

Раздел 2 МОДЕЛИ, СИСТЕМЫ, СЕТИ В ТЕХНИКЕ

Section 2 MODELS, SYSTEMS, NETWORKS IN THE TECHNIQUE

УДК 004.934
doi:10.21685/2227-8486-2021-4-5

СПОСОБ СЕГМЕНТАЦИИ РЕЧЬ/ПАУЗА НА ОСНОВЕ ЭНЕРГЕТИЧЕСКОГО ОПЕРАТОРА ТИГЕРА

А. К. Алимуратов

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

Аннотация. *Актуальность и цели.* Сегментация речи на вокализованные, невокализованные участки и паузы является основной задачей практически для всех речевых приложений. Особенно это важно в системах оценки психоэмоционального состояния человека по речи, так как длительности вокализованных, невокализованных участков и пауз являются информативными параметрами, релевантными естественно выраженным эмоциям человека. *Материалы и методы.* Использовался дифференциальный энергетический оператор Тигера 2-го порядка, обладающий хорошей восприимчивостью к изменению амплитуды и частоты сигнала. Способ реализован посредством программы © Matlab (MathWorks). *Результаты.* Разработан способ сегментации речь/пауза, суть которого заключается в линейном разделении речевого сигнала на фрагменты, вычислении энергетической характеристики с помощью энергетического оператора Тигера, вычислении значений кратковременной энергии и определении статуса «речь/пауза» фрагментов на основе рассчитанных пороговых значений кратковременной энергии. Проведено исследование разработанного способа, в рамках которого оценивалась эффективность сегментации речь/пауза в сравнении с классическим способом на основе анализа кратковременной энергии. *Выводы.* В соответствии с полученными результатами исследований отмечается повышение эффективности сегментации речь/пауза на 5,26 % и 5,51 % для ошибок 1-го и 2-го рода соответственно. Предложенный способ сегментации речь/пауза может успешно тестироваться в системах оценки психоэмоционального состояния человека, так как обладает хорошей восприимчивостью к резким изменениям амплитуды и частоты сигнала в условиях нестабильной моторики речевого аппарата.

Ключевые слова: обработка речевых сигналов, сегментация речи, вокализованная и невокализованная речь, кратковременная энергия, энергетический оператор Тигера

© Алимуратов А. К., 2021. Контент доступен по лицензии Creative Commons Attribution 4.0 License / This work is licensed under a Creative Commons Attribution 4.0 License.

Для цитирования: Алимуратов 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. doi:10.21685/2227-8486-2021-4-5

SPEECH/PAUSE SEGMENTATION METHOD BASED ON TEAGER ENERGY OPERATOR

A.K. Alimuradov

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

Abstract. *Background.* Speech segmentation into voiced, unvoiced sections and pauses is the key task for the majority of speech applications. This is especially important in systems for assessing human psycho-emotional state by speech, since duration of voiced, unvoiced sections and pauses are informative parameters being relevant to naturally expressed human emotions. *Materials and methods.* The second-order differential Teager energy operator was used, which has a good amplitude that is highly susceptible to changes in signal amplitude and frequency. The method is implemented by means of the program © Matlab (MathWorks). *Results.* There has been developed a method for speech/pause segmentation to linearly divide a speech signal into fragments, to calculate the energy characteristic using the Teager energy operator, to calculate the values of short-term energy, and determine the «speech/pause» status of fragments based on the calculated threshold values of the short-term energy. There has been carried out a research on the developed method to assess the effectiveness of speech/pause segmentation over the classical method based on the analysis of short-term energy, has been carried out. *Conclusions.* In accordance with the obtained research results, there is an increase in the efficiency of speech/pause segmentation by 5.26 % and 5.51 % for the 1st and 2nd kind errors, respectively. The proposed speech/pause segmentation method can be effectively tested in systems for assessing human psycho-emotional state due to its good susceptibility to sudden changes in signal amplitude and frequency with unstable vocal motor skills.

