УДК 004.934
doi:10.21685/2307-5538-2022-4-11

НОВЫЙ ПОДХОД СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА

А. К. Алимурадов[1], А. Ю. Тычков[2], П. П. Чураков[3], Д. С. Дудников[4]

[1, 2, 3, 4] Пензенский государственный университет, Пенза, Россия
[1] alansapfir@yandex.ru, [2] tychkov-a@mail.ru, [3] churakov.pp@gmail.com, [4] dmitriy.s.gmpf@gmail.com

Аннотация. *Актуальность и цели.* Точность сегментации речевых сигналов напрямую зависит от параметров, используемых для определения границ начала и окончания информативных фрагментов в слитном потоке речи. Целью работы является повышение эффективности сегментации «речь/пауза» за счет частотно-временного анализа речевых сигналов. Объектом исследования являются параметры, описывающие характеристики речи в частотной и временной областях. Предметом исследования является релевантность информативных параметров речевых сигналов задаче сегментации «речь/пауза». *Материалы и методы.* В работе использовались методы кратковременного анализа спектральных и энергетических характеристик речи на основе дискретного преобразования Фурье и энергетического оператора Тигера. Программная реализация предлагаемого способа была выполнена в среде математического моделирования © Matlab (MathWorks). *Результаты.* Предложен новый оригинальный подход сегментации «речь/пауза» на основе анализа значений средней частоты (в частотной области) и кратковременной энергии функции оператора Тигера (во временной области). Уникальностью предлагаемого подхода является вспомогательный алгоритм исправления ошибок сегментации «речь/пауза», разработанный на основе физиологических особенностей функционирования органов речевого аппарата при формировании слитного потока речи. Представлено краткое описание информативных параметров речевых сигналов, используемых для сегментации «речь/пауза» и подробно описан функционал предлагаемого подхода. Проведено исследование предлагаемого подхода на чистых и зашумленных речевых сигналах в сравнении с известными способами сегментации «речь/пауза». *Выводы.* В соответствии с полученными результатами исследования выявлено, что предлагаемый способ обеспечивает наилучшие результаты сегментации «речь/пауза» чистых и зашумленных речевых сигналов; использование отношения кратковременной энергии функции оператора Тигера к средней частоте в качестве информативного параметра обеспечивает максимальную релевантность к задаче сегментации; применение вспомогательного алгоритма исправления ошибочных статусов повышает эффективность сегментации.

Ключевые слова: обработка речевых сигналов, сегментации «речь/пауза», преобразование Фурье, энергетический оператор Тигера

Для цитирования: Алимурадов А. К., Тычков А. Ю., Чураков П. П., Дудников Д. С. Новый подход сегментации речевых сигналов на основе частотно-временного анализа // Измерение. Мониторинг. Управление. Контроль. 2022. № 4. С. 80–92. doi:10.21685/2307-5538-2022-4-11

NOVEL APPROACH BASED ON TIME-FREQUENCY ANALYSIS FOR SEGMENTATION OF SPEECH SIGNALS

A.K. Alimuradov[1], A.Yu. Tychkov[2], P.P. Churakov[3], D.S. Dudnikov[4]

[1, 2, 3, 4] Penza State University, Penza, Russia
[1] alansapfir@yandex.ru, [2] tychkov-a@mail.ru, [3] churakov.pp@gmail.com, [4] dmitriy.s.gmpf@gmail.com

Abstract. *Background.* The accuracy of speech signal segmentation depends directly on the parameters used to determine the boundaries of the beginning and the end of informative fragments in a continuous speech stream. The purpose of the work is to increase the efficiency of speech/pause segmentation due to the frequency-time analysis of speech signals. The research object is the parameters that describe speech characteristics in the frequency and time domains. The research subject is the relevance of the informative parameters of speech signals to the task of speech/pause segmentation. *Materials and methods.* The methods of short-term analysis of spectral and energy characteristics of speech based on the discrete Fourier transform and the energy Teager operator were used in the work. Software implementation of the proposed method was performed in ©MATLAB mathematical modeling environment produced by MathWorks *Results.* A novel original approach to speech/pause segmentation based on the analysis of the values of the mean frequency (in the frequency domain) and short-term energy of the Teager operator function (in the time domain) is

proposed. The proposed approach is unique due to an auxiliary algorithm to correct speech/pause segmentation errors, developed on the basis of physiological functioning of the respiratory apparatus organs during the formation of a continuous speech stream. A brief overview of speech signal informative parameters used for speech/pause segmentation has been presented, and the proposed approach performance has been detailed. The suggested approach has been compared with the known methods of speech/pause segmentation for pure and noisy speech signals. *Conclusions.* The research findings have evidenced the best results of speech/pause segmentation for pure and noisy speech signals being achieved by the methods based on the proposed approach; the ratio of the short-term energy of the Teager operator function to the mean frequency as an informative parameter ensuring maximum relevance to the segmentation problem; an auxiliary algorithm to correct false states enhancing the efficiency of segmentation.

Keywords: speech signal processing, speech/pause segmentation, Fourier transform, Teager energy operator

For citation: Alimuradov A.K., Tychkov A.Yu., Churakov P.P., Dudnikov D.S. Novel approach based on time-frequency analysis for segmentation of speech signals. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measuring. Monitoring. Management. Control*. 2022;(4):80–92. (In Russ.). doi:10.21685/2307-5538-2022-4-11

Введение

Сегментация речевых сигналов является одной из основных задач предварительной обработки практически для всех приложений в области речевых технологий. В общем понимании сегментация представляет собой деление слитного потока речи на квазистационарные по определенным признакам и характеристикам фрагменты. Точность сегментации речевых сигналов напрямую влияет на эффективность функционирования речевого приложения [1]. В зависимости от предназначения речевого приложения (распознавание речи, голосовое управление, идентификация диктора по голосу, оценка состояния здоровья человека по речи и др.) сегментация осуществляется на разных уровнях. На рис. 1 представлены три уровня сегментации речевых сигналов [2]. Для одних приложений достаточно сегментации «речь/пауза», для других может потребоваться сегментация на вокализованные и невокализованные фрагменты с последующим выделением периодов колебаний голосовых связок в тональных звуках.

