

УДК 004.934

**Д.С. Канищев**

Вятский государственный университет, Киров, Россия

## **АВТОМАТИЧЕСКАЯ СЕГМЕНТАЦИЯ РЕЧИ МЕТОДАМИ КЛАСТЕРИЗАЦИИ И С ПРИМЕНЕНИЕМ МЕТОДА ОЦУ**

Под автоматическим распознаванием речи понимается использование различных как аппаратных, так и программных средств для идентификации и обработки человеческой речи. Сегментация речи – это процесс разбиения речевого сигнала на небольшие сегменты. Этот процесс является важной частью систем автоматической обработки речи. Сегментация речи используется не только в системах автоматического распознавания речи, но также в системах работы с лингвистическими корпусами, системах идентификации говорящего и других системах. Ручная сегментация речи – это крайне неэффективный процесс, таким образом, разработка автоматических методов сегментации речевого сигнала является важной задачей. Базовыми единицами сегментации могут быть слова, фонемы и слоги. Слово – наиболее естественная и интуитивно понятная единица сегментации речевого сигнала. В данной работе предложены три новых метода сегментации речевого сигнала на слова. Они основаны на использовании различных методов кластеризации, таких как метод k-средних и метод нечетких c-средних, а также метода Оцу для анализа спектрограммы речевого сигнала. В работе также описан подход к выявлению точных границ слов. Разработанные методы были реализованы для тестирования в среде MatLab. Средняя точность сегментации подхода, основанного на использовании метода k-средних, составила 90,6 %, подхода, основанного на использовании метода нечетких c-средних, – 91 %, подхода, основанного на использовании метода Оцу, – 91,8 %. Полученные результаты были сравнены с другими методами сегментации речевого сигнала. Наиболее перспективным является дальнейшее развитие подхода, основанного на использовании метода Оцу.

**Ключевые слова:** автоматическое распознавание речи, сегментация речевого сигнала, кластеризация, метод k-средних, метод нечетких c-средних, метод Оцу.

**D.S. Kanishchev**

Vyatka State University, Kirov, Russia

## **AUTOMATIC SPEECH SEGMENTATION THAT USES CLUSTERING METHODS AND OTSU'S METHOD**

Automatic speech recognition is the use of computer hardware and software-based techniques to identify and process human voice. Speech segmentation is a process of breaking down a speech signal into smaller units and it is a very important part of many automatic speech processing systems. Speech segmentation is used not only in speech recognition systems, but also in speech corpus collections, speaker verification systems and other. Manual segmentation is a very ineffective process so development of automatic speech segmentation method is an important task. Base speech segmentation units are words, phonemes or syllables. Word is the most natural unit of segmentation. In the paper three new sentence-to-words speech segmentation approaches are proposed. These approaches are

based on using various clustering methods and the Otsu's method for speech signal spectrogram analysis. Clustering methods are the k-means method and the fuzzy c-means method. In the paper also proposed method for words boundaries detection. Developed methods were implemented in MatLab for testing. The average segmentation accuracy rate of the approach that uses k-means clustering method is 90,6 %, 91 % for the approach that uses fuzzy c-means clustering method and 91,8 % for the approach that uses Otsu's method. Achieved results are compared with other speech segmentation methods. Most promising is a further development of the approach that uses Otsu's method.

**Keywords:** automatic speech recognition, speech segmentation, clustering, k-means method, fuzzy c-means method, Otsu's method.

**Введение.** Автоматическая сегментация речи изучается уже более 30 лет [1] и является важной, если не сказать ключевой, частью многих систем распознавания речи. Под задачей сегментации речи подразумевается разделение речевого сигнала на части. Автоматическая сегментация важна для различных алгоритмов автоматизированной обработки речи: распознавание речи, создание лингвистических корпусов, голосовая верификация, а также для решения исследовательских задач в области обработки естественных языков [2, 3]. Традиционный подход в виде ручной сегментации неприемлем для все возрастающих объемов голосовой информации, которая требует обработки. Большие временные затраты, требовательность к вниманию человека, производящего сегментацию вручную, привели к широкому распространению различных автоматических методов сегментации. Можно назвать следующие из них: сегментация при помощи вейвлет-метода [4], методы, основанные на нечеткой логике [5], применение искусственных нейронных сетей [6], а также использование скрытых марковских цепей [7]. В данной статье будут разобраны способы сегментации, полученные в результате использования модифицированных версий методов k-средних, нечетких c-средних, а также алгоритма Оцу.