Keywords: speech signal processing, speech segmentation, voiced and unvoiced speech, Short-Time Energy, Teager Energy Operator

For citation: Alimuradov A.K. Speech/pause segmentation method based on teager energy operator. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, systems, networks in economics, technology, nature and society.* 2021;(4):52–63. (In Russ.). doi:10.21685/2227-8486-2021-4-5

Введение

В соответствии с физиологией речевого аппарата слитная речь человека состоит из следующих информативных участков: дыхание и паузы, вокализованная и невокализованная речь. Задача сегментации речь/пауза представляет собой классификацию информативных участков, характеризующихся определенной длительностью. Длительности участков дыхания и пауз, вокализованной и невокализованной речи являются важными информативными параметрами речи, релевантными естественно выраженным эмоциям человека [1].

В настоящее время задача сегментации речь/пауза решается разными подходами. Наиболее распространенными являются способы, реализованные с помощью анализа значений:

- кратковременной энергии (Short Time Energy, STE) [2];

– количества пересечения сигнала через нулевую ось (Zero-Crossing Rate, ZCR) [3].

Однако данные способы становятся малоэффективны при повышении уровня окружающего шума. Относительной помехоустойчивостью обладают способы сегментации, реализованные с помощью анализа:

- спектральных характеристик [5];
- мел-частотных кепстральных коэффициентов и их первого и второго приращений (Mel-Frequency Cepstral Coefficients, MFCC) [8];
- мощности в ограниченном диапазоне частот [6];
- отклонения автокорреляционной функции (Autocorrelation Function, ACR) [4];
- одномерного расстояния Махаланобиса (One Dimensional Mahalanobis Distance, ODMD) [7].

Также для повышения устойчивости к шуму в некоторых способах сегментации анализируются несколько упомянутых выше акустических характеристик в сочетании с алгоритмами моделирования, такими как искусственная нейронная сеть (Artificial Neural-Network, ANN) [9] и дерево классификации и регрессии (Classification and Regression Tree, CART) [10].

В данной статье представлен способ сегментации речь/пауза, основанный на анализе фрагментов речи посредством энергетического оператора Тигера (Teager energy operator, TEO) [11] с последующим анализом значений STE. В последнее время TEO широко применяется в задачах обработки речевых сигналов [12, 13].

Статья является результатом научной работы, посвященной разработке новых высокоэффективных способов сегментации речи на информативные участки [14, 15].

Сегментация на основе анализа кратковременной энергии

Значение STE определяется по следующей формуле:

$$E_s = \sum_{n=-\infty}^{\infty} [x(n)w(s-n)]^2, \quad (1)$$

где $x(n)$ – исследуемый речевой сигнал; n – дискретный отсчет времени; s – номер фрагмента; $w(k)$ – функция анализируемого окна.

Для прямоугольного анализируемого окна формула (1) принимает следующий вид:

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

где N – количество дискретных отсчетов во фрагменте.

Способ сегментации речь/пауза на основе анализа STE построен на предположении, что значение STE участков речи больше, чем энергия пауз с фоновым шумом. В работе М. А. Гринвуда и А. Кингхорна [16] представлены и обоснованы пороговые значения STE, соответствующие вокализованной, невокализованной речи и паузам. Точность сегментации составила 65 % в сравнении с сегментацией, осуществленной вручную.

Энергетический оператор Тигера

ТЕО – дифференциальный энергетический оператор 2-го порядка, позволяющий оценивать энергетические характеристики сигнала. Решение задачи сегментации речь/пауза с помощью ТЕО представлено в работах В. Я. Жуйкова и А. Н. Харченко и др. [17, 18]. Для дискретных сигналов функция ТЕО имеет следующий вид:

$$\text{ТЕО}(n) = x(n)^2 - x(n-1)x(n+1). \quad (3)$$

При эмоциональном возбуждении вследствие неполного смыкания голосовых связок работа речевого аппарата человека характеризуется нерегулярностью. Предполагается, что хорошая восприимчивость ТЕО к резкому изменению амплитуды и частоты сигнала обеспечит высокую точность сегментации речь/пауза даже в условиях нестабильной моторики речевого аппарата.