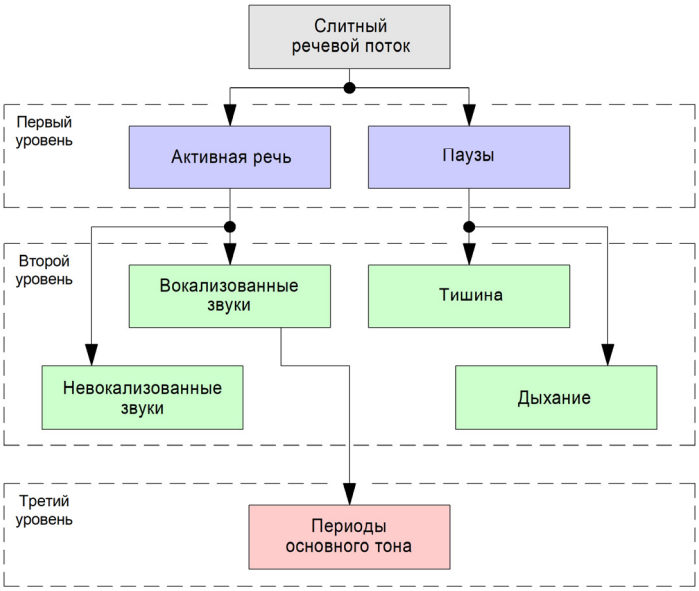

Рис. 1. Уровни сегментации речевых сигналов

Речевой сигнал имеет сложную структуру, амплитудные и частотные характеристики которого быстро изменяются во времени. Также важную роль в структуре речи играют индивидуальные физиологические особенности органов речевого аппарата и эмоциональное состояние человека. По этой причине задачи выбора и обоснования набора информативных параметров для сегментации речевых сигналов являются нетривиальными и представляют большую сложность и важность. Отсюда следует многообразие подходов к решению задач сегментации разного уровня, которые можно разделить на временные, частотные и частотно-временные способы.

Аналитический авторский обзор зарубежной и отечественной литературы выявил следующие широко применяемые способы сегментации [3]:

– во временной области посредством анализа значений одномерного расстояния Махаланобиса (One Dimensional Mahalanobis Distance, ODMD) [4], кратковременной энергии (Short Time Energy, STE) [5] и количества пересечения функции сигнала через нулевую ось (Zero-Crossing Rate, ZCR) [6];

– в частотной области посредством анализа значений мел-частотных кепстральных коэффициентов (Mel-Frequency Cepstral Coefficients, MFCC) [7] и линейно-частотных кепстральных коэффициентов (Linear-Frequency Cepstral Coefficients, LFCC) [8];

– в частотно-временной области посредством комбинирования частотных и временных способов.

Преимуществом способов сегментации во временной области являются быстродействие и малые требования к вычислительным ресурсам. Недостатком является низкая устойчивость к шумам и помехам. И наоборот, преимуществом способов сегментации в частотной области является помехоустойчивость, а недостатком – большие вычислительные затраты.

Исходя из вышеизложенного, можно сделать вывод об актуальности создания новых и совершенствования существующих подходов к решению задачи сегментации речевых сигналов на разных уровнях.

В данной статье представлен новый подход сегментации «речь/пауза» на основе частотно-временного анализа фрагментов речевых сигналов. Оригинальность подхода заключается в использовании в качестве информативных параметров для сегментации «речь/пауза» значений средней частоты [9] (в частотной области) и кратковременной энергии функции оператора Тигера (Teager Energy Operator, TEO) [10] (во временной области). Также уникальностью предлагаемого подхода является авторский вспомогательный алгоритм исправления ошибок сегментации «речь/пауза», разработанный на основе физиологических особенностей функционирования органов речевого аппарата при формировании слитного потока речи.

Данная научная статья подготовлена в рамках проекта № МД-1066.2022.4 «Исследование скрытых паттернов речевых сигналов и разработка способов обнаружения и классификации естественно выраженных психоэмоциональных состояний человека», финансируемого Советом по грантам Президента РФ. Статья является продолжением ранее опубликованных научных работ, посвященных разработке оригинальных способов обработки речевых сигналов для задачи обнаружения и классификации психоэмоциональных состояний человека по речи [11, 12].

Научная статья включает в себя шесть разделов. Второй раздел посвящен описанию информативных параметров речевых сигналов – средней частоты и кратковременной энергии функции оператора Тигера. В третьем разделе представлено описание предлагаемого подхода сегментации «речь/пауза». Четвертый и пятый разделы посвящены исследованию и анализу результатов исследования предлагаемого подхода. Шестой раздел статьи посвящен выводам и перспективам дальнейшей научной работы.

Материалы и методы

Средняя частота

Для сегментации речевых сигналов во временной области часто используется значение количества пересечения функции сигнала через нулевую ось. Частота пересечений может служить простейшей характеристикой спектральных свойств речевого сигнала, хотя обработка осуществляется во временной области [6]. Данный подход в полной мере справедлив для узкополосных сигналов. Речевой сигнал является широкополосным и функция среднего количества пересечений через нулевую ось может быть грубой для оценки спектральных свойств, особенно на фоне посторонних шумов. По этой причине целесообразно использовать значение средней частоты [9], которое в полной мере позволяет оценить спектральные свойства речи. В этом случае обработка речевых сигналов осуществляется в частотной области и вопрос повышения устойчивости к посторонним шумам решается.

Вычисление средней частоты осуществляется по следующей формуле:

$$F_{mean} = \frac{\sum_{k=1}^{F_s/2} x(k) \cdot PSD(k)}{\sum_{k=1}^{F_s/2} PSD(k)}, \quad (1)$$

где F_{mean} – средняя частота речевого сигнала со спектром мощности $PSD(k)$; $x(k)$ – исследуемый речевой сигнал в частотной области; k – дискретный отсчет сигнала в частотной области; F_s – частота дискретизации речевого сигнала.