**Сегментация речи.** Системы распознавания речи требуют разбиения голосового сигнала на дискретные, не перекрывающиеся друг друга звуковые единицы [8, 9]. Это могут быть звуки, слоги, слова, предложения или даже целые диалоги. Слова – наиболее предпочитаемая и естественная единица речи в силу того факта, что слова имеют явно выраженное звуковое представление. Таким образом, мы будем считать слово базовой единицей сегментации в рамках данной работы. Методы автоматической сегментации можно классифицировать по многим различным критериям, но один из самых простых вариантов – разделение на слепые и целевые алгоритмы сегментации [10]. Ключевым различием между этими двумя классами методов является то, насколько

плодотворно метод использует ранее полученную информацию или информацию из внешних источников для обработки речи. Методы слепой сегментации отличаются тем, что не используют никаких источников внешних данных с информацией о языковых особенностях обрабатываемого звукового сигнала. Таким образом, в отсутствие каких-либо внешних данных на первой фазе сегментации эти методы полагаются на звуковые особенности обрабатываемых данных. Вторая фаза же обычно строится на использовании MFCC, LP-коэффициентов или чистого FFT-спектра [11]. В противоположность им целевые методы используют некоторые внешние языковые данные обрабатываемого сигнала для сегментации их на требуемые фрагменты.

**Кластеризация.** Кластеризация – это процесс разбиения набора объектов на несколько групп, именуемых кластерами, таким образом, что объекты в одной группе сильнее похожи друг на друга, чем на объекты из других кластеров. Кластеризация очень широко применяется: распознавание образов [12], интеллектуальный анализ данных [13], машинное обучение [14] и так далее. Алгоритмы кластеризации могут быть классифицированы как четкие, нечеткие, возможностные и вероятностные [15]. Каждая из групп обладает своими особенностями. В данной работе будут использованы четкие и нечеткие алгоритмы кластеризации для получения оптимальных границ сегментов речи. Методы четкой кластеризации жестко ограничивают принадлежность каждого объекта данных к одному кластеру. Один из таких методов – метод *k*-средних. Методы нечеткой кластеризации позволяют объекту быть частью сразу нескольких различных кластеров, причем степень принадлежности варьируется от 0 до 1. Наиболее известный из таких методов – метод нечетких *s*-средних.

**Подготовка к сегментации речевого сигнала.** Как методы кластеризации, так и метод Оцу используются для анализа не оригинального речевого сигнала, а его спектрограммы. В рамках данной работы будет произведен анализ спектрограммы речевого сигнала в наиболее распространенном ее представлении. Это двухмерная диаграмма, вертикальная ось которой является осью частот, горизонтальная – осью времени, интенсивность же каждой точки на спектрограмме характеризует амплитуду речевого сигнала на определенной частоте в определенный момент времени. Итак, оригинальный речевой сигнал представлен на рис. 1.

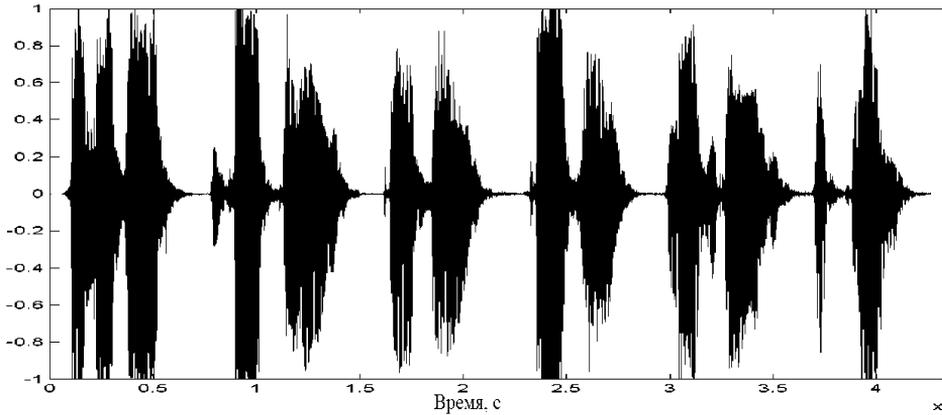


Рис. 1. Оригинальный речевой сигнал

На рис. 2 представлена спектрограмма речевого сигнала. Она представляет из себя изображение, состоящее из оттенков серого. Для того чтобы определить временные границы сегментов, необходимо преобразовать спектрограмму в черно-белое изображение, т.е. определить такой порог, ниже которого точка будет считаться черной, а выше – белой. Для этого могут быть использованы как методы кластеризации, так и метод Оцу.