Описание способа сегментации речь/пауза

Структурно предложенный способ сегментации речь/пауза представлен на рис. 1.

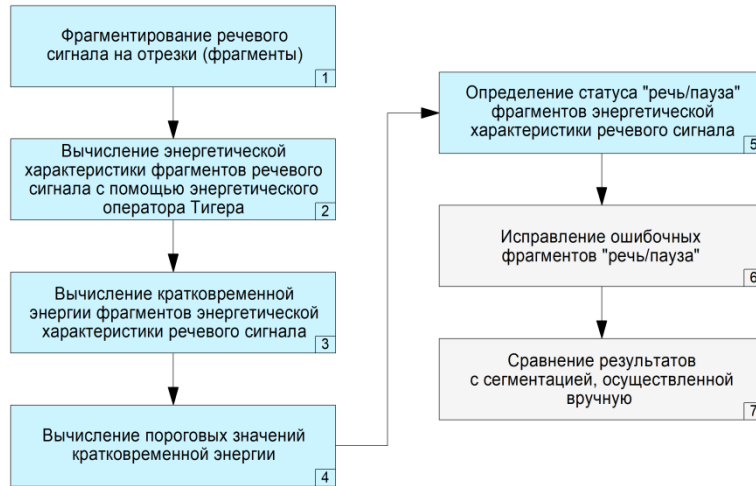


Рис. 1. Структура способа сегментации речь/пауза на основе энергетического анализа фрагментов речевого сигнала с помощью ТЕО и STE

Блок 1. Фрагментирование речевого сигнала на отрезки (фрагменты) равной длительности осуществляется по формулам:

$$S = \frac{x(n)}{L}, \quad (4)$$

$$x_{s+1}[n_{first} : n_{final}] = x[(sL) + 1 : (s+1)L], \quad (5)$$

где S – количество фрагментов; $x(n)$ – исследуемый речевой сигнал; L – количество дискретных отсчетов времени в одном фрагменте; $s = 0, 1, 2, \dots S$ – номер фрагмента, n_{first} – первый дискретный отсчет фрагмента; n_{final} – последний дискретный отсчет фрагмента.

Блоки 2, 3. Вычисление энергетической характеристики речевого сигнала с помощью ТЕО, а также значений STE фрагментов энергетической характеристики осуществляется по формулам (2) и (3) соответственно.

Блок 4. В соответствии с физиологией воспроизведения речи человек перед произношением выдерживает вынужденную начальную паузу длительностью не менее 200 мс, которая соответствует фоновому шуму. Начальная пауза используется в качестве исходных данных для формирования пороговых значений STE. По аналогии с методикой, представленной в работе С. Чакроборти и др. [19], вычисляются математическое ожидание μ_E и дисперсия σ_E значений STE для фрагментов, соответствующих начальной паузе 200 мс (фоновому шуму):

$$\mu_E = \frac{1}{S} \sum_{s=1}^S E_s, \quad (6)$$

$$\sigma_E = \sqrt{\frac{1}{S} \sum_{s=1}^S (E_s - \mu_E)^2}, \quad (7)$$

где E_s – значение STE исследуемого фрагмента; S – количество фрагментов, соответствующих фоновому шуму.

Блок 5. Определение статуса «речь/пауза» фрагментов заключается в проверке следующего условия:

$$|E_s - \mu_E| \geq K \sigma_E, \quad (8)$$

где выражение $|E_s - \mu_E|$ является естественной мерой одномерного расстояния Махаланобиса [19] от текущего значения кратковременной энергии фрагмента E_s к среднему значению кратковременной энергии фрагментов μ_E , соответствующих фоновому шуму; K – коэффициент порога, численная вероятность которого подчиняется следующим выражениям: $|E_s - \mu_E| \leq \sigma = 0,68$, $|E_s - \mu_E| \leq 2\sigma = 0,95$ и $|E_s - \mu_E| \leq 3\sigma = 0,997$ (K всегда больше 1).

