сигнала диктора Г.Зимовиной – женщины. Один из фрагментов (помечено синим) взят из начала рассказа «Дама с собачкой», а другой – из его середины.

Интересно отметить, что вид гистограммы ЧОТ сохраняется даже при чтении диктором какого-либо другого произведения. На рисунке 10 представлены гистограммы двух фрагментов речевого сигнала длительностью порядка одной минуты. Оба фрагмента записаны в исполнении в профессионального диктора И. Змеева. Один из них (помечено синим) взят из рассказа «Дама с собачкой», а другой – из рассказа А.П. Чехова «Попрыгунья».

Представляет интерес гистограммы ЧОТ специфических стилей чтения, например, религиозных текстов. На рисунке 11 приведены гистограммы ЧОТ молитвы и проповеди.

Вид гистограмм, приведенных на рисунке 12, подтверждает известные каждому ощущение монотонности молитвы и напевности проповеди.

Любопытно увидеть также гистограммы ЧОТ речи известных исторических личностей

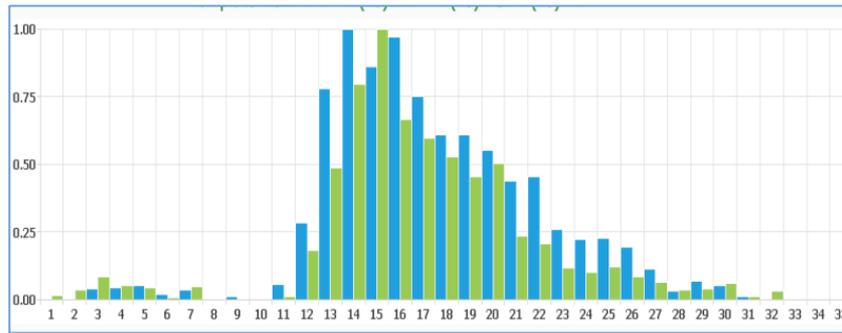


Рис. 10: Гистограммы ЧОТ женского голоса: диктор Г. Зимовина – в начале рассказа «Дама с собачкой» (синие) и в середине (зеленые)

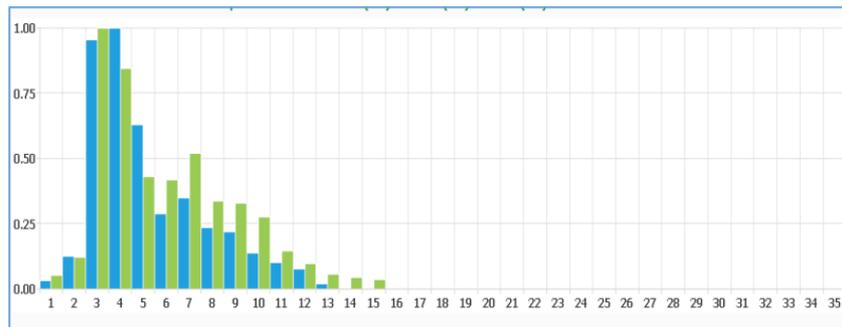


Рис. 11: Гистограммы ЧОТ речи диктора-мужчины И. Змеева – рассказы А.П. Чехова «Дама с собачкой» (синие) и «Попрыгунья» (зеленые)

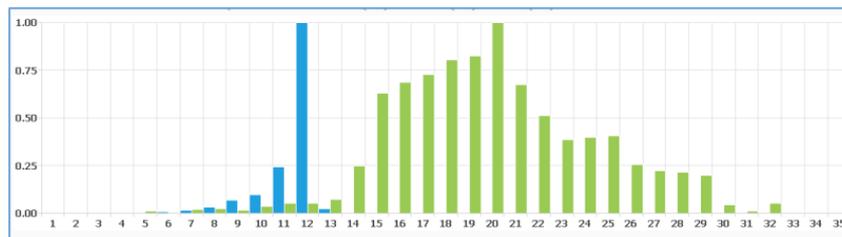


Рис. 12: Гистограммы ЧОТ молитвы (синие) и проповеди (зеленые)

(см. рисунки 13, 14), образцы речи которых приведены в ролике YouTube.com «Как звучали реальные голоса исторических деятелей в начала XX века»