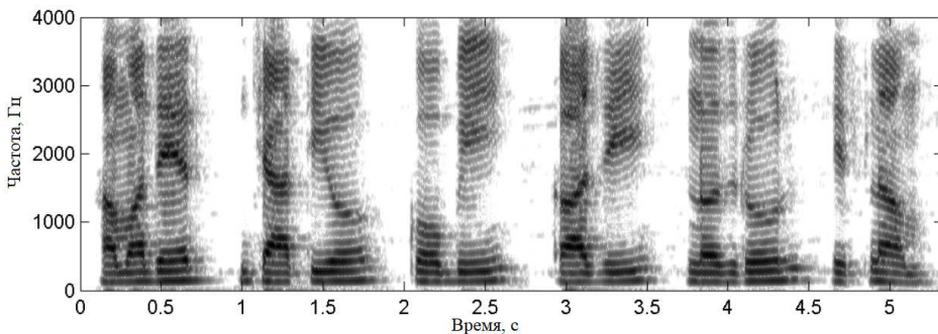


Рис. 2. Спектрограмма речевого сигнала

**Кластеризация методом k-средних.** Метод k-средних является одним из классических методов кластерного анализа. Он используется для распределения  $m$  наблюдений по  $k$  кластерам таким образом, чтобы каждое из наблюдений принадлежало только одному кластеру, к центру которого она наиболее приближено. В данном случае под наблюдениями понимаются точки на спектрограмме, а количество кластеров равно 3. Как показано в работе [16], это наиболее оптимальное

число кластеров для решения подобных задач. Для определения расстояния между наблюдением и центром кластера используется евклидово расстояние:

$$\rho(x, y) = \|x - y\| = \sqrt{\sum_{p=1}^n (x_p - y_p)^2}, x, y \in R^n. \quad (1)$$

Суммарное квадратичное отклонение точек кластеров от их центров должно быть минимальным:

$$\min\left[\sum_{i=1}^k \sum_{x^{(j)} \in S_i} \|x^{(j)} - \mu_i\|^2\right], x^{(j)} \in R^n, \mu_i \in R^n, \quad (2)$$

где  $\mu_i$  – центр кластера  $S_i$ .

Введя все необходимые понятия, можно перейти к формулированию модифицированного для вычисления порога алгоритма:

1. Выбор количества кластеров.
2. Центры данных кластеров выбираются согласно правилу, максимизирующему начальные расстояния между кластерами.
3. Вычисление расстояния от каждого наблюдения до центров всех кластеров.
4. Производим распределение наблюдений по кластерам согласно вычисленным расстояниям (наблюдение теперь относится к ближайшему кластеру).
5. Заново вычисляем центр  $i$ -го кластера, содержащего в себе  $s_i$  элементов:

$$\mu_i = \frac{1}{s_i} \sum_{x^{(j)} \in S_i} x^{(j)}. \quad (3)$$

6. Повторяем с пункта 3 до тех пор, пока центры кластеров, полученные в пункте 4, не изменятся при новом вычислении:

$$\mu_i^{\text{шаг } t} = \mu_i^{\text{шаг } t+1}. \quad (4)$$

7. Вычисляем необходимое нам пороговое значение, в простейшем случае это среднее между конечными значениями центров.

**Кластеризация методом нечетких с-средних.** В то время как метод  $k$ -средних предполагает, что каждый объект принадлежит к одному и только одному кластеру, метод нечетких  $s$ -средних позволяет объекту принадлежать сразу к нескольким кластерам. Сумма степеней принадлежности каждого объекта равняется 1. Чем ближе объект к центру, тем выше эта степень. Для оценки расстояния между  $i$ -м наблюдением и  $j$ -м центром также используется евклидово расстояние

$d_{ij}$ , которое рассчитывается как в (1). Степень принадлежности  $i$ -го наблюдения к  $c$ -му кластеру рассчитывается следующим образом:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}, \quad (5)$$

где  $m$  – экспоненциальный вес. Он влияет на матрицу степеней принадлежности следующим образом: при  $m \rightarrow \infty$  все наблюдения будут принадлежать к каждому кластеру с одинаковой вероятностью. Обычно устанавливается равным 2.

