ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ МЕДИЦИНСКОГО НАЗНАЧЕНИЯ

MEDICAL DEVICES, SYSTEMS AND PRODUCTS

УДК 004.934
doi:10.21685/2307-5538-2021-3-10

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ЭНЕРГЕТИЧЕСКОГО ОПЕРАТОРА ТИГЕРА

А. К. Алимурадов

Пензенский государственный университет, Пенза, Россия
alansapfir@yandex.ru

Аннотация. *Актуальность и цели.* Сегментация речевых сигналов представляет собой обнаружение границ начала и окончания участков вокализованной, невокализованной речи и пауз. Точное обнаружение границ не только повышает качество сегментации речевого сигнала, но и уменьшает количество вычислительных операций. Целью работы является повышение эффективности сегментации на основе энергетического анализа речевых сигналов с помощью энергетического оператора Тигера. *Материалы и методы.* В работе использовался дифференциальный энергетический оператор Тигера 2-го порядка, позволяющий вычислять энергетические характеристики сигнала. Оператор Тигера обладает простотой, эффективностью и хорошей восприимчивостью к изменению амплитуды и частоты сигнала. Программная реализация способа была выполнена в среде математического моделирования © Matlab (MathWorks). *Результаты.* Разработан модернизированный способ сегментации речевых сигналов, обеспечивающий повышение эффективности обнаружения вокализованных, невокализованных участков и пауз. Суть способа заключается в энергетическом анализе фрагментов речевого сигнала с помощью энергетического оператора Тигера; анализе количества пересечений через нулевую ось и кратковременной энергии функции энергетической характеристики. Проведено исследование, в рамках которого оценивалась эффективность и помехоустойчивость модернизированного способа в сравнении с известными способами сегментации. *Выводы.* В соответствии с полученными результатами исследований выявлено, что за счет хорошей восприимчивости энергетического оператора Тигера к резким изменениям амплитуды и частоты сигнала модернизированный способ обеспечивает повышение эффективности сегментации на 2,97 и 2,49 % для ошибок 1-го и 2-го рода соответственно.

Ключевые слова: обработка речи, сегментация речи, вокализованная и невокализованная речь, паузы, энергетический оператор Тигера

Для цитирования: Алимурадов А. К. Повышение эффективности сегментации речевых сигналов на основе энергетического оператора тигера // Измерения. Мониторинг. Управление. Контроль. 2021. № 3. С. 80–92. doi:10.21685/2307-5538-2021-3-10

ENHANCEMENT OF SPEECH SIGNAL SEGMENTATION USING TEAGER ENERGY OPERATOR

A. K. Alimuradov

Penza State University, Penza, Russia
alansapfir@yandex.ru

Abstract. *Background.* Speech signal segmentation is detection of the boundaries of the beginning and the end of sections of voiced and unvoiced speech, and pauses. Accurate detection of the boundaries both improves the quality of

speech signal segmentation, and reduces the number of computational operations. The aim of the work is to improve the efficiency of segmentation based on the energy analysis of speech signals using the Teager energy operator. *Materials and methods.* The second-order differential Teager energy operator, which makes it possible to estimate the energy characteristics of a signal, was used in this work. The Teager operator is simple, efficient, and highly susceptible to changes in signal amplitude and frequency.The software implementation of the method was performed in ©MATLAB (MathWorks) mathematical modeling environment. *Results.* An improved method for speech signal segmentation, providing an increase in the efficiency of detecting voiced and unvoiced areas, and pauses, has been developed. The nature of the method is the energy analysis of speech signal fragments using the Teager energy operator; analysis of zero-crossing rate and short-term energy of the energy characteristic function. Research to assess the efficiency and noise robustness of the improved method over the known segmentation methods, was carried out. *Conclusions.* In accordance with the obtained research results, it was revealed that due to the good susceptibility of the Teager energy operator to sharp changes in signal amplitude and frequency, the improved method provides an increase in the segmentation efficiency by 2.97 % and 2.49 % for the 1st and 2nd kind errors, respectively.

Keywords: speech processing, speech segmentation, voiced and unvoiced speech, pauses, Teager energy operator

For citation: Alimuradov A.K. Enhancement of speech signal segmentation using teager energy operator. *Izmereniya. Monitoring. Upravlenie. Kontrol′ = Measurements. Monitoring. Management. Control.* 2021;(3): 80–92. (In Russ.). doi:10.21685/2307-5538-2021-3-10

Введение

Сегментация речевых сигналов представляет собой обнаружение границ начала и окончания участков вокализованной, невокализованной речи и пауз. На сегодняшний день задача сегментации речевых сигналов решается разными способами, которые можно разделить на частотные и временные. К временным относятся способы на основе анализа количества пересечения через нулевую ось (Zero-Crossing Rate, ZCR) [1], отклонения автокорреляционной функции (Autocorrelation Function, ACR) [2], кратковременной энергии (Short Time Energy, STE) [3], а также одномерного расстояния Махаланобиса (One Dimensional Mahalanobis Distance, ODMD) [4]. К частотным относятся способы на основе анализа мел-частотных кепстральных коэффициентов (Mel-Frequency Cepstral Coefficients, MFCC) [5] и линейно-частотных кепстральных коэффициентов (Linear-Frequency Cepstral Coefficients, LFCC) [6].

В статье представлен способ, позволяющий повысить эффективность сегментации речевых сигналов за счет применения энергетического оператора Тигера (Teager Energy Operator, TEO). Предлагаемый способ представляет собой модернизацию существующего способа сегментации на основе анализа ZCR и STE. Модернизация включает в себя вычисление энергетической характеристики фрагментов речевых сигналов на основе TEO с последующим анализом значений ZCR и STE.