Спектр мощности вычисляется с помощью быстрого преобразования Фурье с размерностью $K = 2048$ дискретных отсчетов. В отличие от количества пересечения функции сигнала через нулевую ось при вычислении средней частоты обязательно учитывается информация о распределении энергии каждого спектрального компонента (k) речевого сигнала (спектральное распределение энергии). Таким образом, значение средней частоты информативнее, так как включает в себя данные о спектральных и энергетических характеристиках речи.

Кратковременная энергия

Для сегментации речевых сигналов разного уровня используется значение кратковременной энергии, которое представляет собой сумму квадратов амплитуд дискретных отсчетов сигнала во временной области:

$$STE = \frac{1}{N} \sum_{n=1}^{N} \left[x(n) \right]^2, \quad (2)$$

где $x(n)$ – исследуемый сигнал во временной области; n – дискретный отсчет сигнала во временной области; N – количество дискретных отсчетов в исследуемом речевом сигнале.

Сегментация «речь/пауза» на основе анализа значений кратковременной энергии построена на предположении, что энергия вокализованных и невокализованных звуков больше, чем энергия участков тишины и дыхания.

Энергетический оператор Тигера

Математический аппарат энергетического оператора Тигера был предложен Х. М. Тигером (H.M. Teager) в рамках научного исследования, посвященного моделированию нелинейного процесса воспроизведения речи. Оператор Тигера – это дифференциальный энергетический оператор второго порядка, позволяющий вычислять энергетические характеристики сигнала [10]. Для дискретных сигналов функция оператора Тигера вычисляется по следующей формуле:

$$TEO(n) = x(n)^2 - x(n-1) \cdot x(n+1). \quad (3)$$

Оператор Тигера обеспечивает отличное временное разрешение, поскольку для вычисления энергии в каждый момент времени требуется всего три дискретных значения. Математический аппарат оператора обладает эффективностью, простотой и хорошей восприимчивостью к мгновенному изменению амплитуды и частоты сигнала. В области обработки речевых сигналов энергетический оператор Тигера также успешно используется для повышения эффективности сегментации разного уровня [13, 14].

Описание нового подхода

На рис. 2 представлена последовательность этапов обработки речевых сигналов предлагаемого подхода сегментации «речь/пауза» на основе частотно-временного анализа. Суть предлагаемого подхода заключается в сегментации речевых сигналов на фрагменты (длительностью 10 мс) (этап 2), вычислении информативных параметров (средней частоты и кратковременной энергии функции оператора Тигера) (этап 3), определении пороговых значений (этап 4) и статусов «речь/пауза» фрагментов (этап 5) с последующим исправлением ошибочных статусов (этап 6) и определением результатов итоговой сегментации (этап 7).

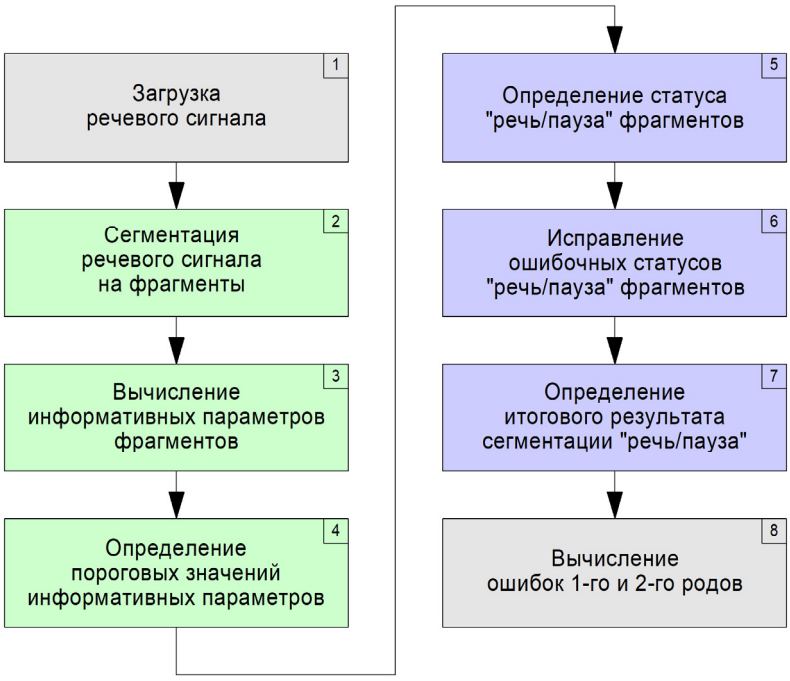

Рис. 2. Последовательность этапов обработки речевых сигналов предлагаемого подхода сегментации «речь/пауза»

Рассмотрим подробнее некоторые этапы обработки предлагаемого подхода. На рис. 3 представлена иллюстрация, поясняющая процесс определения пороговых значений информативных параметров (P) – средней частоты (Fmean) и кратковременной энергии функции оператора Тигера (STETEO).

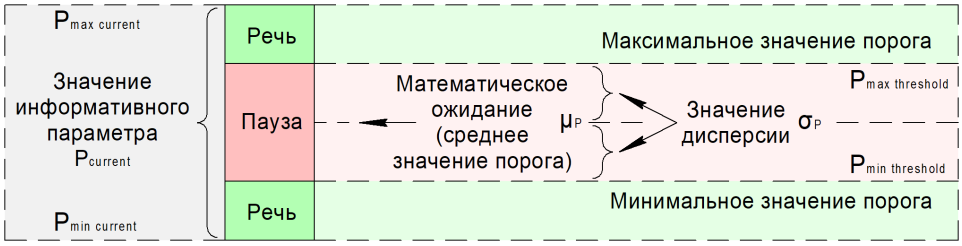

Рис. 3. Определение пороговых значений информативных параметров

Для определения диапазона значений информативного параметра, соответствующего статусу «пауза» (между минимальным ($P_{\min\ threshold}$) и максимальным ($P_{\max\ threshold}$) уровнями порога) используется участок вынужденной начальной паузы, которую человек выдерживает перед произношением (воспроизведением речи). Это связано с физиологией речевого аппарата человека. Обычно вынужденная пауза имеет длительность не более 200 мс и соответствует тишине с фоновым шумом.

