

# Method for Statistical Estimation of the Prosodic Parameters of Speech Tempo (based on Russian speech)

**Boris Lobanov**

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

**Vladimir Zhitko**

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

## Abstract

We propose to describe the rate of speech as a complex prosodic component of intonation, the parameters of which are: 1) Speech rate -  $R_s$ , i.e. the general rate of reading the text; 2) Articulation rate -  $R_a$ , i.e. the rate of generation of the phonemic stream; 3) Pauses score -  $T_p$ , i.e. the average duration of the inter-phrase pauses. To this end, we have developed a new method for determining vowel sounds, which allows us to obtain an effective set of statistical characteristics of speech fragments analyzed:  $N_v$  - number of vowel segments;  $N_c$  - number of consonants plus pause segments;  $T_v$  - total duration of the vowel segments;  $T_{ca}$  - root-mean-square duration of consonants plus pauses;  $T_{cm}$  - median of the duration of consonants plus pauses. On this basis, formulas for calculating the numerical estimates of the complex of parameters ( $R_s$ ,  $R_a$ ,  $T_p$ ) have been developed and a software model for the real time automatic estimation of this complex has been created. The software model is posted on the website <https://intontrainer.by> and is publicly available for free download and testing.

**Keywords:** speech prosody, speech rate, software model, vowels detection, prosodic parameters estimation

**DOI:** 10.28995/2075-7182-2021-20-1120-1129

## Метод статистической оценки просодических параметров темпа речи (на материале русской 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

### 1 Введение

Энциклопедический словарь (см.: <https://bigenc.ru/linguistics/text/4186876>) определяет темп речи как скорость произнесения элементов речи: звуков, слогов или слов. Измеряется числом этих элементов, произносимых в единицу времени. Различают артикуляторный темп, измеряемый только на отрезках говорения, и полный темп, подсчитанный с учетом пауз. Темп речи является одним из просодических компонентов фразовой интонации.

Темп речи является одним из важнейших факторов, влияющих на разборчивость и выразительность речи. Известно, что скорость произношения значительно варьирует среди носителей родного и иностранного языков [4, 2]. Темп речи оказывает большое влияние на понимание речи как носителями родного языка, так и иностранцами [3, 1]. Определенные параметры темпа речи во многом связаны с возрастом, полом и диалектным акцентом говорящего [7], а также с его различными физиологическими нарушениями и болезнями [6].

Темп речи как просодический компонент интонации выполняет следующие функции:

- выражает степень важности содержания (более важное содержание – медленным темпом, менее важное – быстрым темпом);
- привлекает внимание слушающего (посредством замедления или ускорения темпа);
- выражает эмоциональное состояние (положительные эмоции чаще всего ускоряют темп, отрицательные эмоции замедляют его);
- служит средством создания его интонационной завершенности (замедляет темп к концу высказывания).

Существуют различные способы оценки параметров темпа речи. На практике темп речи чаще всего измеряется числом слов, произносимых в одну минуту (в англоязычной литературе обозначается как wpm /words per minute/). В простейшем случае, когда произносимый текст известен (чтение текста подготовленного доклада или прослушивание известной аудиокниги), для того чтобы рассчитать численное значение полного темпа речи  $R_s$  [wpm], нужно разделить общее количество слов в произнесенном тексте  $N_w$  на длительность его звучания  $T_s$  [s]:

$$R_s [\text{wpm}] = N_w * 60 / T_s.$$

Во многих других практически важных ситуациях произносимый текст заранее неизвестен (выступление оратора, телефонный разговор, радио- и ТВ-передачи). В этом случае возникает нетривиальная задача оценки количества звуков, слогов или слов, произнесенных в исследуемом фрагменте речи. В предлагаемом методе оценка темпа речи базируется на автоматическом выделении и подсчете числа гласных звуков  $N_v$  в анализируемом отрезке речи. Как известно (см. <https://slogi.su/>), слоги образуются только за счёт гласных звуков, поэтому число слогов равно числу гласных в слове. Результат деления этого числа на длительность речи является фактически выражением темпа в количестве слогов за единицу времени. В то же время, деление этого результата на статистическую оценку среднего количества гласных в слове  $L_v$  [5] даёт широко распространённую оценку в количестве слов в минуту. В этом случае формула (1) может быть представлена в виде

$$R_s [\text{wpm}] = (L_v * N_v * 60) / T_s.$$

Темп речи авторы предлагают описывать как комплексный просодический компонент фразовой интонации, тремя параметрами которого являются:

1. полный темп (рассчитывается как количество слов, произнесенных в одну минуту за фактическое время звучания речи, т. е. с учетом длительности межфразовых пауз);
2. артикуляторный темп (рассчитывается как количество слов в минуту за суммарное время артикуляции слов, т.е. без учета длительности межфразовых пауз);
3. показатель паузации (рассчитывается как среднеквадратическое значение длительности межсинтагменных и межфразовых пауз).

Отметим, что рассмотренные параметры в значительной степени являются независимыми и выражают комплексную характеристику темпа речи. Например, полный темп может быть низким из-за большой длительности межфразовых пауз, но при этом артикуляторный темп может быть весьма высоким. Показатель паузации также не связан простой зависимостью с первыми двумя параметрами.

## 2 Предварительный анализ и обработка речевого сигнала

С точки зрения фонетики достаточно длительный фрагмент речи представляет собой непрерывную дискретную последовательность гласных, согласных и синтагматических (фразовых) пауз. В этой непрерывной последовательности, по крайней мере, один гласный звук обязательно присутствует в каждом слове (за исключением некоторых служебных слов). Гласные звуки в сравнении с согласными и паузами характеризуются в среднестатистическом смысле значительной интенсивностью (амплитудой колебаний), высокой степенью периодичности и специфическим частотным спектром.

На рисунке 1 представлена осциллограмма и спектрограмма двухсинтагменной фразы «Мама мыла Нину, а Саша кушал кашу». Фонетический состав фразы подобран таким

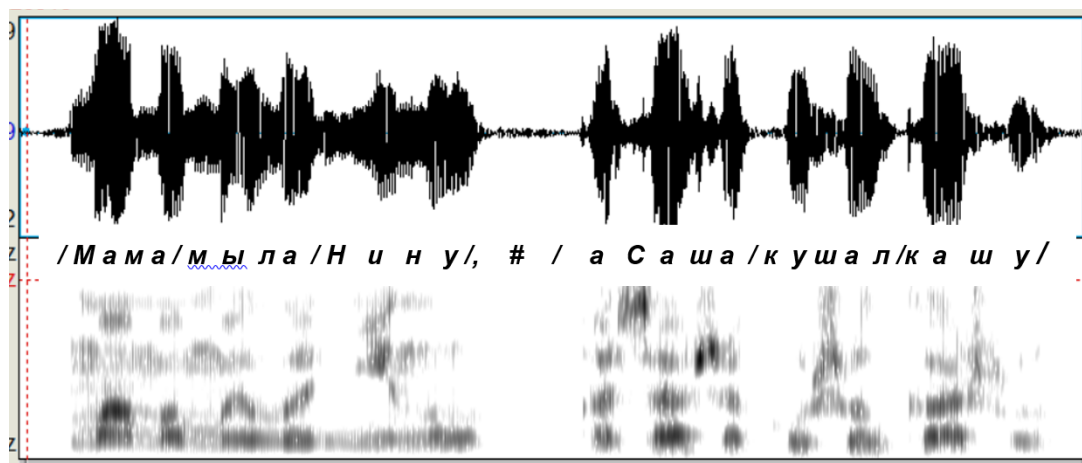


Рис. 1: Осциллограмма и спектрограмма двухсинтагменной фразы «Мама мыла Нину, а Саша кушал кашу».

образом, чтобы продемонстрировать наиболее трудные случаи различения гласных и согласных звуков, а именно: гласных и сонорных (первая синтагма), гласных и фрикативных согласных (вторая синтагма). Во фразе присутствует также межсинтагменная пауза, помеченная значком #. Рисунок 1 иллюстрирует отмеченные выше особенности гласных звуков в сравнении с согласными: их значительную интенсивность и специфический частотный спектр.