Если разница между текущим и средним значениями STE больше или равна $K\sigma_E$, то фрагмент соответствует речи. И наоборот, если условие не выполняется, то фрагмент соответствует паузе.

Блок 6. В основе исправления ошибок сегментации (поиска некорректно определенных фрагментов речь/пауза) заложен следующий физиологический аспект: невозможность человеком кратковременно изменять воспроизводимую речь на паузу и наоборот (в течение 20 мс). На рис. 2 представлен пример, иллюстрирующий ошибки сегментации речь/пауза.

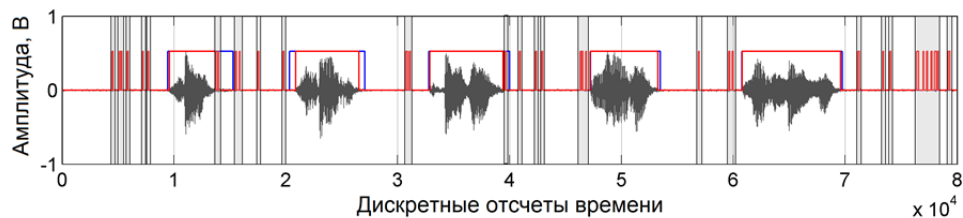


Рис. 2. Ошибки сегментации речь/пауза (линией красного цвета обозначен результат сегментации, линией синего цвета – результат сегментации, осуществленной вручную, фоном серого цвета обозначены некорректно определенные фрагменты)

Блок 7. Вручную сегментация речевых сигналов на информативные участки осуществлялась посредством многоплатформенного аудиоредактора звуковых файлов «Audacity» (производитель «Audacity Team», audacityteam.org).

Исследование способа сегментации речь/пауза

Для оценки предложенного способа сегментации речь/пауза сформирована база речевых сигналов. Количество дикторов – 20 человек из числа мужчин и женщин. Регистрация осуществлялась в обычном лабораторном помещении без специальных средств звукоизоляции и шумоподавления. Психологическое состояние дикторов – условно нейтральное. Параметры и пространственные характеристики микрофона не изменялись для всех регистрируемых дикторов.

Для оценки эффективности предложенного способа сегментации речь/пауза вычислялись значения ошибок 1-го (α) и 2-го (β) рода. В рамках исследования предложенного способа оценивалось влияние коэффициента порога на эффективность сегментации речь/пауза в сравнении со способом на основе анализа значений STE.

В табл. 1 представлены усредненные значения ошибок 1-го и 2-го рода для способа сегментации речь/пауза на основе энергетического анализа с помощью STE и предложенного способа. Фоном зеленого цвета отмечены наилучшие достигнутые результаты сегментации речь/пауза. Фоном синего цвета отмечены значения ошибок 1-го и 2-го рода, которые необходимо детализировать.