Для определения $P_{\min\ threshold}$ и $P_{\max\ threshold}$ вычисляется математическое ожидание (μ_P) и дисперсия (σ_P) значений информативного параметра первых 20 фрагментов вынужденной начальной паузы (при длительности анализируемых фрагментов 10 мс):

$$\mu_P = \frac{1}{20}\sum_{s=1}^{20} Pcurrent_s, \tag{4}$$

$$\sigma_P = \sqrt{\frac{1}{20}\sum_{s=1}^{20}\left(Pcurrent_s - \mu_P\right)^2}, \tag{5}$$

где $Pcurrent_s$ – текущее значение информативного параметра фрагментов речевого сигнала; s – номер фрагмента.

Определение статуса «речь/пауза» фрагментов речевого сигнала заключается в проверке следующих условий:

$$\left|Pcurrent_s - \mu_P\right| \geq U \times \sigma_P, \quad (6)$$

где U – значение коэффициента порога.

Коэффициент порога U введен для расширений динамического диапазона значений информативного параметра между $P_{\min\ threshold}$ и $P_{\max\ threshold}$. Коэффициент принимает значения от 1 до 15 для грубой настройки порога и от 0,1 до 1,0 для мягкой настройки.

Если разница по модулю между текущим значением $P_{current}$ и средним значением порога μ_P информативного параметра больше или равна значению дисперсии σ_P, то фрагмент соответствует речи. И наоборот, если условие не выполняется, то фрагмент соответствует паузе.

На рис. 4 представлена иллюстрация, поясняющая принцип работы вспомогательного алгоритма исправления ошибочных статусов «речь/пауза». Вспомогательный алгоритм исправления ошибок сегментации основан на физиологических особенностях функционирования органов речевого аппарата при формировании слитного потока речи. Поток слитной речи представляет собой нестационарный случайный процесс, быстро изменяющийся во времени. Однако из-за инерции органов речевого аппарата характеристики слитной речи не могут изменяться мгновенно. Данная физиологическая особенность обеспечивает кратковременную стационарность речи длительностью не более 40 мс, т.е. вне зависимости от скорости изменения амплитудных и частотных характеристик состояние речи на участках длительностью 40 мс (4 фрагмента по 10 мс) будет неизменным. В левой части рис. 4 представлены варианты статусов «речь/пауза» фрагментов до исправления, в правой части – после исправления. Как видно из рисунка, вспомогательный алгоритм обеспечивает исправление ошибок сегментации «речь/пауза» на участках длительностью 40 мс, учитывая кратковременную стационарность речи.

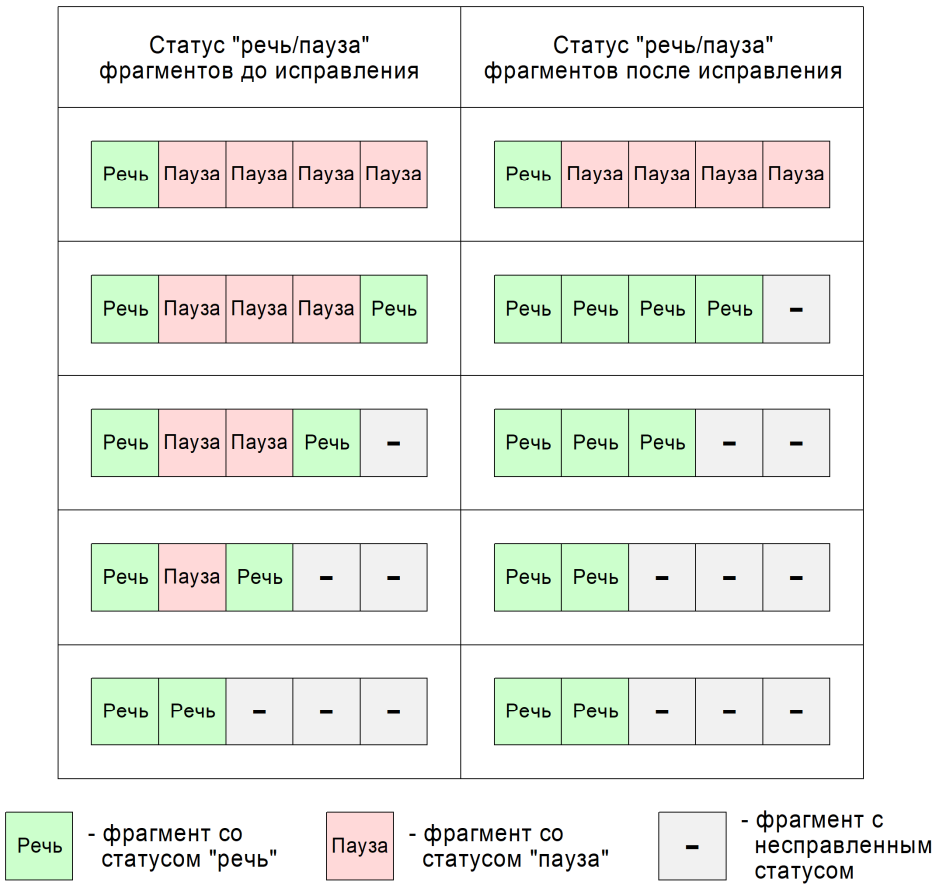

Рис. 4. Принцип работы вспомогательного алгоритма исправления ошибочных статусов «речь/пауза»

Исследование нового подхода

Исследование предлагаемого подхода заключалось в оценке эффективности сегментации «речь/пауза» чистых речевых сигналов в зависимости от значений коэффициента порога и зашумленных белым шумом речевых сигналов с разными отношениями сигнал/шум (ОСШ).

Для исследования предлагаемого подхода сформирована речевая база данных. Речевые сигналы длительностью не более 10 с были зарегистрированы посредством специализированных методики и технических средств. Подготовленные дикторы в количестве 20 человек воспроизводили речь на русском языке, содержащую публицистический текст (30 записей), текст из литературного произведения (30 записей) и счет чисел от 0 до 100 (10 записей). Общее количество зарегистрированных речевых сигналов – 1400.