Основными задачами анализа и обработки речевого сигнала являются сегментация сигнала на участки гласных, согласных и пауз, измерение длительности этих участков и расчет статистических характеристик их длительностей на анализируемом фрагменте речи. Процедура анализа и обработки речевого сигнала состоит из следующих пяти этапов.

На первом этапе производится фильтрация речевого сигнала с помощью фильтра, частотная характеристика которого согласована со среднестатистическим спектром гласных звуков. При практической реализации программной модели оказалось достаточным использование обычного полосового фильтра в диапазоне частот 450 – 1800 Гц. В результате применения такого рода фильтрации в выходном сигнале максимально подчеркиваются гласные звуки в сравнении с согласными для множества мужских и женских голосов.

На втором этапе осуществляется вычисление кривой интенсивности звуков речи  $E(n)$ . При вычислении кривой интенсивности должны быть выполнены два требования:

1. сглаживание возможных колебаний кривой интенсивности с частотой основного тона  $F_0$ ,
2. сохранение динамики изменения интенсивности звуков речи.

Для выполнения этих требований необходимо выбрать подходящую длительность окна анализа  $T_w = q * (1 / F_{0min})$ . При выборе практически приемлемых значений ( $F_{0min} = 50$  Гц и  $q = 1,5$ ) получим значение  $T_w = 0,03$  с.

Вычисление кривой интенсивности звуков речи  $E(n)$  для каждого  $n$ -го отсчета сигнала выполняется следующим образом:

1. Рассчитывается временное окно анализа  $N_{frame}$  речевого сигнала  $s(n)$ :

$$N_{frame} = F_d * T_w, \quad (3)$$

где  $F_d$  – частота дискретизации сигнала;  $T_w$  – длительность окна анализа. При выбранных  $F_d = 8000$  Гц и  $T_w = 0,03$  с получаем  $N_{frame} = 240$  отсчетов.

2. Для получения  $n$ -го отсчета кривой интенсивности  $E(n)$  речевого сигнала  $S(i)$  рассчитывается среднее значение модуля сигнала  $|s(i)|$  на интервале окна анализа  $[i, i+N_{frame}]$ .

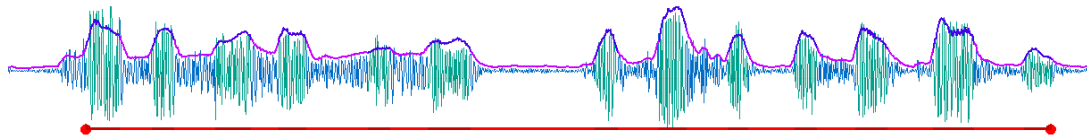


Рис. 2: Результат вычисления кривой интенсивности звуков  $E(n)$ .

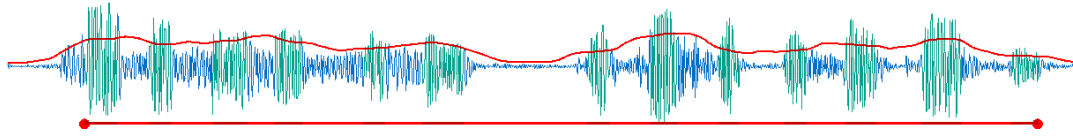


Рис. 3: Результат вычисления сглаженной кривой интенсивности  $E(m)$ .

3. Для получения следующего отсчета окно анализа смещается на количество отсчетов равное  $N_{\text{shift}}$ , а затем повторяется предыдущий шаг вплоть до конца анализируемого сигнала.

С точки зрения достижения приемлемой точности при минимуме вычислительных затрат, оптимальным может быть рекомендован выбор значения  $N_{\text{shift}} = \frac{1}{2} * N_{\text{frame}}$ .

На рисунке 2 представлен график результата вычисления огибающей  $E(n)$  речевого сигнала для фразы «Мама мыла Нину, а Саша кушал кашу».