Статья является результатом научной работы, посвященной разработке эффективных алгоритмов и способов обработки речевых сигналов на основе новых частотно-временных методов анализа [7–9].

Структурно статья состоит из семи разделов. Второй и третий разделы посвящены краткому обзору известных способов сегментации речевых сигналов, а также вычислению энергетической характеристики речи с помощью TEO. Четвертый и пятый разделы посвящены описанию и исследованию модернизированного способа. В шестом разделе представлен анализ результатов исследований. Последний раздел посвящен выводам и перспективам дальнейшей научной работы.

Сегментация речевых сигналов

Способы сегментации речевых сигналов на основе анализа ZCR и STE применяются ограниченно. Ограничения связаны с выбором и обоснованием корректных пороговых значений, соответствующих вокализованной, невокализованной речи и паузам.

Функция ZCR основана на сравнении знаков соседних дискретных отсчетов времени и определяется по следующей формуле:

$$\mathrm{ZCR}_s = 0{,}5\sum_{n=1}^{N-1}\left|\mathrm{sgn}\left(x\left(s-1\right)N+n+1\right)-\mathrm{sgn}\left(x\left(s-1\right)N+n\right)\right|, \quad (1)$$

где $x(n)$ – исследуемый сигнал; n – дискретный отсчет времени; s – номер фрагмента; N – количество дискретных отсчетов в исследуемом фрагменте; $sgn(x)$ – знаковая функция ($\mathrm{sgn}(x) = 1$ при $x \geq 0$ и $\mathrm{sgn}(x) = -1$ при $x \leq 0$).

Функция STE представляет собой сумму квадратов амплитуд дискретных отсчетов сигнала для короткой последовательности (фрагмента) и определяется по следующей формуле:

$$E_s = \sum_{n=1}^{N} \left[x(s-1)N + n \right]^2 . \quad (2)$$

Анализ ZCR построен на предположении, что количество пересечений функции сигнала с нулевой осью для пауз с фоновым шумом больше по сравнению с вокализованной и невокализованной речью. Аналогично построен способ на основе анализа STE – кратковременная энергия вокализованной и невокализованной речи больше, чем энергия пауз с фоновым шумом. Однако данные предположения не совсем корректны. Не решен главный вопрос – насколько текущие значения ZCR и STE должны быть больше, чем пороговые для корректной сегментации речевых сигналов. Кроме того, известно, что пороговые значения могут варьироваться для каждого конкретного анализируемого речевого сигнала. В работе [10] была предпринята попытка выбрать и обосновать пороговые значения ZCR и STE, соответствующие вокализованной, невокализованной речи и паузам. В соответствии с выводами в работе [10] точность сегментации составила 65 % в сравнении с сегментацией, осуществленной вручную.

Способ сегментации речевых сигналов на основе анализа ODMD построен на статистических свойствах фонового шума [4]. В соответствии с физиологией воспроизведения речи человек перед произношением выдерживает вынужденную начальную паузу, длительностью не менее 200 мс, которая соответствует фоновому шуму. Предполагается, что фоновый шум, регистрируемый во время начальной паузы, имеет Гауссовский характер, а остальные информативные участки вокализованной и невокализованной речи имеют другое распределение. В этом случае функция плотности вероятности распределения фонового шума является критерием сегментации речевых сигналов.

В основе вычисления ODMD лежит функция плотности вероятности нормального распределения [4]:

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} , \quad (3)$$

где μ и σ – математическое ожидание и стандартное отклонение независимых случайных величин y.

Аналитическое выражение ODMD имеет следующий вид:

$$r = \frac{|y-\mu|}{\sigma}, \quad (4)$$

где выражение $|y-\mu|$ является естественной мерой расстояния от y к среднему значению μ.

В работе [11] представлен подробный сравнительный анализ результатов сегментации речевых сигналов, полученных с помощью способов на основе анализа ZCR, STE и ODMD. В соответствии с выводами в работе [11] способ на основе анализа ODMD эффективнее для отдельных словосочетаний, чем вышеупомянутые способы на 5,6 и 13,18 % соответственно. Для слитной речи повышение эффективности составляет на 8,88 и 9,59 % соответственно.

В соответствии с вышеупомянутым математическим описанием проведены исследования способов, основанных на анализе ZCR, STE и ODMD. В табл. 1 представлены усредненные данные ошибок 1-го (α) и 2-го рода (β), полученные по результатам сегментации с помощью вышеупомянутых способов. Основной задачей сегментации является точное обнаружение границ начала и окончания информативных участков вокализованной и невокализованной речи, поэтому ошибкой 1-го рода считалось ошибочное присваивание речевому фрагменту статуса «пауза». Ошибкой 2-го рода считалось ошибочное присваивание фрагменту паузы статуса «речь». Ошибки 1-го и 2-го рода определялись в сравнении с результатом сегментации, осуществленной вручную.

Таблица 1

Усредненные данные ошибок 1-го и 2-го рода, полученные по результатам сегментации способами на основе анализа ZCR, STE и ODMD

Способ сегментации речевых сигналов	Ошибки 1-го и 2-го рода, %	
	α	β
Способ на основе анализа ODMD	21,97	0,89
Способ на основе анализа ZCR	23,11	3,02
Способ на основе анализа STE	10,53	3,2
Способ на основе анализа ZCR и STE	7,32	5,33

В соответствии с полученными результатами в табл. 1 сделан вывод, что целесообразным является модернизация способа сегментации речевых сигналов на основе анализа ZCR и STE.

Энергетический оператор Тигера

TEO – это дифференциальный энергетический оператор 2-го порядка, позволяющий вычислять энергетические характеристики сигнала [12]. TEO обладает простотой, эффективностью и хорошей восприимчивостью к изменению амплитуды и частоты сигнала.