Введя необходимые дополнительные понятия, можно перейти к формулированию модифицированного для вычисления порога алгоритма:

1. Выбор количества кластеров.
2. Центры данных кластеров выбираются согласно правилу, максимизирующему начальные расстояния между кластерами.
3. Вычисление расстояния от каждого наблюдения до центров всех кластеров.
4. Расчет матрицы степеней принадлежности  $U = (\mu_{ij})_{n \times c}$ .
5. Заново вычисляем центр  $j$  – го кластера, содержащего в себе  $s_i$  элементов:

$$v_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m}. \quad (6)$$

6. Повторяем с пункта 3 до тех пор, пока не будет удовлетворено одно из условий:

$$\left| |U^{k+1} - U^k| \right| < e, \quad (7)$$

где  $e$  – некоторое пороговое значение, выбираемое заранее. Второе условие: функция потерь будет не отличаться от рассчитанной на предыдущем шаге:

$$J = \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ij}. \quad (8)$$

**Метод Оцу.** Для применения метода необходимо заранее подготовить гистограмму спектрограммы речевого сигнала  $P(i)$ . Если допустить, что порог бинаризации находится на уровне  $t$ , то относительные частоты для каждого из классов, на которые этот порог делит все точки изображения:

$$w_1(t) = \sum_{i=1}^t P(i), \quad (9)$$

$$w_2(t) = \sum_{i=t+1}^L P(i), \quad (10)$$

где  $L$  – максимальное значение, которое может принять  $t$ . После этого необходимо вычислить средние уровни классов:

$$\mu_1(t) = \sum_{i=1}^t \frac{iP(i)}{w_1(t)}, \quad (11)$$

$$\mu_2(t) = \sum_{i=t+1}^L \frac{iP(i)}{w_2(t)}. \quad (12)$$

Дисперсия внутри каждого из классов:

$$\sigma_1^2(t) = \sum_{i=1}^t [i - \mu_1(t)]^2 \frac{P(i)}{w_1(t)}, \quad (13)$$

$$\sigma_2^2(t) = \sum_{i=t+1}^L [i - \mu_2(t)]^2 \frac{P(i)}{w_2(t)}. \quad (14)$$

Одним из важных моментов, на которых базируется метод Оцу, является тот факт, что минимизация дисперсии внутри класса равносильна максимизации дисперсии между классами.

Дисперсия внутри классов может быть рассчитана как взвешенная сумма дисперсий каждого отдельного класса:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t). \quad (15)$$

Дисперсия между двумя классами:

$$\sigma_b^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2. \quad (16)$$

После чего данная цепочка вычислений проводится для каждого  $t$ . Оптимальный порог соответствует максимальному значению дисперсии между классами.

Разобрав базовые моменты, связанные с методом Оцу, можно перейти к формулированию модифицированного для вычисления порога алгоритма:

1. Подготовка гистограммы спектрограммы речевого сигнала.
2. Выбор порогового значения бинаризации (в простейшем случае вначале выбирается минимальное возможное значение – 1).
3. Расчет относительных частот для классов.
4. Расчет среднего уровня каждого класса.
5. Расчет дисперсий – как внутриклассовой, так и между классами.
6. Выбор нового порога бинаризации (например,  $t_{n+1} = t_n + 1$ ).
7. Повторяем предыдущие пункты, начиная с 3, до тех пор, пока не рассмотрим все возможные пороги бинаризации ( $t = L$ ).
8. В качестве необходимого для дальнейшей обработки спектрограммы выбираем то пороговое значение, при котором было максимальное значение дисперсии между двумя классами и минимальным – значение внутриклассовой дисперсии.

**Расчет границ сегментов.** Используя разработанные выше алгоритмы бинаризации спектрограммы речевого сигнала, можно получить пригодные для дальнейшего анализа изображения.



Рис. 3. Спектрограмма после бинаризации, порог выбран при помощи метода k-средних



Рис. 4. Спектрограмма после бинаризации, порог выбран при помощи метода с-средних

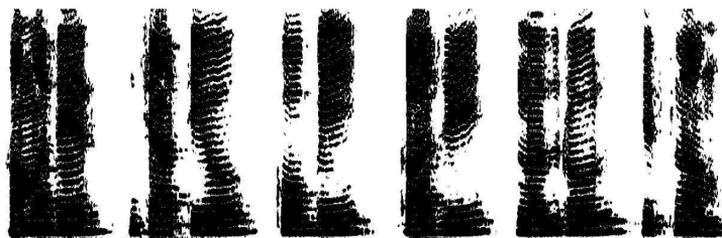


Рис. 5. Спектрограмма после бинаризации, порог выбран при помощи метода Оцу

На рис. 3 – спектрограмма, обработанная при помощи метода k-средних, на рис. 4 – при помощи нечетких с-средних, рис. 5 – при помощи метода Оцу.