Зашумление чистых речевых сигналов осуществлялось программно посредством наложения сгенерированного белого шума в программе аудиоредактирования «Audacity». База зашумленных речевых сигналов сформирована с различными значениями ОСШ от –5 до 15 дБ с шагом 5 дБ.

Эффективность сегментации оценивалась в соответствии с полученными значениями ошибок 1-го (α) и 2-го (β) рода. Основной задачей сегментации «речь/пауза» считалось определение фрагментов речи среди всех фрагментов речевого сигнала. Ошибкой α считалась ситуация, когда фрагменту речи присваивался статус «пауза». Ошибкой β считалась ситуация, когда фрагменту паузы присваивался статус «речь». Ошибки определялись по результатам сопоставления полученных данных сегментации с данными сегментации, осуществленной вручную.

Сегментация «речь/пауза» осуществлялась посредством предлагаемого подхода на основе анализа:

– средней частоты и применения вспомогательного алгоритма исправления ошибочных статусов «речь/пауза» (Fmean+correction);

– отношения кратковременной энергии функции оператора Тигера к средней частоте (STETEO/Fmean);

– отношения кратковременной энергии функции оператора Тигера к средней частоте и применения вспомогательного алгоритма исправления ошибочных статусов «речь/пауза» (STETEO/Fmean+correction).

Результаты оценивались в сравнении с известными способами сегментации «речь/пауза» на основе анализа:

– количества пересечений функции сигнала через нулевую ось (ZCR);

– кратковременной энергии (STE);

– одномерного расстояния Махаланобиса (ODMD);

– количества пересечений сигнала через нулевую ось функции оператора Тигера (ZCRTEO);

– кратковременной энергии функции оператора Тигера (STETEO);

– количества пересечений сигнала через нулевую ось и кратковременной энергии (ZCR+STE);

– количества пересечений сигнала через нулевую ось и кратковременной энергии функции оператора Тигера (ZCRTEO + STETEO).

В табл. 1–4 представлены результаты проведенного исследования – усредненные значения ошибок α и β в зависимости от значений коэффициента порога, полученные по результатам сегментации чистых и зашумленных речевых сигналов способами ZCR, STE, ODMD, ZCRTEO, STETEO, ZCR + STE, ZCRTEO + STETEO, Fmean+correction, STETEO/Fmean и STETEO/Fmean + correction.

Анализ результатов исследования

В соответствии с данными в табл. 1 и 2 наилучший результат сегментации «речь/пауза» с ошибками α = 2,08 % и β = 2,12 % при U = 4 достигается способом STETEO/Fmean + + correction на основе предлагаемого подхода. Максимально близкими по эффективности являются способы STETEO и ZCRTEO + STETEO с ошибками α = 2,08 % и β = 3,88 % при U = 14. Однако здесь важно отметить, что наилучший результат сегментации во всем диапазоне зна-

чений коэффициента порога от 1 до 15 обеспечивается только способом STETEO/Fmean + + correction. Объясняется это максимальной релевантностью информативного параметра (отношение кратковременной энергии функции оператора Тигера к средней частоте) к задаче сегментации и эффективностью применения вспомогательного алгоритма исправления ошибочных статусов «речь/пауза».

Таблица 1

Усредненные значения ошибок α и β, полученные по результатам сегментации чистых речевых сигналов способами ZCR, STE, ODMD, ZCRTEO и STETEO

U	ZCR		STE		ODMD		ZCRTEO		STETEO	
	α, %	β, %	α, %	β, %	α, %	β, %	α, %	β, %	α, %	β, %
1	7,16	35,98	5,31	21,52	21,71	1,59	7,85	29,45	0,23	34,57
2	24,48	14,46	7,39	10,05	21,71	1,59	17,09	9,70	0,46	21,52
3	36,95	8,29	9,47	5,82	21,71	1,59	29,10	3,53	0,69	16,58
4	56,12	4,23	10,86	3,35	21,71	1,59	43,88	2,12	0,69	13,40
5	73,67	2,65	11,32	2,47	21,71	1,59	59,58	1,94	0,69	10,94
6	81,06	1,94	12,47	1,94	21,71	1,59	79,45	1,59	0,92	8,99
7	86,37	1,94	13,63	1,94	21,71	1,59	98,15	1,59	0,92	7,41
8	87,53	1,76	15,24	1,59	21,71	1,59	100,00	1,59	1,15	6,00
9	87,99	1,59	16,40	1,59	21,71	1,59	100,00	1,59	1,15	5,47
10	88,45	1,59	17,09	1,59	21,71	1,59	100,00	1,59	1,39	4,59
11	89,38	1,59	18,01	1,59	21,71	1,59	100,00	1,59	1,62	4,41
12	89,84	1,59	18,48	1,59	21,71	1,59	100,00	1,59	2,08	4,23
13	89,84	1,59	19,17	1,59	21,71	1,59	100,00	1,59	2,08	4,23
14	90,30	1,59	19,86	1,59	21,71	1,59	100,00	1,59	2,08	3,88
15	90,30	1,59	20,09	1,59	21,71	1,59	100,00	1,59	2,08	3,88

Таблица 2

Усредненные значения ошибок α и β, полученные по результатам сегментации способами ZCR+STE, ZCRTEO+STETEO, Fmean+correction, STETEO/Fmean и STETEO/Fmean+correction