Для дискретных сигналов аналитическое выражение TEO имеет следующий вид:

$$\mathrm{TEO}(n) = x(n)^2 - x(n-1)x(n+1). \quad (5)$$

На сегодняшний день TEO получил широкое практическое применение в задачах обработки речевых сигналов [13, 14]. Задача сегментации речевых сигналов с помощью TEO частично решается в работах [15, 16]. Алгоритм на основе вейвлет-преобразования и TEO с высокой точностью сегментирует речь на вокализованные, невокализованные и переходные участки [15]. Применение TEO позволяет улучшить различимость вокализованной речи в присутствии сильного фонового шума [16].

Описание модернизированного способа сегментации речевых сигналов

На рис. 1 структурно представлен модернизированный способ сегментации речевых сигналов на основе энергетического анализа фрагментов речевого сигнала с помощью TEO и последующего анализа значений ZCR и STE. Суть работы способа заключается в линейном разделении речевого сигнала на фрагменты (блок 1), вычислении энергетической характеристики речевого сигнала с помощью TEO (блок 2), вычислении значений ZCR и STE фрагментов энергетической характеристики (блок 3, 4) и определении статуса «речь/пауза» фрагментов (блок 7) на основе вычисленных пороговых значений ZCR и STE (блок 5, 6). Блоки 8 и 9 не относятся к модернизированному способу и предназначены для постобработки ошибок сегментации, а также для сравнения результатов с сегментацией, осуществленной вручную. Рассмотрим подробнее некоторые этапы обработки модернизированного способа.

Блок 1. Фрагментирование представляет собой линейное разделение речевого сигнала на отрезки (фрагменты) равной длительности. Фрагментирование основано на кратковременном анализе, в рамках которого фрагменты обрабатываются так, как если бы они были короткими речевыми сигналами с отличающимися свойствами. Фрагментирование речевого сигнала осуществляется по следующим формулам:

$$S = \frac{x(n)}{L}, \quad (6)$$

где S – количество фрагментов в исследуемом речевом сигнале $x(n)$; L – количество дискретных отсчетов времени в одном фрагменте;

$$x_{s+1}[n_{first} : n_{final}] = x\left[(sL)+1 : (s+1)L\right], \quad (7)$$

где $s = 0, 1, 2, \ldots S$ – номер фрагмента, n_{first} – первый дискретный отсчет фрагмента; n_{final} – последний дискретный отсчет фрагмента.

Рис. 1. Структура модернизированного способа сегментации речевых сигналов на основе энергетического анализа фрагментов речевого сигнала с помощью TEO и последующего анализа ZCR и STE

Блоки 2–4. Вычисление энергетической характеристики фрагментов речевого сигнала с помощью TEO, а также значений ZCR и STE фрагментов энергетической характеристики осуществляется по формулам (1), (2) и (5) соответственно.

Особенностью модернизации способа является анализ значений ZCR и STE фрагментов энергетической характеристики речевого сигнала, вычисленной с помощью TEO. Известно, что TEO обеспечивает хорошую восприимчивость к изменению амплитуды и частоты сигнала, поэтому предполагается, что энергетическая характеристика содержит полную и достоверную информацию о значениях ZCR и STE.

На рис. 2 представлен пример, иллюстрирующий результат вычисления энергетической характеристики вокализованного участка и участка паузы с помощью TEO. На рис. 2,*а*,*б* представлены осциллограммы участков длительностью 30 мс (240 дискретных отсчета при частоте дискретизация 8000 Гц). На рис. 2,*в*,*г* представлены соответствующие участки функции энергетической характеристики. Как видно из рис. 2, функция энергетической характеристики в действительности обеспечивает полную информативность значений ZCR и STE, анализ которых позволит повысить эффективность сегментации речевых сигналов. Особенно полнота информации наблюдается при анализе значений ZCR для участка паузы. Осциллограмма участка паузы представляет собой фоновый шумовой сигнал с резким изменением значений амплитуды дискретных отсчетов времени (рис. 2,*б*). Значение ZCR для осциллограммы сигнала равно 8. Амплитудные значения дискретных отсчетов функции энергетической характеристики участка паузы также имеют резкие перепады (рис. 2,*г*), однако значение ZCR в этом случае равно 59.

Также необходимо отметить разницу значений ZCR, STE вокализованного участка и паузы, полученных для функции энергетической характеристики речевого сигнала. Если для значений ZCR разница между вокализованным участком (ZCR = 30) и паузой (ZCR = 59) примерно вдвое, то для значений STE разница составляет шесть порядков (STE = $9{,}82 \cdot 10^{-3}$ и STE = $7{,}83 \cdot 10^{-9}$). Таким образом, предположительно анализ значений ZCR и STE энергетической характеристики речевого сигнала, вычисленной с помощью TEO, должен обеспечить повышение эффективности сегментации речевых сигналов.