На третьем этапе осуществляется вычисление сглаженной кривой интенсивности  $E(m)$  путем дополнительной обработки кривой  $E(n)$ , полученной на предшествующем этапе. При этом должны быть выполнены два требования:

1. сглаживание возможных колебаний с частотой изменения последовательности звуков речи,
2. сохранение динамики изменения средней интенсивности отдельных синтагм и фраз.

Для выполнения этих требований необходимо выбрать подходящую длительность окна анализа  $T_m$ , исходя из условия:

$$T_p > T_m > T_w, \quad (4)$$

где  $T_p$  и  $T_w$  – ожидаемые длительности фраз и слов соответственно. Вычисление сглаженной кривой интенсивности  $E(m)$ , в соответствии с полученным для  $E(n)$  интервалом ее дискретизации  $N_{\text{shift}}$ , выполняется следующим образом:

1. Рассчитывается дополнительное (сглаживающее) временное окно анализа  $N_{\text{smooth}}$  кривой интенсивности звуков речи  $E(n)$ :  

$$N_{\text{smooth}} = F_d * T_m, \quad (5)$$
 где:  $F_d$  – частота дискретизации сигнала,  $T_m$  – длительность окна анализа. При выбранных  $F_d = 8000$  Гц и  $T_m = 0,3$  сек., получаем  $N_{\text{smooth}} = 2400$  отсчетов.
2. Для получения  $m$ -го отсчета сглаженной кривой интенсивности речевого сигнала –  $E(m)$  – рассчитывается среднее значение кривой интенсивности –  $E(n)$  на интервале окна анализа  $[n, n + N_{\text{smooth}}]$ .
3. Для получения следующего отсчета окно анализа смещается на количество отсчетов, равное  $N_{\text{shift}}$ , а затем повторяется предыдущий шаг вплоть до конца анализируемого сигнала.

На рисунке 3 представлен график результата вычисления сглаженной кривой интенсивности  $E(m)$  речевого сигнала для фразы «Мама мыла Нину, а Саша кушал кашу».

На четвертом этапе осуществляется определение границ сегментов гласных звуков. Для этого производится сравнение двух показанных выше кривых интенсивности звуков  $E(n)$  (см. рисунок 2) и сглаженной кривой интенсивности  $E(m)$  (см. рисунок 3). Точки пересече-

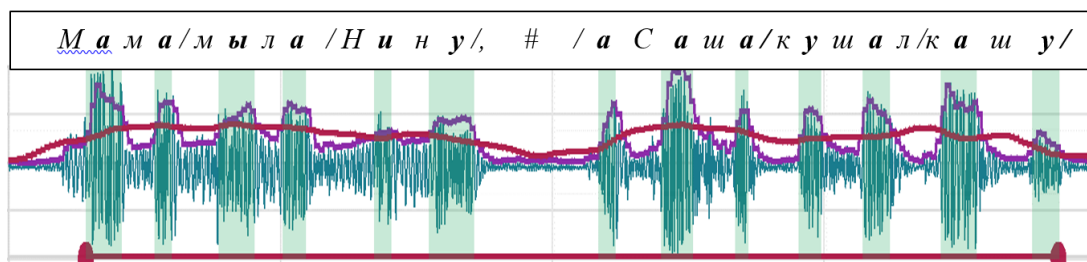


Рис. 4: Иллюстрация способа определения границ гласных сегментов.

ния этих кривых показывают границы гласных звуков, а интервалы, на которых выполняется неравенство  $E(n) > E(m)$ , соответствуют временным интервалам присутствия гласных звуков (см. рисунок 4).

В результате выполнения указанных операций на выходе четвертого этапа обработки речевого сигнала получаем непрерывную последовательность двоичных сигналов 1,0,1,0, ... ,0 различной длительности. При этом временные интервалы единичных значений сигнала будут соответствовать интервалам гласных (см. рисунок 4), а нулевые значения – интервалам согласных и пауз.