U	ZCR+STE		ZCRTEO+STETEO		Предлагаемый подход					
					Fmean + correction		STETEO/Fmean		STETEO/Fmean + + correction	
	α, %	β, %	α, %	β, %	α, %	β, %	α, %	β, %	α, %	β, %
1	0,69	51,50	0,23	50,62	4,62	17,99	0,92	23,46	0,46	8,82
2	1,62	22,75	0,46	27,34	11,32	8,29	1,62	16,05	1,39	4,59
3	4,16	12,52	0,46	18,34	19,40	2,65	2,31	11,46	1,62	4,06
4	6,00	6,00	0,46	13,93	31,41	2,29	2,77	8,29	2,08	2,12
5	6,93	3,53	0,69	11,29	40,65	1,59	3,46	6,35	2,31	2,12
6	8,08	2,29	0,92	8,99	46,65	1,59	3,93	5,29	2,31	2,12
7	10,39	2,29	0,92	7,41	55,43	1,59	4,62	4,94	2,31	2,12
8	12,24	1,76	1,15	6,00	70,90	1,59	5,31	4,06	3,70	2,12
9	13,63	1,59	1,15	5,47	79,22	1,59	5,54	3,70	4,16	2,12
10	14,55	1,59	1,39	4,59	87,53	1,59	6,24	3,17	4,39	1,59
11	16,40	1,59	1,62	4,41	87,53	1,59	6,47	3,00	4,39	1,59
12	17,32	1,59	2,08	4,23	89,61	1,59	6,93	2,82	5,08	1,59
13	18,01	1,59	2,08	4,23	89,61	1,59	7,39	2,47	6,93	1,59
14	18,71	1,59	2,08	3,88	89,61	1,59	8,08	2,47	6,93	1,59
15	18,94	1,59	2,08	3,88	89,61	1,59	8,31	2,29	7,16	1,59

При сравнении результатов сегментации способов Fmean+correction (α = 11,32 % и β = 8,29 % при $U = 3$), STETEO/Fmean (α = 4,62 % и β = 4,94 % при $U = 7$) и STETEO/Fmean + + correction (α = 2,08 % и β = 2,12 % при $U = 4$) на основе предлагаемого подхода необходимо отметить повышение эффективности при совместном анализе значений средней частоты и

кратковременной энергии функции оператора Тигера с применением вспомогательного алгоритма исправления ошибочных статусов «речь/пауза».

В соответствии с данными в табл. 3 и 4 наилучшие результаты сегментации «речь/пауза» зашумленных речевых сигналов также обеспечиваются способами на основе предлагаемого подхода:

- ОСШ –5 дБ, STETEO/Fmean + correction, α = 10,86 % и β = 10,58 % при $U = 2$;
- ОСШ 0 дБ, STETEO/Fmean + correction, α = 9,70 % и β = 3,17 % при $U = 2$;
- ОСШ 5 дБ, Fmean + correction, α = 8,78 % и β = 3,35 % при $U = 1$;
- ОСШ 10 дБ, STETEO/Fmean + correction, α = 8,55 % и β = 1,59 % при $U = 3$;
- ОСШ 15 дБ, STETEO/Fmean + correction, α = 4,16 % и β = 5,82 % при $U = 2$.

Таблица 3

Усредненные значения α и β, полученные по результатам сегментации чистых и зашумленных речевых сигналов способами ZCR, STE, ODMD, ZCRTEO и STETEO

ОСШ, дБ	ZCR			STE			ODMD			ZCRTEO			STETEO		
	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U
–5	37,64	30,16	1	39,03	2,65	3	100,00	1,59	1	50,12	42,33	1	26,56	15,52	1
0	25,17	25,57	2	21,71	9,88	3	100,00	1,59	1	46,42	35,63	1	20,32	11,64	1
5	20,32	4,94	2	13,86	2,12	3	60,28	1,59	1	36,03	31,39	1	11,78	3,35	2
10	18,94	4,06	2	9,70	3,88	3	42,03	1,59	1	31,87	44,62	1	12,93	2,82	2
15	13,16	8,82	2	10,16	4,59	3	35,57	1,59	1	36,49	32,98	1	9,70	3,53	2
Чистый сигнал	24,48	14,46	2	9,47	5,82	3	21,71	1,59	1	17,09	9,70	2	2,08	3,88	14

Таблица 4

Усредненные значения ошибок α и β, полученные по результатам сегментации чистых и зашумленных речевых сигналов способами ZCR + STE, ZCRTEO + STETEO, Fmean + correction, STETEO/Fmean и STETEO/Fmean+correction

ОСШ, дБ	ZCR + STE			ZCRTEO + + STETEO			Предлагаемый подход								
							Fmean + + correction			STETEO/Fmean			STETEO/Fmean + + correction		
	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U	α, %	β, %	U
–5	29,79	15,17	2	22,63	25,75	2	23,56	12,52	1	27,71	9,35	2	10,86	10,58	2
0	24,25	5,47	4	18,48	14,99	2	26,33	3,35	1	21,48	6,53	2	9,70	3,17	2
5	13,86	2,47	3	12,93	3,00	3	8,78	3,35	1	11,78	5,11	3	10,39	1,76	3
10	9,70	3,88	3	11,78	4,06	3	6,93	14,64	1	10,86	4,41	2	8,55	1,59	3
15	9,93	4,59	3	9,70	3,53	3	6,47	14,64	1	6,70	9,70	2	4,16	5,82	2
Чистый сигнал	6,93	3,53	5	2,08	3,88	14	11,32	8,29	2	4,62	4,94	7	2,08	2,12	4

Помехоустойчивость способов сегментации «речь/пауза» на основе предлагаемого подхода достигается за счет анализа информативных параметров речевых сигналов в частотно-временной области. Особенно важно отметить хорошие результаты сегментации при низких значениях ОСШ –5 и 0 дБ. Максимально близкие по помехоустойчивости способы обеспечивают результаты сегментации «речь/пауза» в 1,5–2 раза хуже, чем способы на основе предлагаемого подхода.

На рис. 5 представлен пример наилучшей сегментации «речь/пауза» одного зашумленного речевого сигнала из сформированной речевой базы данных. Светло-серым цветом отмечен зашумленный речевой сигнал, темно-серым – исходный чистый сигнал. Линией синего цвета отмечен результат сегментации, выполненной вручную. Линией красного цвета отмечен результат сегментации, выполненной способом STETEO/Fmean + correction на основе анализа отношения кратковременной энергии функции оператора Тигера к средней частоте и применения вспомогательного алгоритма исправления ошибочных статусов повышает эффективность сегментации «речь/пауза».