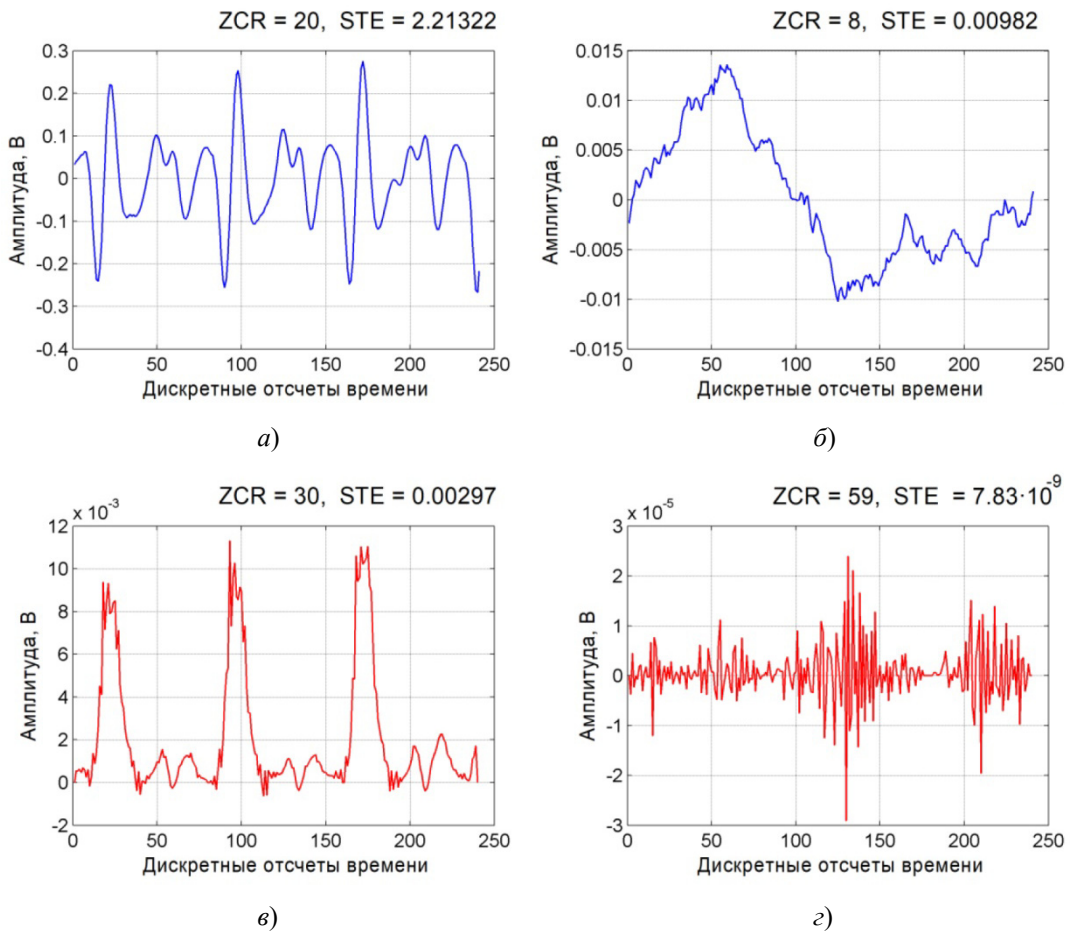


Рис. 2. Результат вычисления энергетической характеристики вокализованного участка и участка паузы с помощью ТЕО:
а, *б* – осциллограммы вокализованного участка и паузы соответственно;
в, *г* – функция энергетической характеристики вокализованного участка и паузы соответственно

Блоки 5, 6. Для корректной сегментации речевых сигналов в модернизированном способе представлено решение проблемы выбора пороговых значений ZCR и STE. По аналогии со способом [11] предлагается использовать начальную паузу в качестве исходных данных для формирования пороговых значений ZCR и STE. В соответствии с методикой в работе [11] вычисляются математическое ожидание μ_E, μ_{ZCR} и дисперсия σ_E, σ_{ZCR} значений ZCR и STE для фрагментов, соответствующих начальной паузе 200 мс (фоновому шуму):

$$\mu_{ZCR} = \frac{1}{S}\sum_{s=1}^{S}\mathrm{ZCR}_s; \tag{8}$$

$$\mu_E = \frac{1}{S}\sum_{s=1}^{S}E_s; \tag{9}$$

$$\sigma_{ZCR} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\mathrm{ZCR}_s - \mu_{ZCR}\right)^2}; \tag{10}$$

$$\sigma_E = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(E_s - \mu_E\right)^2}, \tag{11}$$

где ZCR_s, E_s – значения ZCR и STE исследуемого фрагмента соответственно; S – количество фрагментов, соответствующих фоновому шуму.

Блок 7. Определение статуса «речь/пауза» фрагментов заключается в проверке следующих условий:

$$\left|\mathrm{ZCR}_s - \mu_{ZCR}\right| \geq K\sigma_{ZCR}; \quad (12)$$

$$\left|E_s - \mu_E\right| \geq K\sigma_E, \quad (13)$$

где выражения $\left|\mathrm{ZCR}_s - \mu_{ZCR}\right|$, $\left|E_s - \mu_E\right|$ являются естественной мерой ODMD от текущих значений ZCR и STE к средним значениям, соответствующим фоновому шуму; K – коэффициент порога, предназначенный для изменения пороговых значений ZCR и STE (K всегда больше 1).

Если разница между текущим и средним значениями ZCR больше или равна произведению коэффициента порога и дисперсии $K\sigma_{ZCR}$, то фрагмент соответствует паузе. И наоборот, если условие не выполняется, то фрагмент соответствует речи. Аналогично если разница между текущим и средним значениями STE больше или равна произведению $K\sigma_E$, то фрагмент соответствует речи. И наоборот, если условие не выполняется, то фрагмент соответствует паузе.

Исследование модернизированного способа сегментации речевых сигналов

В рамках исследования модернизированного способа оценивалось:

– влияние коэффициента порога на эффективность сегментации речевых сигналов в сравнении с классическим способом на основе анализа ZCR и STE;

– помехоустойчивость в сравнении со способами на основе анализа ZCR, STE и ODMD.

Для проведения исследования сформирована база чистых речевых сигналов. Регистрация сигналов осуществлялась посредством специально разработанной методики, в рамках которой дикторы зачитывали следующий текстовый материал:

– статья из рекламно-информационной газеты, включающая публицистический текст на русском языке (не менее 200 слов);

– краткое детское литературное произведение, включающее фонетически сбалансированный текст на русском языке (не менее 200 слов);

– счет чисел от 0 до 99 на русском языке (100 слов).