На пятом этапе осуществляется расчет набора статистических характеристик длительности сегментов, используемый для оценки параметров темпа анализируемого отрезка речи. Необходимым и достаточно эффективным оказался следующий набор статистических характеристик распределений длительности сегментов:

1.  $N_c$  – количество сегментов согласных и пауз;
2.  $T_s$  – длительность анализируемого отрезка речи;
3.  $T_v$  – суммарная длительность гласных сегментов;
4.  $T_{ca}$  – среднеквадратическая длительность согласных и пауз;
5.  $T_{cm}$  – среднемедианная длительность согласных и пауз.

Способы расчета перечисленных характеристик по известной последовательности двоичных сигналов различной длительности, получаемой на выходе четвертого этапа обработки, достаточно очевидны и выполняются с использованием известных процедур и математических формул.

На рисунке 5 представлена блок-схема алгоритма, реализующего процедуры анализа и обработки речевого сигнала на описанных выше пяти этапах.

### 3 Численная оценка параметров темпа речи

В данном разделе рассматриваются вопросы применения перечисленных выше статистических характеристик для расчета численных оценок комплекса просодических параметров темпа речи, а именно:

- Полного темпа речи  $R_s$  (wpm);
- Артикуляторного темпа  $R_a$  (wpm);
- Показателя паузации  $T_p$  (sec).

#### 3.1 Оценка полного темпа речи $R_s$

Перепишем формулу (2) для полного темпа речи в виде

$$R_s [\text{wpm}] = k_1 * N_v * 60 / T_s, \quad (6)$$

где  $N_v$  и  $T_s$  – известные величины, рассчитываемые в блоке 5 (см. рисунок 5),  $k_1$  – неизвестная константа.

Для статистической оценки значения  $k_1$  создается набор эталонных речевых образцов с известным количеством слов, начитанных различными дикторами. Таким образом, для



Рис. 5: Алгоритм предварительной обработки речевого сигнала.

каждого выбранного речевого образца нам известны количество произнесенных слов  $N_w$  и длительность  $T_s$ . Эти данные, в соответствии с формулой (1), определяют полный темп речи  $R_s$ .

Таким образом, зная  $R_s$ ,  $N_w$ ,  $T_s$  и применяя формулу (6) для каждого  $i$  – го речевого образца, можно определить набор значений  $k_{1i}$ :

$$k_{1i} = (R_{si} * T_{si}) / (N_{wi} * 60). \quad (7)$$

Расчет набора коэффициентов  $k_{1i}$  по формуле (7) производится многократно при различных значениях параметров используемой программной модели  $N_{frame}$  и  $N_{shift}$  (см. раздел 2). В результате такой итерационной процедуры определяются оптимальные значения этих параметров. Оптимальными  $N_{frame}$  и  $N_{shift}$  будут считаться такие их значения, при которых достигается минимальный разброс (дисперсия) множества значений  $k_{1i}$ , рассчитанных по формуле (7). Окончательно величина  $k_1$ , используемая в формуле (6) для численной оценки полного темпа речи, определяется как средняя величина полученного множества  $k_{1i}$ .

### 3.2 Оценка артикуляторного темпа речи – $R_a$

Артикуляторный темп – это количество слов в минуту, подсчитанное за период артикуляторной активности (без учета длительности межфразовых пауз  $T_p$ ). Таким образом, артикуляторный темп определяется как количество слов, которое могло бы быть произнесено в одну минуту, если бы длительность пауз равнялась нулю. Исходя из такого определения, по аналогии с (6) можно записать

$$R_a [wpm] = k_2 * N_v * 60 / T_a, \quad (8)$$

где  $k_2$  – неизвестная константа,  $N_v$  – подсчитанное число гласных в анализируемом от-



резке речи,  $T_a$  – суммарная длительность артикуляторной активности.

Для нахождения статистической оценки значения  $T_a$  предлагается использовать следующую формулу:

$$T_a = T_v + (N_c * T_{cm}), \quad (9)$$

где  $T_v$  – суммарная длительность гласных сегментов,  $N_c$  – общее количество сегментов согласных и пауз,  $T_{cm}$  – средне-медианное значение длительности согласных и пауз.