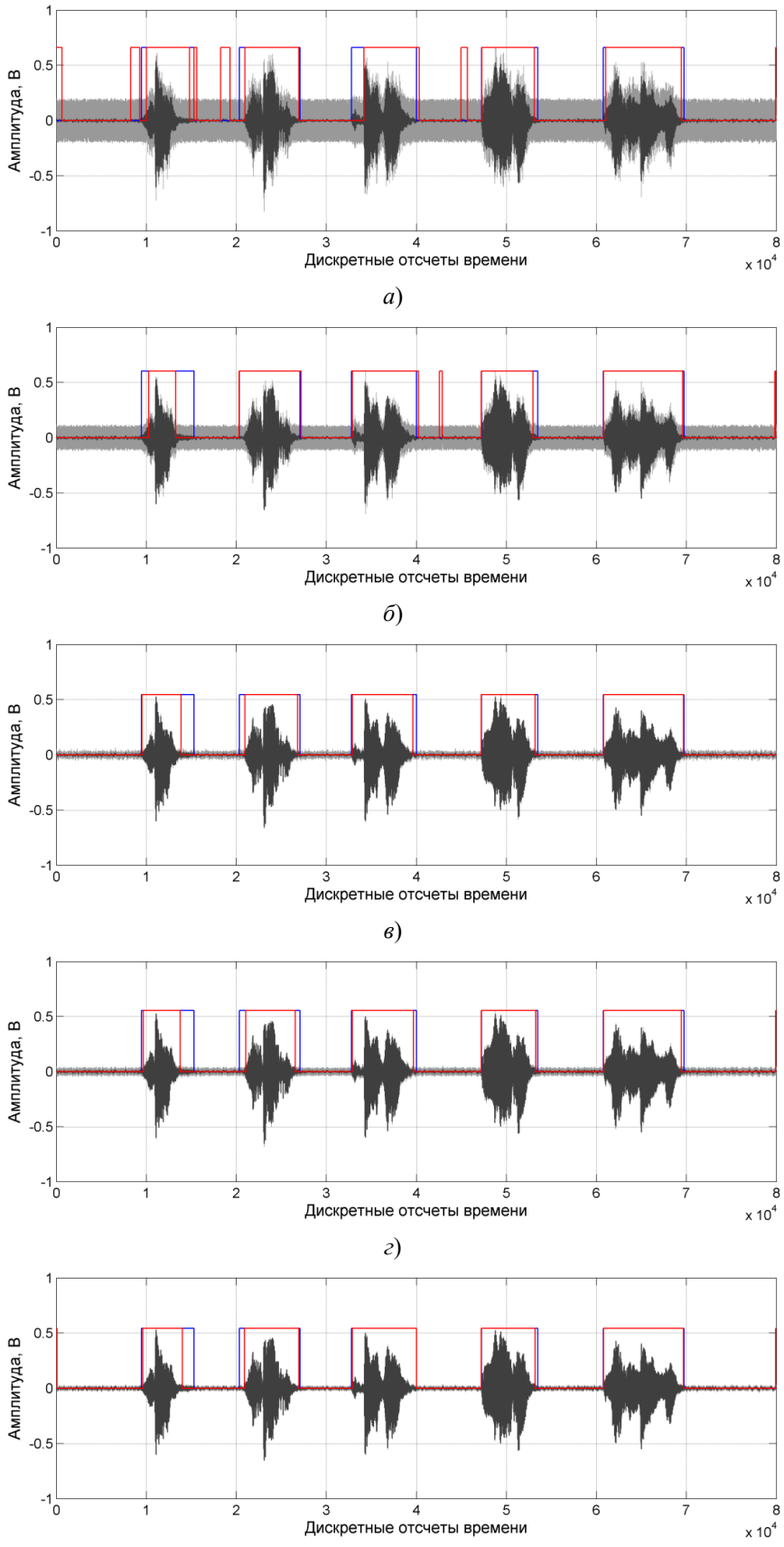

Рис. 5. Пример сегментации «речь/пауза» зашумленного речевого сигнала способом STETEO/Fmean+correction на основе предложенного подхода:
а – ОСШ = –5 дБ; *б* – ОСШ = 0 дБ; *в* – ОСШ = 5 дБ; *г* – ОСШ = 10 дБ; *д* – ОСШ = 15 дБ

Заключение

В соответствии с анализом результатов исследования сделаны следующие краткие выводы:

1. Предлагаемый подход на основе частотно-временного анализа фрагментов речевых сигналов обеспечивает наилучший результат в сравнении с известными способами сегментации «речь/пауза».

2. Использование отношения кратковременной энергии функции оператора Тигера к средней частоте в качестве информативного параметра обеспечивает максимальную релевантность к задаче сегментации «речь/пауза».

3. Применение вспомогательного алгоритма исправления ошибочных статусов повышает эффективность сегментации «речь/пауза».

4. Наилучшие результаты сегментации «речь/пауза» во всем диапазоне значений коэффициента порога обеспечиваются только способами на основе предлагаемого подхода.

5. Способы на основе предлагаемого подхода обладают наилучшей помехоустойчивостью в сравнении с известными способами сегментации «речь/пауза».

В перспективе коллективом авторов планируется провести дополнительные исследования быстродействия способов сегментации «речь/пауза» на основе предлагаемого подхода, а также исследовать устойчивость предлагаемого подхода к коричневому и розовому шумам.