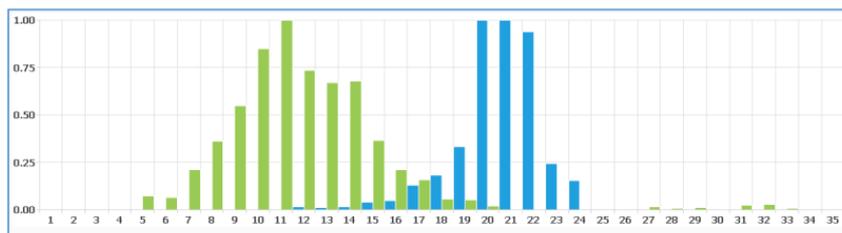


Рис. 13: Гистограммы ЧОТ выступлений: В. Ленин (синие) и И. Сталин (зеленые)

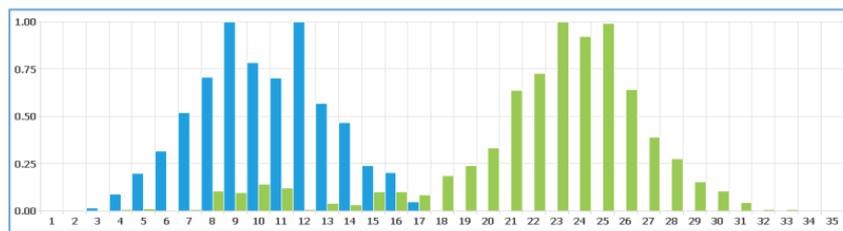


Рис. 14: Гистограммы ЧОТ выступлений: У. Черчилль (синие) и Б. Муссолини (зеленые)

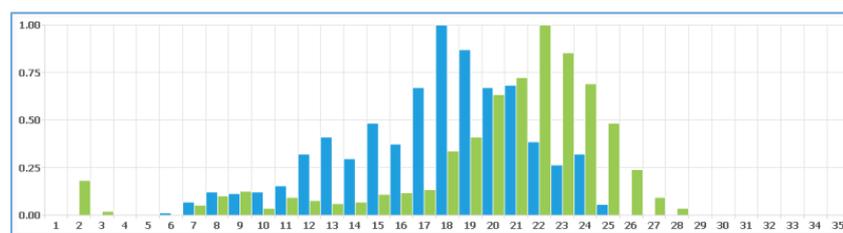


Рис. 15: Гистограммы ЧОТ выступлений: А. Вышинский (синие) и Л. Берия (зеленые)

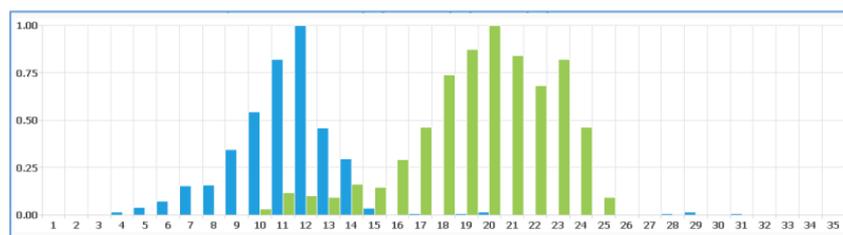


Рис. 16: Гистограммы ЧОТ выступлений: Ю. Левитан (синие) и Ю. Гагарин (зеленые)

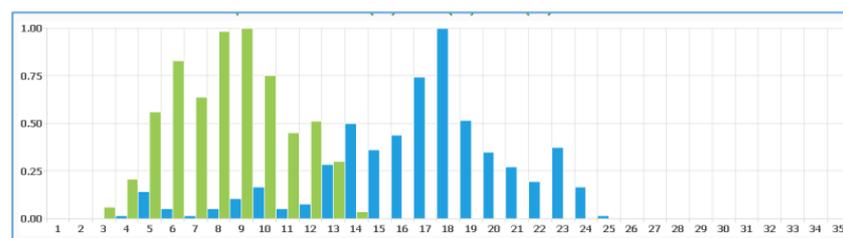


Рис. 17: Гистограммы ЧОТ выступлений: Н. Хрущев (синие) и Л. Брежнев (зеленые)

## 5 Оценка и сравнение статистических характеристик мелодики речи

Наиболее очевидным способом сравнения статистических характеристик мелодики речи двух дикторов является непосредственное сопоставление их гистограмм ЧОТ, вычисляя, например, векторное расстояние между ними или взаимную корреляцию. Для полимодальных распределений определенный интерес представляет также описание гистограмм коэффициентами Фурье.