Здесь  $T_v$ ,  $N_c$  и  $T_{cm}$  – известные величины, рассчитываемые в блоке 5 (см. рисунок 5).

В формуле (9) второе слагаемое является по существу статистической оценкой суммарной длительности согласных без учета длительности межфразовых пауз. Действительно, известно [5], что длительность согласных в среднем варьирует в пределах 100-150 мс, а длительность межфразовых пауз – в пределах 500 – 1500 мс. При этом количество согласных на достаточно длительном отрезке речи значительно превышает количества пауз. Отсюда следует, что медиана  $T_{cm}$ , являющаяся средним значением в отсортированной по возрастанию последовательности чисел, может считаться типичным представителем длительности согласных. Их общая длительность ( $T_{cm} * N_c$ ) может считаться достаточно правдоподобной оценкой суммарной длительности согласных без учета длительности межфразовых пауз.

Таким образом, формула (8) окончательно может быть представлена в следующем виде:

$$R_a [wpm] = k_2 * N_v * 60 / (T_v + N_c * T_{cm}). \quad (10)$$

Численное значение константы  $k_2$  находится экспериментально с использованием процедуры, подобной процедуре определения константы  $k_1$ , описанной в разделе 3.1. Единственное отличие заключается в том, что в каждом из эталонных речевых образцов необходимо удалить все межфразовые паузы при сохранении общего количества слов в каждом из них.

### 3.3 Оценка показателя паузации речи $T_p$

Для оценки показателя паузации рассчитывается степенное среднее длительностей сегментов согласных и пауз анализируемого фрагмента речевого сигнала. Воспользуемся известным математическим понятием степенного среднего множества чисел. При степени  $n = 1$  его величина называется среднеарифметическим значением, при степени  $n = 2$  – среднеквадратическим и т. д. Известно также, что на величину степенного среднего существенно влияет величина наибольшего из чисел, причем при  $n$ , стремящемся к бесконечности, степенное среднее стремится к значению, равному наибольшему из множества чисел. В рассматриваемом случае анализируется набор чисел, определяющих длительности сегментов согласных и пауз. Как было сказано выше, средняя длительность пауз на порядок превышает длительность согласных. Следовательно, величина степенного среднего для множества длительностей сегментов согласных и пауз в определенной мере отображает значение средней длительности пауз на анализируемом отрезке речи.

Для определения искомого значения показателя паузации  $T_p$ , используем рассчитанные в блоке 5 (рис. 2) значение  $T_{ca}$  (среднеквадратическая длительность согласных и пауз) и  $T_{cm}$  (медиана множества длительностей сегментов согласных и пауз).  $T_{cm}$  считается типичным представителем средней длительности согласных, а  $T_{ca}$  – монотонной функцией от искомой длительности пауз. Учитывая это, авторы предлагают следующую формулу для статистической оценки среднего значения длительности межфразовых пауз, принимаемого как показатель паузации:

$$T_p = k_3 * (T_{ca} - T_{cm}). \quad (11)$$

Численное значение константы  $k_3$  находится экспериментально с использованием указанного выше набора эталонных речевых образцов. Используя формулу (11) для каждого  $i$  – го речевого образца можно определить:

$$k_3 i = T_{pi} / (T_{cai} - T_{cmi}). \quad (12)$$

В этой формуле статистические характеристики  $T_{cai}$  и  $T_{cmi}$  известные величины, рассчитываемые в блоке 5 (см. рисунок 5). Для определения реальных значений средней длительности пауз  $T_{pi}$  проводится серия измерений и расчетов для каждого  $i$  – го образца из