Произношение диктором – размеренное. Параметры и пространственные характеристики микрофона не изменялись для всех регистрируемых дикторов. Количество дикторов – 20 человек из числа мужчин и женщин. Регистрация осуществлялась в обычном лабораторном помещении без специальных средств шумоподавления.

Для оценки помехоустойчивости сформирована база зашумленных речевых сигналов с различными отношениями сигнал/шум (ОСШ): от 20 до –5 дБ с шагом 5 дБ. Зашумление зарегистрированных чистых речевых сигналов осуществлялось программно посредством наложения сгенерированного белого шума.

В табл. 2 представлены усредненные значения ошибок 1-го и 2-го рода для классического способа сегментации речевых сигналов на основе анализа ZCR, STE и модернизированного способа. В табл. 3 представлены усредненные значения ошибок 1-го и 2-го рода, полученные по результатам сегментации зашумленных речевых сигналов классическим способом на основе анализа ZCR, STE и модернизированным способом. Светло-серым фоном отмечены наилучшие достигнутые результаты сегментации речевых сигналов. Темно-серым фоном отмечены значения ошибок 1-го и 2-го рода, которые необходимо детализировать.

Таблица 2

Усредненные значения ошибок 1-го и 2-го рода для классического способа сегментации речевых сигналов на основе анализа ZCR, STE и модернизированного способа на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE

Способ сегментации речевых сигналов		Значение коэффициента порога														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Классический способ	α	1,37	1,60	5,03	6,64	7,09	8,24	11,67	12,59	14,42	16,02	17,39	18,08	18,76	18,99	19,45
	β	36,06	15,10	4,44	1,78	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
Модернизированный способ	α	0,92	0,92	0,92	0,92	0,92	1,37	1,37	1,83	1,83	2,06	2,29	2,52	2,52	2,52	2,52
	β	47,25	22,03	13,68	10,48	7,46	5,51	3,91	2,66	2,66	1,95	1,95	1,95	1,95	1,95	1,95

Таблица 3

Усредненные значения ошибок 1-го и 2-го рода, полученные по результатам сегментации зашумленных речевых сигналов классическим способом на основе анализа ZCR, STE и модернизированным способом на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE

Отношение сигнал/шум, дБ	Ошибка, %	Значение коэффициента порога														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Классический способ																
20	α	4,81	8,01	10,76	11,90	12,59	13,27	14,19	16,02	17,62	18,31	18,99	19,22	19,91	20,37	20,60
	β	29,31	4,09	1,95	1,24	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
15	α	2,06	8,47	10,53	11,90	13,27	14,19	15,10	16,48	18,31	18,76	18,99	20,14	20,37	21,28	21,51
	β	50,27	4,80	1,60	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
10	α	2,52	8,01	10,53	12,82	14,65	16,48	17,16	17,62	18,99	20,14	20,60	21,05	21,51	22,20	22,43
	β	56,84	5,15	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
5	α	7,09	11,90	14,42	17,85	19,22	19,91	21,28	22,20	22,43	22,88	23,57	24,03	24,26	24,26	24,71
	β	42,81	1,78	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
0	α	2,52	15,10	21,28	24,26	26,09	27,23	29,98	31,81	33,18	35,01	36,16	37,07	37,76	39,36	40,50
	β	83,30	28,60	5,51	1,07	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
–5	α	7,78	31,35	37,99	44,85	49,89	53,78	57,90	60,18	62,24	67,96	70,48	74,14	75,29	78,26	84,21
	β	60,75	4,80	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
Модернизированный способ																
20	α	0,69	2,52	4,58	7,09	7,78	8,92	9,61	10,07	11,44	11,90	12,36	12,59	12,82	13,73	13,96
	β	33,93	8,35	1,60	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
15	α	2,97	8,24	10,53	12,13	12,36	13,50	15,10	16,02	17,16	17,39	18,54	18,99	19,68	20,14	20,82
	β	40,32	3,55	1,24	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
10	α	2,75	8,01	12,36	13,50	14,19	17,85	19,22	20,37	20,37	21,74	22,20	22,20	22,43	22,65	22,88
	β	71,23	9,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
5	α	6,41	11,21	13,50	15,33	17,39	18,08	19,91	20,60	20,82	21,51	21,74	21,97	22,20	22,20	22,88
	β	52,58	2,31	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
0	α	5,26	18,31	22,88	24,49	27,00	31,12	32,04	33,18	33,41	35,24	36,38	39,13	40,05	40,96	41,42
	β	64,65	5,51	1,24	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89
–5	α	6,41	22,65	33,18	38,22	43,25	48,28	50,11	53,78	56,06	56,98	59,50	60,64	61,56	64,07	66,13
	β	71,76	15,45	1,42	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89

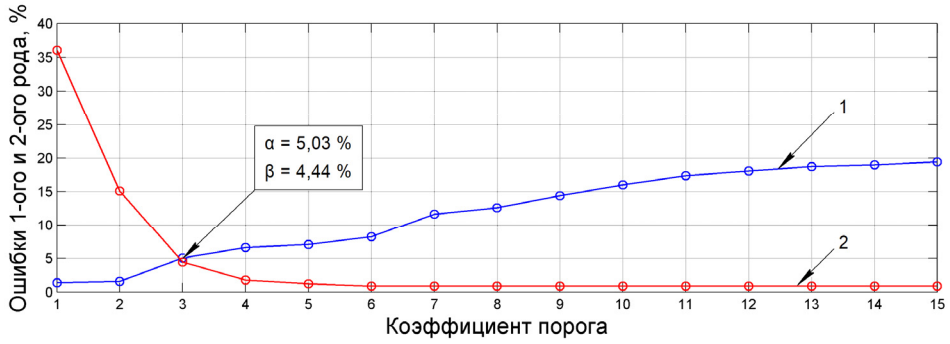
Анализ результатов исследований

В соответствии с данными из табл. 2 на рис. 3 представлены кривые зависимости ошибок 1-го и 2-го рода от коэффициента порога.