В данной работе мы ограничимся рассмотрением способа оценки одномодальных распределений, характерных для большинства случаев. Наиболее простым в реализации и достаточно информативным является оценка гистограммы набором следующих 3-х параметров, предлагаемых в данной работе:

- центр концентрации распределения – Регистр – R,
- эффективный диапазон – Диапазон» - D,
- асимметрия распределения – Асимметрия – A.

Одним из наиболее простых и статистически эффективных способов определения центра гистограммы является расчет локализации центра «тяжести» гистограммы ЧОТ –  $C(0)$ , представленной 35-ю отсчетами: 1, 2, ... n ... 35.

Тогда:  $C(0) = [\text{SUM}(n * F0n)] / [\text{SUM}(F0n)]$ , (1)

Где  $F0n$  – значения ЧОТ на n-м отсчете гистограммы. Рассчитанное значение  $C(0)$  – условный мелодический регистр исследуемого речевого сигнала - R.

Аналогичным образом рассчитываются левый и правый центры тяжести гистограммы:  $C(0-)$  и  $C(0+)$ . А именно:  $C(0-)$  – от 1-го значения до  $[C(0)-1]$ ,  $C(0+)$  – от  $[C(0)+1]$  до последнего 35-го значения.

Результатом расчёта являются:  $C(0)$ ,  $C(0-)$  и  $C(0+)$  – целые числа, соответствующие номерам отсчетов гистограммы. Используя полученные значения, рассчитываются диапазон – D и асимметрии гистограммы – A:

$$D = C(0+) - C(0-). \quad (2)$$

$$A = [C(0+) + C(0-)] - 2C(0) \quad (3)$$

Отметим, что в результате расчета асимметрия A может принимать положительное значение, если в гистограмме преобладают высокочастотные компоненты и отрицательное – если низкочастотные.

Таким образом, используя формулы (1) – (3), получаем искомые оценки статистических характеристик мелодики речи дикторов.

На рисунке 18 представлены гистограммы ЧОТ записанных с помощью диктофона выступлений по ТВ В. Зеленского и А. Лукашенко, а на рисунке 19 – А. Невзорова и В. Жириновского, используемых в качестве примеров результатов расчета статистических характеристик мелодики речи.

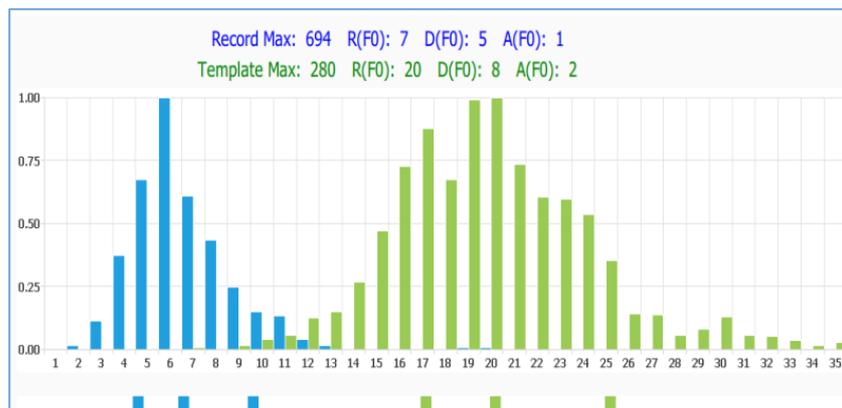


Рис. 18: Гистограммы ЧОТ выступления по ТВ В. Зеленского (синие) и А. Лукашенко (зеленые). Вверху рисунка показаны рассчитанные значения R, D, A. Внизу рисунка показаны метки позиций R, D, A на гистограммах

Полученные результаты расчётов R, D, A для указанных выше 4-х дикторов графически отображены на диаграмме (рисунок 20).

Рисунок 20 наглядно подтверждает достаточно высокую эффективность использования предложенных признаков в качестве статистических оценок персональных дикторских различий в мелодике речи.