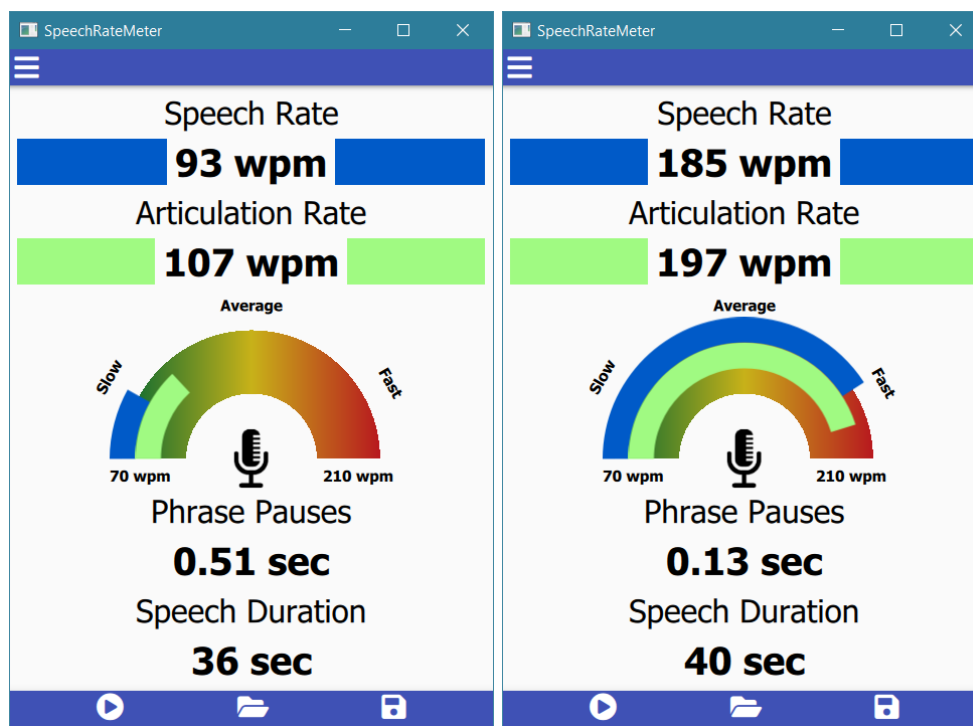


Рис. 6: Информационное окно SRM. а) Медленный темп б) Быстрый темп

используемого набора речевых образцов. Окончательное значение  $k_3$ , используемое в формуле (11), определяется как среднее значение чисел –  $k_{3i}$ , полученных в результате расчетов по формуле (12).

Таким образом, используя формулы (6), (10), (11), и подставляя в них соответствующие расчетные значения статистических характеристик распределений длительности сегментов (см. раздел 2), получаем искомый набор численных оценок параметров темпа речи:

- полный темп речи  $R_s$  /wpm/;
- артикуляторный темп  $R_a$  /wpm/;
- показатель паузации  $T_p$  /s/.

#### 4 Программная модель и результаты тестирования

Программная модель рассмотренного метода реализована в ОС Windows (7, 8, 10) как отдельное приложение Speech Rate Meter (SRM).

Сразу после запуска программы пользователю предоставляются следующие возможности:

- Запись через микрофон достаточно длительного отрезка анализируемой речи (желательно не менее 10 с);
- Вызов одного из заранее подготовленных тестовых аудиофайлов, хранящихся в папке “data-tests”.

После того как анализируемый речевой сигнал введен с микрофона или из папки “data-tests”, осуществляется расчет параметров темпа анализируемой речи и открытие информационного окна SRM (рис. 6) с экранным отображением результатов расчета.

В средней части окна SRM графически показаны параметры Speech Rate ( $R_s$ ) и Articulation Rate ( $R_a$ ) в виде двух подвижных дуг разного цвета: синего  $R_s$  и зеленого  $R_a$ . Длина дуг нормирована относительно задаваемых пользователем минимального и максимального значений этих параметров. Выбор минимального и максимального значений



определяется конкретными условиями использования SRM, при которых измеренное значение темпа речи будет считаться медленным Slow, средним Average или быстрым Fast. Выбранные минимальное и максимальное значения /wpm/ показаны на концах дуги полуокружности.