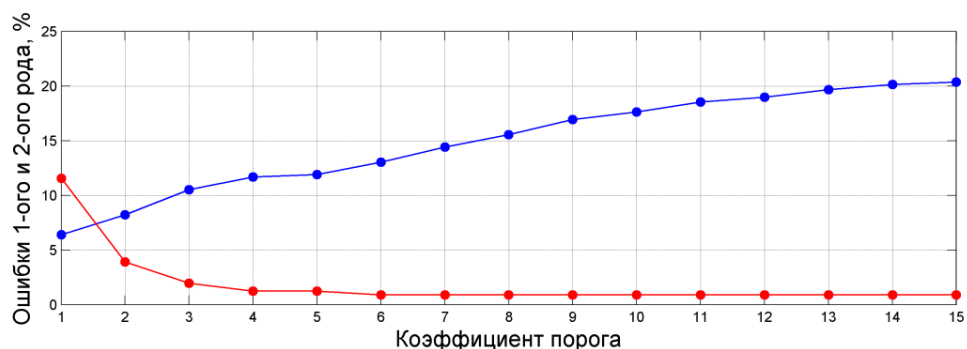
Результаты и обсуждение

В соответствии с данными из табл. 1 на рис. 3 представлены кривые зависимости ошибок 1-го и 2-го рода от коэффициента порога для способа сегментации речь/пауза на основе энергетического анализа с помощью STE и предложенного способа.

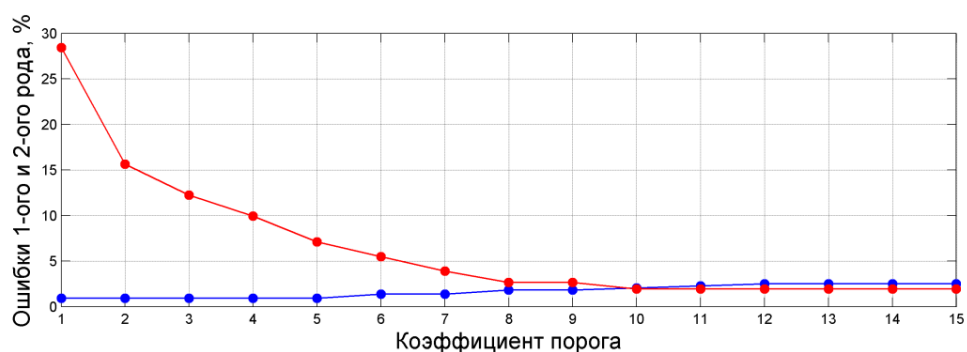
Таблица 1

Усредненные значения ошибок 1-го и 2-го рода для способа сегментации речь/пауза на основе энергетического анализа с помощью STE и способа сегментации речь/пауза на основе энергетического анализа с помощью ТЕО и STE

Способ сегментации	Ошибка, %	Значение коэффициента порога														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Способ сегментации речь/пауза на основе энергетического анализа с помощью STE	α	6,41	8,24	10,53	11,67	11,90	13,04	14,42	15,56	16,93	17,62	18,54	18,99	19,68	20,14	20,37
	β	11,55	3,91	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
Способ сегментации речь/пауза на основе энергетического анализа с помощью ТЕО и STE	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	28,42	15,63	12,26	9,95	7,10	5,51	3,91	2,66	2,66	1,95	1,95	1,95	1,95	1,95	1,95



а)



б)

Рис. 3. Зависимость ошибок 1-го и 2-го рода от коэффициента порога:
 а – способ сегментации речь/пауза на основе энергетического анализа с помощью STE; б – способ сегментации речь/пауза на основе энергетического анализа с помощью ТЕО и STE (линией синего цвета обозначены значения ошибок 1-го рода, линией красного цвета – значения ошибок 2-го рода)

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

Для способа сегментации речь/пауза на основе энергетического анализа с помощью STE оптимальные значения ошибок 1-го и 2-го рода достигаются при значении коэффициента порога от 1 до 2 и находятся в промежутках 6,41–8,24 % и 3,91–11,55 % соответственно. В табл. 2 представлены детализированные усредненные значения ошибок 1-го и 2-го рода для способа сегментации речь/пауза на основе энергетического анализа с помощью STE для значений коэффициента порога от 1 до 2 с шагом 0,1.

В соответствии с данными в табл. 2 и кривыми зависимости на рис. 4 следует, что наиболее оптимальные значения ошибок 1-го (7,32 %) и 2-го рода (7,46 %) для способа сегментации речь/пауза на основе энергетического анализа с помощью STE достигаются при значении коэффициента порога равном 1,7.