Предложенный набор признаков может быть с успехом использован также в качестве статистических оценок эмоциональных различий в мелодике речи дикторов. На рисунках 21 и 22 представлены гистограммы ЧОТ выступлений на ТВ Ольги Скабеевой и Владимира

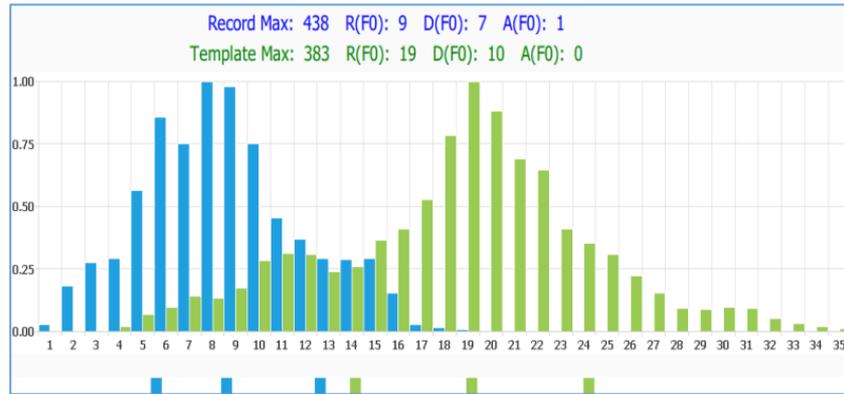


Рис. 19: Гистограммы ЧОТ выступления по ТВ А. Невзорова (синие) и В. Жириновского (зеленые). Вверху рисунка показаны рассчитанные значения R, D, A.

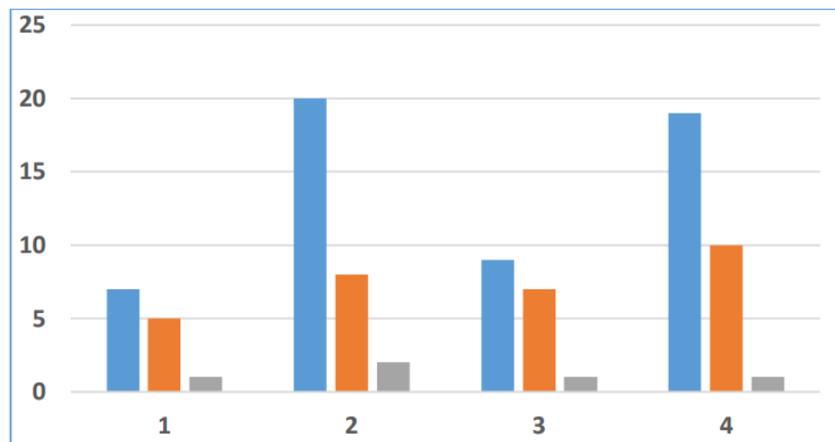


Рис. 20: Диаграмма расчетных значений R (ряд 1), D (ряд 2), A (ряд 3). Выступления по ТВ: 1 – В. Зеленский, 2 – А. Лукашенко, 3 – А. Невзоров, 4 – В. Жириновский

Соловьева в двух профессионально выраженных эмоциональных состояниях: нейтральном и в ярко-выраженном негативном.

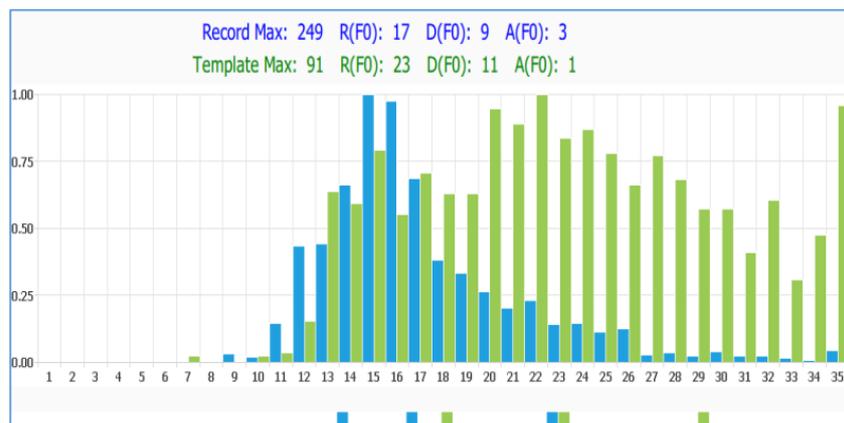


Рис. 21: Гистограммы ЧОТ выступления по ТВ Ольги Скабеевой в нейтральном эмоциональном состоянии (синие) и в ярко-выраженном негативном (зеленые)