Пользователю, исходя из его собственных научных или практических интересов, предоставляется возможность записи в папку “tests” неограниченного количества речевых аудио файлов от различных источников. В качестве примера в папке “tests” размещены 10 образцов естественной речи:

- Примеры (1, 2) – образцы медленной и быстрой речи на английском языке;
- Пример (3) – образец радиопередачи о погоде на белорусском языке;
- Примеры (4, 5) – аудио записи двух одинаковых фрагмента рассказа А.П. Чехова «Дама с собачкой» в исполнении преподавательницы русского языка и профессионального актера;
- Примеры (6, 7, 8) – фрагменты ТВ-передач на русском языке: реклама и монологи женщины и мужчины;
- Примеры (9, 10) – медленное и быстрое чтение фрагмента стихотворения на русском языке.

Перечисленные выше примеры тестовых аудио файлов имеют сравнительно небольшой объем: максимальная длительность до 500 с и содержат до 900 слов (примеры 4, 5). Дополнительно были проведены также испытания с использованием аудиофайлов значительно большего объема: длительность до 3000 с, их состав - до 5000 слов. Анализ результатов тестирования проведен на материале русскоязычных аудиокниг и озвученных упражнения для уроков русского языка. В используемом тестовом материале были известны текстовое и звуковое содержание речевых образцов. Это дает возможность оценить для каждого из образцов их реальный полный темп  $R_s$  путем деления числа слов на длительность записи. Расчет реальных значений артикуляторного темпа  $R_a$  и показателя паузации  $T_p$  осуществлялся путем подсчета количества и длительности межсинтагменных пауз путем анализа осциллограммам сигналов. Полученные данные принимались за эталонные, которые сравнивались с измеренными данными с помощью разработанного программного комплекса SRM. На основе полученных данных вычислялась усредненная погрешность численной оценки каждого из параметров темпа речи. Анализ полученных результатов тестирования показал, что погрешность численной оценки каждого из параметров темпа речи достаточна низкая и лежит в пределах до 25 процентов.

## Заключение

Разработана программная реализация описанного метода – «Speech Rate Meter» - для OS Windows. Программный модуль размещен на сайте <https://intontrainer.by> и общедоступен для бесплатной загрузки и тестирования.

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

- При подготовке выступлений и презентаций для самоконтроля и обучения желаемому темпу речи;
- Для совершенствования навыков свободного чтения на родном и иностранном языке;
- В качестве средства самоконтроля для устранения дефектов речи, связанных с различными физиологическими нарушениями и болезнями;
- Для индивидуального тренинга устно-речевых навыков в ряде профессий таких, как: операторы колл-центров, дикторы радио, ТВ и др.;
- Как средство мониторинга темпа речи в реальных условиях работы операторов колл-центра, дикторов радио и телевидения.

К настоящему времени завершена также разработка мобильной версии системы Speech Rate Meter в OS Android, которая доступна для скачивания и установки в Google Play

(<https://play.google.com/store/apps/details?id=by.intoncore.SpeechRateMeter>).

## References

- [1] Bradlow D. Pisoni. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors // *The Journal of the Acoustic Society of America*. — 1999. — P. 74—85.
- [2] J. Murray M. Tracey. The effects of speaking rate on listener evaluations of native and foreign-accented speech // *Language Learning*. — 1998. — Vol. 48. — P. 159—182.
- [3] J.Yuan M. Liberman C. Cieri. Towards an integrated understanding of speaking rate in conversation // *Proceedings of Interspeech*. — 2006. — P. 541—544.
- [4] Kenneth K. The effect of foreign accent and speaking rate on native speaker comprehension // *Language Learning*. — 1981. — Vol. 38. — P. 561—613.
- [5] Pellowski Mark W. Pauses and the temporal structure of speech // E. Keller (Ed.) *Fundamentals of speech synthesis and speech recognition*. — 1994. — P. 41—62.
- [6] Pellowski Mark W. Speech-Language Pathologists'. Knowledge of Speaking Rate and Its Relationship to Stuttering // *Towson University, Towson, MD, Contemporary issues*. — 2010. — Vol. 37. — P. 50—57.
- [7] Tsao Ying-Chiao Weismer Gary Iqbal Kamran. Interspeaker variation in habitual speaking rate: Additional evidence // *Journal of Speech, Language, and Hearing Research*. — 2006. — Vol. 49. — P. 1156—1164.