Таблица 2

Детализированные усредненные значения ошибок 1-го и 2-го рода
для способа сегментации речь/пауза на основе
энергетического анализа с помощью STE

Способ сегментации	Ошибка, %	Значение коэффициента порога										
		1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
Способ сегментации речь/пауза на основе энергетического анализа с помощью STE	α	7,32	7,32	7,32	7,32	7,32	7,32	7,32	7,32	7,78	7,78	7,78
	β	7,82	7,82	7,82	7,82	7,82	7,82	7,82	7,46	7,28	7,28	7,10

На рис. 4 представлены кривые зависимости ошибок 1-го и 2-го рода от коэффициента порога.

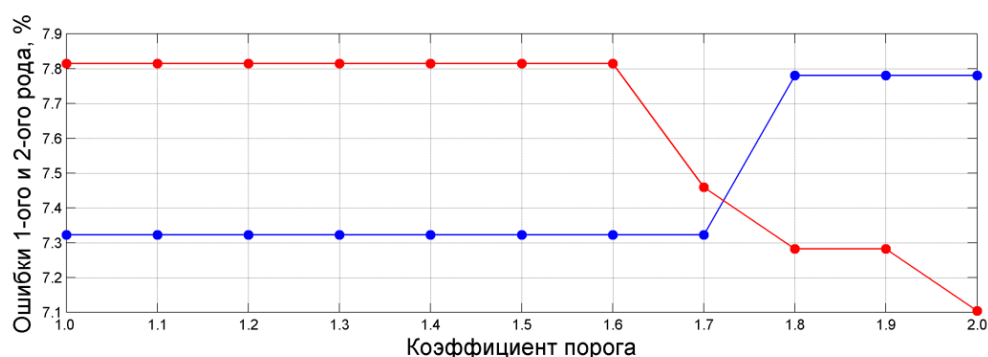


Рис. 4. Зависимость ошибок 1-го и 2-го рода от коэффициента порога для способа сегментации речь/пауза на основе энергетического анализа с помощью STE (линией синего цвета обозначены значения ошибок 1-го рода, линией красного цвета – значения ошибок 2-го рода)

На рис. 5 представлен пример, иллюстрирующий результаты сегментации речь/пауза.

Выводы и перспективы

Подводя итоги анализа, можно сделать следующие выводы:

1. При сравнении оптимальных значений ошибок 1-го и 2-го рода предложенный способ на основе энергетического анализа с помощью ТЕО и STE обеспечивает повышение эффективности сегментации речь/пауза на 5,26 % и 5,51 % соответственно. Это обеспечивается за счет хорошей восприимчивости ТЕО к резким изменениям амплитуды и частоты сигнала.

2. Детализированный анализ результатов сегментации речь/пауза, выявил, что ошибки сегментации 1-го рода в основном наблюдаются в пограничных областях между участками речи и пауз (см. рис. 5). Ошибочно сегментированные участки имеют длительность менее 20 мс. Следовательно, мелкими ошибками сегментации в пограничных областях можно пренебречь.

3. Предложенный способ сегментации речь/пауза может успешно тестироваться в системах оценки психоэмоционального состояния человека, так как обладает хорошей восприимчивостью к резким изменениям амплитуды и частоты сигнала в условиях нестабильной моторики речевого аппарата.