Список литературы

1. Schuller B. W., Batliner A. M. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. New York : Wiley, 2013. P. 344.
2. Huang X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Developmen. New Jersey : Prentice Hall, 2001. 980 p.
3. Алимурадов А. К., Тычков А. Ю., Чураков П. П. [и др.]. Способ повышения эффективности сегментации речь/пауза на основе метода декомпозиции на эмпирические моды // Известия высших учебных заведений. Поволжский регион. Технические науки. 2021. № 2. С. 24–43.
4. Duda R. O., Hart P. E., Strok D. G. Pattern Classification. 2nd ed. New Jersey : A Wiley-Interscience Publ. John Wiley & Sons, Inc., 2001. 688 p.
5. Childers D. G., Hand M., Larar J. M. Silent and voiced/unvoied/ mixed excitation (four-way), classification of speech // IEEE Transaction on ASSP. 1989. Vol. 37, № 11. P. 1771–1774.
6. Atal B., Rabiner L. R. A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition // IEEE Trans. Acoust. Speech Signal Process. 1976. Vol. 24, № 3. P. 201–212.
7. Martin A., Charlet D., Mauuary L. Robust speech/non-speech detection using LDA applied to MFCC // IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221) (ICASSP2001) (May 7–11, 2001). Salt Lake City, UT, USA, 2001. Vol. 1. P. 237–240.
8. Hlavnička J., Čmejla R., Tykalová T. et al. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder // Scientific Reports. 2017. Vol. 7. 13 p.
9. Sharma R., Prasanna S. R. M. Characterizing glottal activity from speech using empirical mode decomposition // 21st National Conference on Communications (NCC) (27 February – 1 March, 2015). Mumbai, India, 2015. P. 1–6.
10. Kaiser J. F. On a simple algorithm to calculate the 'energy' of a signal // International Conference on Acoustics, Speech, and Signal Processing (April 3–6, 1990). Albuquerque, NM, USA, 1990. Vol. 2. P. 381–384.
11. Алимурадов А. К., Тычков А. Ю., Чураков П. П. [и др.]. Способ обработки речевых сигналов на основе метода декомпозиции на эмпирические моды // Измерение. Мониторинг. Управление. Контроль. 2022. № 2 (40). С. 75–89.
12. Алимурадов А. К. Способ сегментации речь/пауза на основе энергетического оператора Тигера // Модели, системы, сети в экономике, технике, природе и обществе. 2021. № 4. С. 52–63.
13. Zhuikov V. Ya., Kharchenko A. N. Algorithm for speech signal segments classification // Electronics and Communications. Special issue on Electronics and Nanotechnology. 2009. Part 1, № 2-3. P. 130–137.
14. Bahoura M., Rouat J. Wavelet speech enhancement based on the teager energy operator // IEEE Signal Processing Letter. 2001. Vol. 8, № 1. P. 10–12.

References

1. Schuller B.W., Batliner A.M. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. New York: Wiley, 2013:344.

2. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Developmen*. New Jersey: Prentice Hall, 2001:980.
3. Alimuradov A.K., Tychkov A.Yu., Churakov P.P. et al. A way to increase the efficiency of speech/pause segmentation based on the method of decomposition into empirical modes. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = News of Higher Educational Institutions. Volga region. Technical sciences*. 2021;(2):24–43. (In Russ.)
4. Duda R.O., Hart P.E., Strok D.G. *Pattern Classification. 2nd ed*. New Jersey: A Wiley-Interscience Publ. John Wiley & Sons, Inc., 2001:688.
5. Childers D.G., Hand M., Larar J.M. Silent and voiced/unvoied/ mixed excitation (four-way), classification of speech. *IEEE Transaction on ASSP*. 1989;37(11):1771–1774.
6. Atal B., Rabiner L.R. A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition. *IEEE Trans. Acoust. Speech Signal Process*. 1976;24(3):201–212.
7. Martin A., Charlet D., Mauuary L. Robust speech/non-speech detection using LDA applied to MFCC. *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221) (ICASSP2001) (May 7–11, 2001)*. Salt Lake City, UT, USA, 2001;1:237–240.
8. Hlavnička J., Čmejla R., Tykalová T. et al. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. *Scientific Reports*. 2017;7:13.
9. Sharma R., Prasanna S.R.M. Characterizing glottal activity from speech using empirical mode decomposition. *21st National Conference on Communications (NCC) (27 February – 1 March, 2015)*. Mumbai, India, 2015:1–6.
10. Kaiser J.F. On a simple algorithm to calculate the 'energy' of a signal. *International Conference on Acoustics, Speech, and Signal Processing (April 3–6, 1990)*. Albuquerque, NM, USA, 1990;2:381–384.
11. Alimuradov A.K., Tychkov A.Yu., Churakov P.P. et al. Method of processing speech signals based on the method of decomposition into empirical modes. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measuring. Monitoring. Management. Control.* 2022;(2):75–89. (In Russ.)
12. Alimuradov A.K. Segmentation method speech/pause based on the energy operator Tigera. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, systems, networks in economics, technology, nature and society*. 2021;(4):52–63. (In Russ.)
13. Zhuikov V.Ya., Kharchenko A.N. Algorithm for speech signal segments classification. *Electronics and Communications. Special issue on Electronics and Nanotechnology*. 2009;1(2-3):130–137.
14. Bahoura M., Rouat J. Wavelet speech enhancement based on the teager energy operator. *IEEE Signal Processing Letter*. 2001;8(1):10–12.

Информация об авторах / Information about the authors

Алан Казанферович Алимурадов
кандидат технических наук,
директор студенческого научно-производственного бизнес-инкубатора,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: alansapfir@yandex.ru

Александр Юрьевич Тычков
доктор технических наук,
профессор кафедры радиотехники и радиоэлектронных систем,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: tychkov-a@mail.ru

Петр Павлович Чураков
доктор технических наук,
профессор кафедры информационно-измерительной техники и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: churakov-pp@mail.ru

Alan K. Alimuradov
Candidate of technical sciences,
director of the student research and production business incubator,
Penza State University
(40 Krasnaya street, Penza, Russia)

Alexander Yu. Tychkov
Doctor of technical sciences,
professor of the sub-department of radio engineering and radioelectronic systems,
Penza State University
(40 Krasnaya street, Penza, Russia)

Petr P. Churakov
Doctor of technical sciences,
professor of the sub-department of information and measuring equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Дмитрий Сергеевич Дудников
студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: dmitriy.s.gmpf@gmail.com

Dmitriy S. Dudnikov
Student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.

Поступила в редакцию/Received 14.04.2022
Поступила после рецензирования/Revised 16.05.2022
Принята к публикации/Accepted 20.06.2022