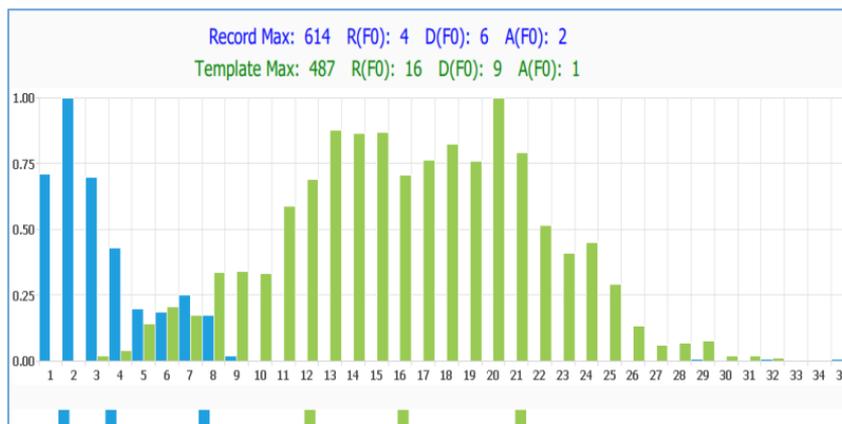


Рис. 22: Гистограммы ЧОТ выступления по ТВ Владимира Соловьева в нейтральном состоянии (синие) и в ярко-выраженном негативном (зеленые).

Полученные результаты расчётов  $R$ ,  $D$ ,  $A$  для указанных дикторов и их эмоционального состояния графически отображены на диаграмме (рисунок 23).

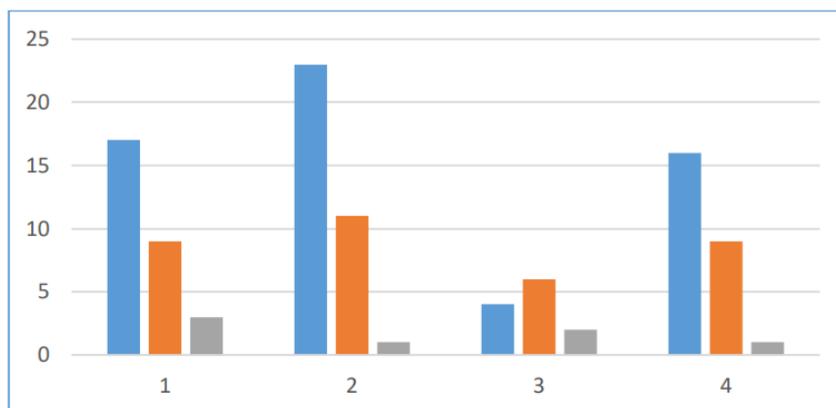


Рис. 23: Диаграмма расчетных значений  $R$  (ряд 1),  $D$  (ряд 2),  $A$  (ряд 3): 1 – О.С. (нейтр.), 2 – О.С. (негат.), 3 – В.С. (нейтр.), 4 – В.С. (негат.)

Рисунок 23 наглядно подтверждает достаточно высокую эффективность использования предложенных признаков в качестве статистических оценок эмоциональных дикторских различий в мелодике речи.

## 6 Программная реализация предложенного метода

Приведенные в разделах 2 – 5 результаты получены с использованием разработанного прототипа системы Speech Melody Meter (SMM), являющейся программной реализацией рассмотренного метода оценки статистических характеристик мелодики речи.

В основу программной реализации метода положен тот же подход, который был использован при разработке системы Speech Rate Meter (SRM) для статистической оценки просодических параметров темпа речи (В. Lobanov, 2021).

Прототип системы работает следующим образом. Сразу после запуска программы пользователю предоставляются следующие возможности (см. рисунок 24):

- Запись через микрофон достаточно длительного отрезка анализируемой речи (желательно не менее 30 с);
- Вызов пары заранее подготовленных тестовых аудиофайлов, хранящихся в папке “data-tests”.



Рис. 24: Стартовое окно интерфейса программы

После того как анализируемый речевой сигнал введен с микрофона или из папки “data-tests”, осуществляется построение гистограммы статистического распределения значений ЧОТ речевого сигнала и расчет статистические характеристики ЧОТ: R, D, A с экранном отображением результатов расчета (см. рисунки 18, 19 и 21, 22).

Алгоритмы расчёта и обработки данных написаны на языке “C” с использованием библиотек GNU Scientific Library, Speech Signal Processing Toolkit (SPTK) и OpenAL. Разработан прототип системы для OS Windows – «Speech Melody Meter (SMM)». Программный модуль будет размещен на сайте <https://intontrainer.by> и общедоступен для бесплатной загрузки и тестирования. Планируется также разработка мобильной версии системы SMM в OS Android.