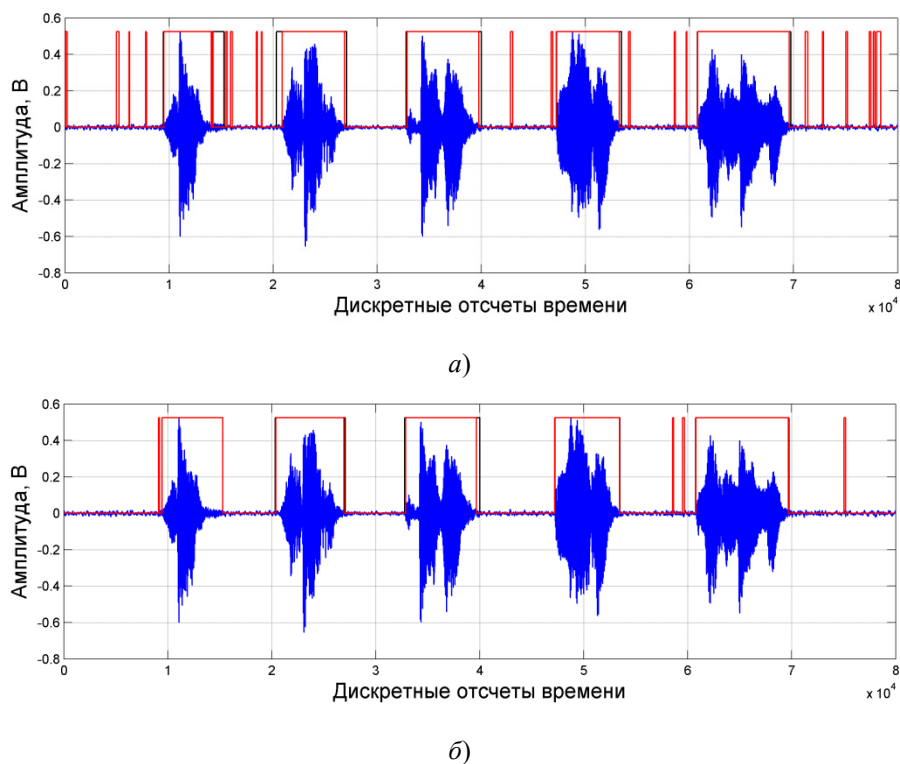


Рис. 5. Пример, иллюстрирующий результаты сегментации речь/пауза: а – способ сегментации речь/пауза на основе энергетического анализа с помощью STE; б – способ сегментации речь/пауза на основе энергетического анализа с помощью TEO и STE (линией красного цвета обозначены достигнутые результаты сегментации, линией черного цвета – результат сегментации, осуществленной вручную)

В перспективе планируется провести дополнительное исследование помехоустойчивости и быстродействия предложенного способа сегментации речь/пауза.

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. Childers D. G., Hand M., Larar J. M. Silent and voiced/unvoiced/ mixed excitation (four-way), classification of speech // IEEE Transaction on ASSP. 1989. Vol. 37, № 11. P. 1771–1774.
3. Atal B., Rabiner L. R. A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition // IEEE Transactions on Speech and Audio Processing. 1976. Vol. 24, № 3. P. 201–212.

4. Kristjansson T., Deligne S., Olsen P. Voicing features for robust speech detection // INTERSPEECH 2005 – Eurospeech, 9th European Conference on Speech Communication and Technology (September 4–8, 2005). Lisbon, Portugal : IEEE, 2005. P. 369–372.
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. Marzinik M., Kollmeier B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics // IEEE Transactions on Speech and Audio Processing. 2002. № 10. P. 109–118.
7. Huang X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Developmen. New Jersey : Prentice Hall, 2001. 980 p.
8. 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.
9. Shin W. H., Lee B. S., Lee Y. K., Lee J. S. Speech/non-speech classification using multiple features for robust endpoint detection // IEEE International Conference on Acoustics, Speech, and Signal Processing : proceedings (Cat. No.00CH37100) (June 5–9, 2000). Istanbul, Turkey : IEEE, 2000. Vol. 1. P. 1399–1402.
10. Wuand G. D., Lin C. T. Word boundary detection with mel scale frequency bank in noisy environment // IEEE Transactions on Speech and Audio Processing. 2000. Vol. 8, № 5. P. 541–554.
11. 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 : IEEE, 1990. Vol. 2. P. 381–384.
12. 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 : IEEE, 1999. Vol. 1. P. 135–138.
13. 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.
14. Алимуратов А. К., Тычков А. Ю., Чураков П. П., Торгашин С. И. Способ сегментации речевого сигнала для систем оценки психогенных состояний // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 3. С. 81–93.
15. Алимуратов А. К., Тычков А. Ю., Чураков П. П. Способ автоматизированной сегментации речевых сигналов для определения временных паттернов естественно выраженных психоэмоциональных состояний // Измерение. Мониторинг. Управление. Контроль. 2019. № 3. С. 48–60.
16. 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.
17. Жуйков В. Я., Харченко А. Н. Алгоритм классификации сегментов речевого сигнала // Электроника и Связь, тематический выпуск «Электроника и нанотехнологии». 2009. Ч. 1, № 2–3. С. 130–137.
18. Bahoura M., Rouat J. Wavelet speech enhancement based on the teager energy operator // IEEE Signal Processing Letter. 2001. Vol. 8, № 1. P. 10–12.
19. 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 : IEEE, 2005. P. 51–61.