Финальный этап – определение границ отдельных сегментов – происходит следующим образом. Каждый столбец исследуется на предмет нахождения в нем черных и белых точек, после чего в зависимости от их соотношения принимается решение о замене столбца на полностью черный или полностью белый. На рис. 6 – пример подобной трансформации. Итоговое сравнение дано на рис. 7.

Был проведен ряд экспериментов для каждого из методов, в ходе которых обрабатывалось 100 предложений от 5 различных человек, в каждом предложении по 5 слов. Результаты представлены в таблице.



Рис. 6. Трансформированная спектрограмма речевого сигнала

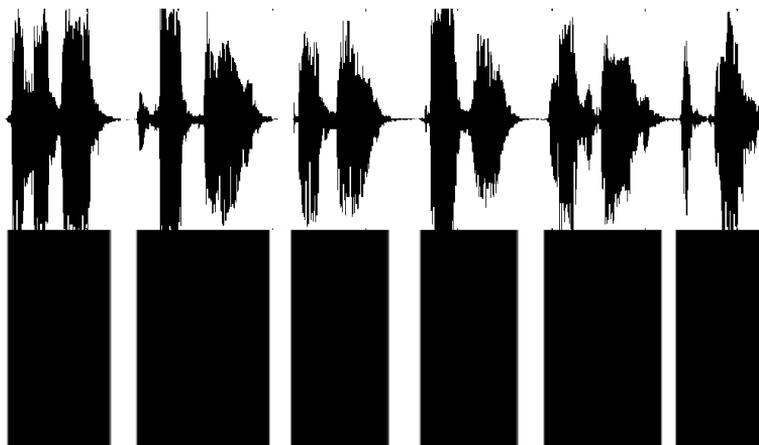


Рис. 7. Исходный речевой сигнал и обработанной спектрограммы

### Результаты экспериментального применения методов

Номер говорящего	Число предложений	Общее число слов	Число правильно выделенных слов и процент точности					
			К-средних		Нечетких с-средних		Оцу	
			#	%	#	%	#	%
№1	100	500	453	90,6	451	90,2	425	85,0
№2	100	500	453	90,6	454	90,8	458	91,6
№3	100	500	454	90,8	452	90,4	466	93,2
№4	100	500	451	90,2	460	92,0	472	94,4
№5	100	500	453	90,6	459	91,8	474	94,8
Всего	500	2500	2264	90,6	2276	91,0	2295	91,8

Средняя точность для метода k-средних 90,6 %, для метода нечетких с-средних 91,0 %, для метода Оцу 91,8 %. Для сравнения с полученными результатами можно обратиться к имеющимся в открытом доступе методикам сегментации речевого сигнала на слова и их эф-

фективности. Подход, основанный на использовании скрытых марковских моделей, показывает эффективность в 90 % правильно выделенных слов [17].

В своей работе авторы используют Hidden Markov Model Toolkit (НТК) для сегментации речевого сигнала. Использование энергетических параметров для выделения слов позволило достичь эффективности в 85 % правильно выделяемых слов [18].

Авторами этой [19] работы было предложено использовать преобразование Гилберта для решения задачи определения участков активной речи. Данный подход показал результат в 88,15 % правильно выделенных слов. Еще один из предложенных вариантов – использование вейвлет-преобразования и показателя энтропии речевого сигнала (Perceptual wavelet entropy neighbor slope, PWENS) для определения участков активной речи [20] – показывает эффективность в 91 %.

**Выводы.** В данной статье были предложены 3 метода сегментации речевого сигнала, основанные на анализе его спектрограммы. Наибольшую эффективность показал метод, построенный на использовании алгоритма Оцу. Его использование повышает эффективность корректного выделения слов в речевом сигнале на 1–3 % относительно существующих современных методов сегментации речи. Предлагается использование разработанного метода для построения программных комплексов, которые, в свою очередь, могут быть использованы и как отдельные системы сегментации речевого сигнала, и как модули более крупных систем анализа речевого сигнала.