## 7 Заключение

- Предложен метод оценки статистических характеристик мелодики речи и описана процедура построения гистограмм распределения частот встречаемости дискретных значений частоты основного тона (ЧОТ) на достаточно длительных интервалах речи.
- Отличительной особенностью предложенного метода является то, что дискретные значения ЧОТ речи измеряются только на интервалах гласных звуков исследуемого речевого фрагмента.
- Выбраны два варианта шкал ЧОТ, используемые при анализе характеристик мелодики, а именно: линейная шкала для анализа речи и логарифмическая (музыкальная) – для пения.
- Предложен метод оценки трех параметров гистограммы ЧОТ: регистр – R, диапазон – D, асимметрия – A и показана их достаточно высокая различительная способность.
- Приведены многочисленные примеры, показывающие эффективность предложенного метода при оценке индивидуальности мелодики речи диктора, а также его эмоционального состояния.
- Дано краткое описание прототипа системы Speech Melody Meter (SMM), являющейся программной реализацией предложенного метода оценки статистических характери-

стик мелодики речи (см. также: <https://intontrainer.by>).

Программная реализация разработанного метода может быть полезной в следующих практически важных ситуациях:

- Для индивидуального тренинга при подготовке выступлений и презентаций для самоконтроля и обучения желаемым характеристикам мелодики речи;
- В качестве средства самоконтроля для устранения дефектов речи, связанных с различными физиологическими нарушениями и болезнями;
- Как средство мониторинга мелодики речи в реальных условиях работы операторов колл-центра, дикторов радио и телевидения.
- Определенный интерес представляет использование разработанной программы SMM для непрерывного самоконтроля особенностей мелодики речи – одного из важных показателей эмоционального интеллекта человека.

В заключение заметим, что в данной работе предприняты лишь первые шаги в направлении численной оценки экспрессивной функции мелодики речи. Дальнейшее развитие данного направления может быть связано как с использованием более тонких методов статистического анализа траекторий ЧОТ, так и с более тонкими информативными признаками гистограммам, например, описание коэффициентами ряда Фурье.

## References

- S. Marchina G. Schlaug A. Norton, L. Zipse. 2009. *Melodic intonation therapy: Shared insights on how it is done and why it might help*. Annals of the New York Academy of Sciences.
- V. Zhitko B. Lobanov. 2021. *Method for Statistical Estimation of the Prosodic Parameters of Speech Tempo*.
- et al C. Leyns. 2010. *Age and gender differences in Belgian Dutch intonation*, volume 4, No. 1. Majlesi Journal of Electrical Engineering.
- M. Cahill. 2016. *Intonation and emotions in Knni: A preliminary study*. In Doris L. Payne, Sara Pacchiarotti Mokaya Bosire (eds.), *Diversity in African languages*, Berlin. Language Science Press.
- et al. Celine De Looze. 2017. *Effects of cognitive impairment on prosodic parameters of speech production planning in multiple sclerosis*. Journal of Neuropsychology. The British Psychological Society.
- M. Janipour D. Gharavian, M. Sheikhan. 2010. *Pitch in Emotional Speech and Emotional Speech Recognition Using Pitch Frequency*, volume 4, No. 1. Majlesi Journal of Electrical Engineering.
- SPTK Working Group. 2021. *pitch*.
- Dik J. Hermes and Joost C. van Gestel. 1991. *The frequency scale of speech intonation*, volume 90, No. 1. The Journal of the Acoustical Society of America.
- J. Hernando M. Farru´s. 2009. *Using Jitter and Shimmer in speaker verification*, volume 3, Iss. 4. IET Signal Process, Englewood Cliffs, NJ.
- B. Yegnanarayana M. Leena. 2008. *Extraction and representation of prosodic features for language and speaker recognition*, volume 50. Speech Communication.
- S. Peppé. 2009. *Why is prosody in speech-language pathology so difficult?*, volume 11 (4). International Journal of Speech-Language Pathology.
- K. Scherer T. Banziger. 2005. *The role of intonation in emotional expressions*, volume 46. Speech Communication.
- D. Talkin. 1995. *A robust algorithm for pitch tracking*. Speech Coding and Synthesis.
- D. Talkin. 2015. *REAPER: Robust epoch and pitch estimator*.

## 8 ПРИЛОЖЕНИЕ

Аудио файлы примеров, использованных при расчете приведенных в докладе гистограмм. См: <https://intontrainer.by/papers/PaperAudioAssets.zip>