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. Childers D.G., Hand M., Larar J.M. Silent and voiced/unvoiced/ mixed excitation (four-way), classification of speech. *IEEE Transaction on ASSP*. 1989;37(11):1771–1774.
3. Atal B., Rabiner L.R. A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition. *IEEE Transactions on Speech and Audio Processing*. 1976;24(3):201–212.
4. Kristjansson T., Deligne S., Olsen P. Voicing features for robust speech detection. *INTERSPEECH 2005 – Eurospeech, 9th European Conference on Speech Communication and Technology (September 4–8, 2005)*. Lisbon, Portugal: IEEE, 2005:369–372.
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. Marzinik M., Kollmeier B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing*. 2002;(10):109–118.
7. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Developmen*. New Jersey: Prentice Hall, 2001:980.
8. 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.
9. Shin W.H., Lee B.S., Lee Y.K., Lee J.S. Speech/non-speech classification using multiple features for robust endpoint detection. *IEEE International Conference on Acoustics, Speech, and Signal Processing: proceedings (Cat. No.00CH37100) (June 5–9, 2000)*. Istanbul, Turkey: IEEE, 2000;1:1399–1402.
10. Wuand G.D., Lin C.T. Word boundary detection with mel scale frequency bank in noisy environment. *IEEE Transactions on Speech and Audio Processing*. 2000;8(5):541–554.
11. 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: IEEE, 1990;2:381–384.
12. 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: IEEE, 1999;1:135–138.
13. Kvedalen E. *Signal Processing Using the Teager Energy Operator and Other Nonlinear Operators: PhD dissertation, Department of Informatics*. Oslo: University of Oslo, 2003:121.
14. Alimuradov A.K., Tychkov A.Yu., Churakov P.P., Torgashin S.I. A method of segmentation of a speech signal for systems of assessment of psychogenic states. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = The institutions of higher education. Volga region. Technical sciences*. 2017;(3):81–93. (In Russ.)
15. Alimuradov A.K., Tychkov A.Yu., Churakov P.P. Method of automated segmentation of speech signals for determining time patterns of naturally expressed psychoemotional states. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control*. 2019;(3):48–60. (In Russ.)
16. 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.

17. Zhuykov V.Ya., Kharchenko A.N. Algorithm of classification of speech signal segments. *Elektronika i Svyaz', tematicheskiy vypusk «Elektronika i nanotekhnologii» = Electronics and Communications, thematic issue "Electronics and Nanotechnologies"*. 2009;1(2–3):130–137. (In Russ.)
18. Bahoura M., Rouat J. Wavelet speech enhancement based on the teager energy operator. *IEEE Signal Processing Letter*. 2001;8(1):10–12.
19. 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: IEEE, 2005:51–61.

Информация об авторах / Information about the authors

Алан Казанферович Алимуратов

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 07.05.2021

Поступила после рецензирования/Revised 08.09.2021

Принята к публикации/Accepted 24.09.2021