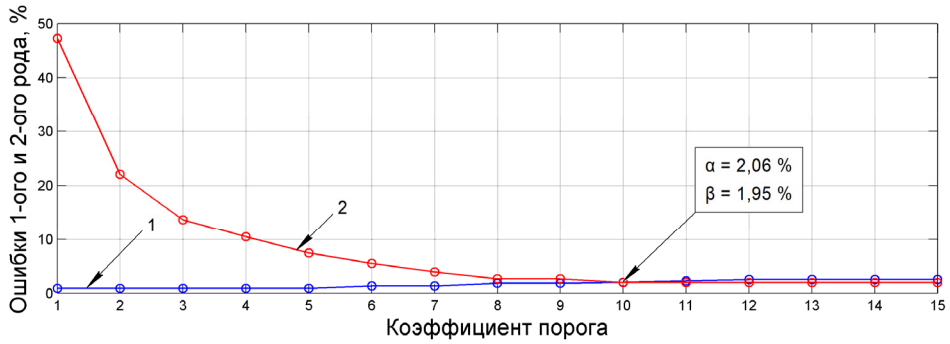
Анализ полученных результатов в табл. 2 и кривых зависимостей на рис. 3 выявил, что наиболее оптимальные значения ошибок 1-го и 2-го рода достигаются модернизированным способом – 2,06 и 1,95 % соответственно при значении коэффициента порога равном 10.

Наиболее оптимальные значения ошибок 1-го и 2-го рода для классического способа на основе анализа ZCR и STE – 5,03 и 4,44 % соответственно достигаются при значении коэффициента порога равном 3.

На рис. 4 представлен пример, иллюстрирующий результаты сегментации речевого сигнала длительностью 10 с, представляющего собой сочетание следующих слов на русском языке: шанс, шар, баян, Лара, нормально. Слова подобраны таким образом, чтобы в них содержались разные по способу образования звуки: гласные, сонорные, шумные смычные (взрывные, фрикативные) и шумные щелевые.

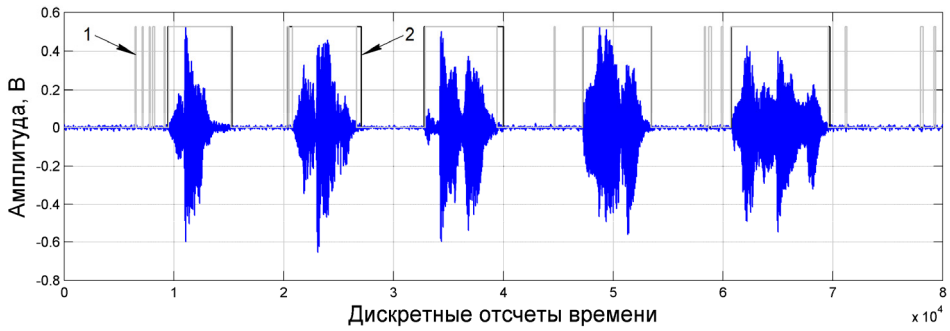


а)

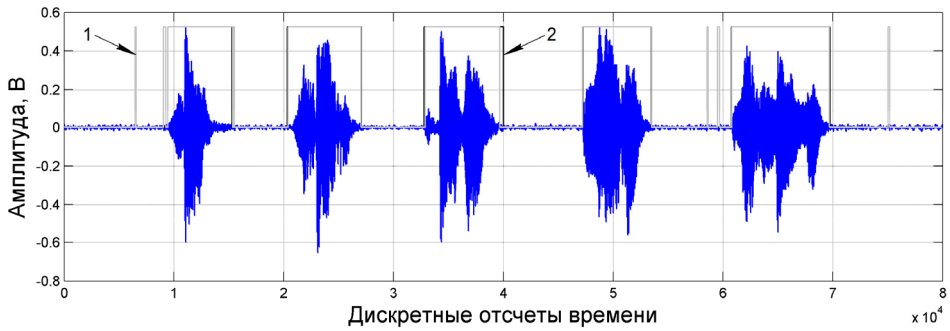


б)

Рис. 3. Зависимость ошибок 1-го и 2-го рода от коэффициента порога:
а – классический способ сегментации речевых сигналов на основе анализа ZCR и STE;
б – модернизированный способ сегментации речевых сигналов на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE;
1 – значения ошибок 1-го рода; *2* – значения ошибок 2-го рода



а)



б)

Рис. 4. Пример, иллюстрирующий результаты сегментации речевого сигнала:
а – классический способ сегментации речевых сигналов на основе анализа ZCR и STE;
б – модернизированный способ сегментации речевых сигналов на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE; *1* – достигнутые результаты сегментации; *2* – результат сегментации, осуществленной вручную

В соответствии с данными в табл. 3, полученными по результатам сегментации зашумленных речевых сигналов, оптимальные значения ошибок 1-го и 2-го рода достигаются при значениях коэффициента порога от 1 до 3. В табл. 4 представлены детализированные усредненные значения ошибок 1-го и 2-го рода для коэффициента порога от 1,5 до 2,5 с шагом 0,1. Как видно из результатов в табл. 4, оптимальные значения ошибок 1-го и 2-го рода достигаются при разных значениях коэффициента порога. Для классического и модернизированного способов оптимальные значения коэффициента порога находятся в пределах от 1,9 до 2,2.