### **Библиографический список**

1. Rasanen O. Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture. Helsinki University of Technology. – 2007. – P. 94.

2. Cherif A., Bouafif L., Dabbabi T. Pitch Detection and Formant Analysis of Arabic Speech Processing // Applied Acoustics. – 2001. – Vol. 62. – P. 1129–1140. DOI: 10.1016/S0003-682X(01)00007-X

3. Sharma M., Mammone R. Subword-based text-dependent speaker verification system with user-selectable passwords // IEEE International Conference on Acoustics, Speech and Signal Processing. – 1996. – Vol. 1. – P. 93–96. DOI: 10.1109/ICASSP.1996.540298

4. Hioka Y., Hamada N. Voice activity detection with array signal processing in the wavelet domain // 11<sup>th</sup> European Signal Processing Conference. – 2002. – P. 1–4.

5. Beritelli F., Casale S. Robust voiced/unvoiced speech classification using fuzzy rules // IEEE Workshop on Speech Coding for Telecommunications. – 1997. – P. 5–6. DOI: 10.1109/SCFT.1997.623868

6. Qi Y., Hunt B. R. Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier // IEEE Transactions on Speech and Audio Processing. – 1993. – Vol. 1. – P. 250–255. DOI: 10.1109/89.222883

7. Basu S. A linked-HMM model for robust voicing and speech detection // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03). – 2003. – Vol. 1. – P. 816–819.

8. Thangarajan R., Natarajan M., Selvam M. Syllable modeling in continuous speech recognition for Tamil language // International Journal of Speech Technology. – 2009. – Vol. 12. – P. 47–57. DOI: 10.1007/s10772-009-9058-0

9. Kvale K. Segmentation and Labeling of Speech // Norwegian Institute of Technology. – 1993. – P. 271.

10. Rahman M., Bhuiyan A. Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches // International Journal of Advanced Computer Science and Application (IJACSA). – 2012. – Vol. 3. – P. 131–138.

11. SaiJayram A.K.V., Ramasubramanian V., Sreenivas T.V. Robust parameters for automatic segmentation of speech // IEEE International Conference on Acoustics, Speech and Signal Processing. – 2002. – Vol. 1. – P. 513–516. DOI: 10.1109/ICASSP.2002.5743767

12. Webb A. Statistical Pattern Recognition // John Wiley & Sons, New Jersey. – 2002. – P. 496. DOI: 10.1002/0470854774

13. Tan P.N., Steinbach M., Kumar V. Introduction to Data Mining // Addison-Wesley, Boston. – 2005. – P. 769.

14. Alpaydin E. Introduction to Machine Learning // MIT Press, Cambridge. – 2016. – P. 206. DOI: 10.1017/S0269888906220745

15. Hathway R.J., Bezdek J. Optimization of Clustering Criteria by Reformulation // IEEE Transaction on Fuzzy Systems. – 1995. – Vol. 3. – P. 241–245. DOI: 10.1109/91.388178

16. Philipose S.S. A Triclass Image Segmentation using Adaptive K-means Clustering and Otsu's Method // *International Journal of Engineering Research and General Science*. – 2015. – Vol. 3. – P. 134–138.

17. Shanthi T., Chelva L. Isolated word speech recognition system using HTK // *International Journal of Computer Science Engineering and Information Technology Research*. – 2014. – Vol. 4. – P. 81–86.

18. Hossain A., Nahid N., Khan N.N., Gomes D.C., Mugab S.M. Automatic silence/unvoiced/voiced classification of Bangla velar phonemes: New approach // 8<sup>th</sup> International Conference on Computer and Information Technology, Dhaka. – 2011.

19. Ortiz D., Villa L., Salazar C., Quintero O.L. A simple but efficient voice activity detection algorithm through Hilbert transform and dynamic threshold for speech pathologies // *Journal of Physical: Conference Series*. – 2016. – Vol. 705. – P. 9. DOI: 10.13140/RG.2.1.2140.0406

20. Lee G., Na S.D., Cho J., Kim M.N. Voice activity detection algorithm using perceptual wavelet entropy neighbor slope // *Bio-Medical Materials and Engineering*. – 2014. – Vol. 24. – P. 3295–3301. DOI: 10.3233/BME-141152

## References

1. Rasanen O. *Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture*. Helsinki University of Technology. 2007. 94 p.