Таблица 4

Детализированные усредненные значения ошибок 1-го и 2-го рода, полученные по результатам сегментации зашумленных речевых сигналов классическим способом на основе анализа ZCR, STE и модернизированным способом на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE

Отношение сигнал/шум, дБ	Ошибка, %	Значение коэффициента порога										
		1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3	2,4	2,5
Классический способ												
20	α	7,09	7,78	7,78	7,78	8,01	8,01	8,01	9,15	9,38	9,61	10,07
	β	9,41	7,46	6,04	4,80	4,09	4,09	3,37	3,37	2,84	1,95	1,95
15	α	5,95	7,09	8,01	8,01	8,47	8,47	8,47	8,47	8,70	9,15	9,61
	β	14,92	11,19	8,17	7,46	5,15	4,80	3,37	3,37	3,37	3,37	2,84
10	α	5,95	6,18	6,41	7,32	7,55	8,01	8,70	9,38	9,38	9,84	9,84
	β	21,85	17,23	13,32	9,77	6,75	5,15	4,62	4,09	3,02	2,66	2,31
5	α	9,84	10,30	10,53	10,98	11,90	11,90	12,36	12,36	12,36	12,36	12,36
	β	10,12	5,33	4,62	2,66	2,49	1,78	1,78	1,42	1,24	1,24	1,24
0	α	17,39	17,85	19,45	19,68	20,60	21,28	21,28	21,74	22,20	22,43	22,43
	β	15,10	14,39	9,95	9,95	7,10	6,04	5,51	3,73	3,55	3,55	2,66
–5	α	23,11	24,49	26,32	29,06	30,44	31,35	32,72	33,87	35,01	35,47	35,93
	β	24,69	15,99	13,50	8,17	7,46	4,80	4,26	2,31	1,60	1,42	1,42
Модернизированный способ												
20	α	1,83	1,83	2,52	2,52	2,52	2,52	2,75	2,97	3,66	1,83	1,83
	β	13,85	13,50	11,37	9,95	9,59	8,35	6,22	5,68	4,97	13,85	13,50
15	α	5,49	6,64	7,09	8,01	8,01	8,24	8,24	8,47	8,92	5,49	6,64
	β	11,01	8,53	6,93	4,97	4,09	3,55	2,84	2,13	2,13	11,01	8,53
10	α	6,64	6,64	7,09	7,55	7,78	8,01	9,15	9,38	9,61	6,64	6,64
	β	30,20	28,77	22,20	14,57	12,08	9,24	6,75	4,62	4,09	30,20	28,77
5	α	8,70	8,92	9,84	10,07	10,53	11,21	11,21	11,67	12,36	8,70	8,92
	β	13,14	11,72	7,46	4,26	4,26	2,31	2,31	1,60	1,42	13,14	11,72
0	α	12,36	14,42	15,56	17,16	18,31	18,31	20,60	20,82	21,28	12,36	14,42
	β	19,54	15,45	13,50	8,70	6,93	5,51	4,09	3,37	2,49	19,54	15,45
–5	α	12,13	14,65	16,48	17,16	21,05	22,65	24,71	26,09	26,55	12,13	14,65
	β	37,48	32,68	29,31	25,22	15,99	15,45	13,50	9,95	7,99	37,48	32,68

В табл. 5 представлены усредненные данные значений ошибок 1-го и 2-го рода, полученные по результатам сегментации чистого и зашумленных речевых сигналов. В соответствии с данными в табл. 5 практически для всех значений ОСШ модернизированный способ демонстрирует лучшую помехоустойчивость в сравнении со способами на основе анализа ZCR, STE и ODMD. Необходимо отметить, что в зависимости от требований к точности сегментации речевых сигналов, модернизированный способ обеспечивает вариативность значений ошибок 1-го и 2-го рода за счет изменения коэффициента порога (см. табл. 4).

Таблица 5

Усредненные данные значений ошибок 1-го и 2-го рода, полученные по результатам сегментации чистого и зашумленных речевых сигналов способами на основе анализа ZCR, STE, ODMD, а также модернизированным способом на основе энергетического анализа с помощью TEO и последующего анализа значений ZCR и STE

Способ сегментации речевых сигналов	Чистый сигнал		Зашумленный речевой сигнал											
			Отношение сигнал/шум, дБ											
			20		15		10		5		0		–5	
	Ошибка, %													
	α	β	α	β	α	β	α	β	α	β	α	β	α	β
Способ на основе анализа ODMD	21,97	0,89	22,88	0,89	35,01	0,89	41,19	0,89	59,50	0,89	99,77	0,89	99,77	0,89
Способ на основе анализа ZCR	23,11	3,02	27,23	2,13	71,85	1,42	65,90	1,95	75,29	1,07	97,48	0,89	90,85	1,78
Способ на основе анализа STE	10,53	3,20	9,61	1,95	7,09	4,97	9,38	3,55	12,59	1,42	16,25	22,20	37,07	1,42
Способ на основе анализа ZCR и STE	5,03	4,44	8,01	3,37	8,47	3,37	8,01	5,15	11,90	1,78	21,74	3,73	31,35	4,80
Модернизированный способ	2,06	1,95	2,75	6,22	8,24	2,84	8,01	9,24	11,67	1,60	18,31	5,51	21,05	15,99

Заключение

Подводя итоги анализа результатов исследований, можно сделать следующие выводы:

1. При сравнении оптимальных значений ошибок 1-го и 2-го рода модернизированный способ обеспечивает повышение эффективности сегментации на 2,97 и 2,49 % соответственно. Это обеспечивается за счет хорошей восприимчивости TEO к резкому изменению амплитуды и частоты речевых сигналов.

2. За счет применения TEO модернизированный способ обеспечивает наилучшие результаты сегментации зашумленной речи. В сравнении с наиболее помехоустойчивым классическим способом сегментации на основе анализа ZCR и STE отмечаются следующие изменения значений ошибок 1-го и 2-го рода:

- ОСШ = 20 дБ улучшение на 5,26 % и ухудшение на 2,85 %;
- ОСШ = 15 дБ улучшение на 0,23 % и 0,53 %;
- ОСШ = 10 дБ без изменений и ухудшение на 4,09 %;
- ОСШ = 5 дБ улучшение на 0,23 % и 0,18 %;
- ОСШ = 0 дБ улучшение на 3,43 % и ухудшение на 1,78 %;
- ОСШ = –5 дБ улучшение на 10,3 % и ухудшение на 11,19 %.

3. В зависимости от приоритета решаемой задачи сегментации речевых сигналов, у исследователей имеется возможность выбирать между необходимыми значениями коэффициента порога, обеспечивающими необходимые значения ошибок 1-го и 2-го рода.

В перспективе планируется провести дополнительное исследование быстродействия модернизированного способа сегментации речевых сигналов.

Список литературы

1. 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.
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. 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.
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. 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.
6. 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.
7. Алимурадов А. К., Квитка Ю. С., Чураков П. П., Тычков А. Ю. Повышение точности измерения частоты основного тона на основе оптимизации процесса декомпозиции речевых сигналов на эмпирические моды // Измерение. Мониторинг. Управление. Контроль. 2018. № 4. С. 53–65.
8. Алимурадов А. К. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов // Труды МФТИ. 2015. Т. 7, № 3. С. 56–68.
9. Алимурадов А. К., Тычков А. Ю. Применение метода декомпозиции на эмпирические моды для исследования вокализованной речи в задаче обнаружения стрессовых эмоций человека // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2020. № 3. С. 7–29.
10. Greenwood M.A., Kinghorn A. SUVing: automatic silence/unvoiced/voiced classification of speech // Undergraduate Coursework, Department of Computer Science, The University of Sheffield, UK, 1999. 4 p.
11. Saha G., Chakroborty S., Senapat S. A new silence removal and endpoint detection algorithm for speech and speaker recognition applications // Eleventh National Conference on Communications (NCC-2005) (Jan. 28–30, 2005). Kharagpur, India, 2005. P. 51–61.
12. 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.
13. Abu-Shikhah N., Deriche M. A novel pitch estimation technique using the Teager energy // International Symposium on Signal Processing and Its Applications (ISSPA) (IEEE Cat. No.99EX359) (Aug. 22–25, 1999). Brisbane, Queensland, Australia, 1999. Vol. 1. P. 135–138.
14. Kvedalen E. Signal Processing Using the Teager Energy Operator and Other Nonlinear Operators : PhD dissertation, Department of Informatics. Oslo : University of Oslo, 2003. 121 p.
15. Жуйков В. Я., Харченко А. Н. Алгоритм классификации сегментов речевого сигнала // Электроника и связь. Тем. вып. «Электроника и нанотехнологии». 2009. Ч. 1. С. 130–137.
16. 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. 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.
2. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Developmen*. New Jersey: Prentice Hall, 2001:980.
3. 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.
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. 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.
6. 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.
7. Alimuradov A.K., Kvitka Yu.S., Churakov P.P., Tychkov A.Yu. Improving the accuracy of measuring the pitch frequency based on optimizing the process of decomposition of speech signals into empirical modes. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control*. 2018; (4):53–65. (In Russ.)
8. Alimuradov A.K. Investigation of frequency-selective properties of decomposition methods into empirical modes for estimating the frequency of the fundamental tone of speech signals. *Trudy MFTI = Works of MIPT*. 2015;7(3):56–68. (In Russ.)
9. Alimuradov A.K., Tychkov A.Yu. Application of the method of decomposition into empirical modes for the study of vocalized speech in the task of detecting stressful human emotions. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye*

tekhnologii, sistemy upravleniya = Bulletin of Perm National Research Polytechnic University. Electrical engineering, information technology, control systems. 2020;(3):7–29. (In Russ.)
10. Greenwood M.A., Kinghorn A. *SUVing: automatic silence/unvoiced/voiced classification of speech*. Undergraduate Coursework, Department of Computer Science, The University of Sheffield, UK, 1999:4.
11. Saha G., Chakroborty S., Senapat S. A new silence removal and endpoint detection algorithm for speech and speaker recognition applications. *Eleventh National Conference on Communications (NCC-2005) (Jan. 28–30, 2005)*. Kharagpur, India, 2005:51–61.
12. 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.
13. Abu-Shikhah N., Deriche M. A novel pitch estimation technique using the Teager energy. *International Symposium on Signal Processing and Its Applications (ISSPA) (IEEE Cat. No.99EX359) (Aug. 22–25, 1999)*. Brisbane, Queensland, Australia, 1999;1:135–138.
14. Kvedalen E. *Signal Processing Using the Teager Energy Operator and Other Nonlinear Operators: PhD dissertation, Department of Informatics*. Oslo: University of Oslo, 2003:121.
15. Zhuykov V.Ya., Kharchenko A.N. Algorithm for classification of speech signal segments. *Elektronika i svyaz'. Tem. vyp. «Elektronika i nanotekhnologii» = Electronics and Communications. Topic. issue. "Electronics and nanotechnology".* 2009;1:130–137. (In Russ.)
16. 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

Alan K. Alimuradov
Candidate of technical sciences,
director of student research
and production business incubator,
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 12.05.2021
Поступила после рецензирования/Revised 19.05.2021
Принята к публикации/Accepted 20.05.2021