2. Cherif A., Bouafif L., Dabbabi T. Pitch Detection and Formant Analysis of Arabic Speech Processing. *Applied Acoustics*, 2001, vol. 62, pp. 1129-1140. DOI: 10.1016/S0003-682X(01)00007-X

3. Sharma M., Mammone R. Subword-based text-dependent speaker verification system with user-selectable passwords. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1996, vol. 1, pp. 93-96. DOI: 10.1109/ICASSP.1996.540298

4. Hioka Y., Hamada N. Voice activity detection with array signal processing in the wavelet domain. *11<sup>th</sup> European Signal Processing Conference*, 2002, pp. 1-4.

5. Beritelli F., Casale S. Robust voiced/unvoiced speech classification using fuzzy rules. *IEEE Workshop on Speech Coding for Telecommunications*, 1997, pp. 5-6. DOI: 10.1109/SCFT.1997.623868

6. Qi Y., Hunt B. R. Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier. *IEEE Transactions on Speech and Audio Processing*, 1993, vol. 1, pp. 250-255. DOI: 10.1109/89.222883

7. Basu S. A linked-HMM model for robust voicing and speech detection. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, 2003, vol. 1, pp. 816-819.

8. Thangarajan R., Natarajan M., Selvam M. Syllable modeling in continuous speech recognition for Tamil language. *International Journal of Speech Technology*, 2009, vol. 12, pp. 47-57. DOI: 10.1007/s10772-009-9058-0

9. Kvale K. Segmentation and Labeling of Speech. Norwegian Institute of Technology, 1993, p. 271.

10. Rahman M., Bhuiyan A. Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches. *International Journal of Advanced Computer Science and Application (IJACSA)*, 2012, vol. 3, pp. 131-138.

11. SaiJayram A.K.V., Ramasubramanian V., Sreenivas T.V. Robust parameters for automatic segmentation of speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2002, vol. 1, pp. 513-516. DOI: 10.1109/ICASSP.2002.5743767

12. Webb A. Statistical Pattern Recognition. John Wiley & Sons, New Jersey, 2002. 496 p. DOI: 10.1002/0470854774

13. Tan P. N., Steinbach M., Kumar V. Introduction to Data Mining. Addison-Wesley, Boston. 2005. p.769.

14. Alpaydin E. Introduction to Machine Learning. MIT Press, Cambridge, 2016. p. 206. DOI: 10.1017/S0269888906220745

15. Hathway R.J., Bezdek J. Optimization of Clustering Criteria by Reformulation. *IEEE Transaction on Fuzzy Systems*, 1995, vol. 3, pp. 241-245. DOI: 10.1109/91.388178

16. Philipose S.S. A Triclass Image Segmentation using Adaptive K-means Clustering and Otsu's Method. *International Journal of Engineering Research and General Science*, 2015, vol. 3, pp. 134-138.

17. Shanthy T., Chelva L. Isolated word speech recognition system using HTK. *International Journal of Computer Science Engineering and Information Technology Research*, 2014, vol. 4, pp. 81-86.

18. Hossain A., Nahid N., Khan N.N., Gomes D.C., Mugab S.M. Automatic silence/unvoiced/voiced classification of Bangla velar phonemes:

New approach. *8<sup>th</sup> International Conference on Computer and Information Technology*, Dhaka, 2011.

19. Ortiz D., Villa L., Salazar C., Quintero O.L. A simple but efficient voice activity detection algorithm through Hilbert transform and dynamic threshold for speech pathologies. *Journal of Physical: Conference Series*, 2016, vol. 705. 9 p. DOI: 10.13140/RG.2.1.2140.0406

20. Lee G., Na S.D., Cho J., Kim M.N. Voice activity detection algorithm using perceptual wavelet entropy neighbor slope. *Bio-Medical Materials and Engineering*, 2014, vol. 24, pp. 3295-3301. DOI: 10.3233/BME-141152

### **Сведения об авторе**

**Канищев Даниил Сергеевич** (Киров, Россия) – аспирант кафедры «Электронные вычислительные машины» Вятского государственного университета (610009, Киров, Московская, 36, e-mail: adelantekang@gmail.com).

### **About the author**

**Kanishchev Daniil Sergeevich** (Kirov, Russian Federation) is a Graduate Student at the department “Computer Engineering” Vyatka State University (610009, Kirov, 36, Moskovskaya str., e-mail: adelantekang@gmail.com).

Получено: 15.04.2019