Системы, сети и устройства телекоммуникаций

УДК 621.396.67

ДЕТЕКТОР ГОЛОСОВОЙ АКТИВНОСТИ В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ ТИПА АКУСТИЧЕСКОГО ШУМА

Кравцов Сергей Андреевич
аспирант кафедры инфокоммуникаций и радиофизики
Ярославского государственного университета им. П.Г. Демидова.
E-mail: whiteyar@yandex.ru.

Топников Артем Игоревич
кандидат технических наук, доцент кафедры инфокоммуникаций и радиофизики
Ярославского государственного университета им. П.Г. Демидова».
E-mail: topnikov@gmail.com.

Приоров Андрей Леонидович
доктор технических наук, доцент кафедры инфокоммуникаций и радиофизики
Ярославского государственного университета им. П.Г. Демидова»
E-mail: andcat@yandex.ru.

Адрес: 150003, Ярославль, ул. Советская, д. 14, каб. 309.

Аннотация: Исследуется возможность повышения точности автоматической классификации акустических шумов, находящихся в аддитивной смеси с речевыми сигналами, за счёт применения детектора голосовой активности. Рассматриваются классификаторы акустических шумов на основе моделей гауссовых смесей и метода опорных векторов. В обоих случаях в качестве признаков речевых сигналов используются мел-частотные кепстральные коэффициенты. Для проведения исследования реализовано два детектора голосовой активности. Первый основан на сравнении логарифма энергии сигнала в окне с порогом. Он выступает в качестве идеализированного детектора, так как в рамках исследования при проведении разметки сигналов использует незашумленный сигнал, недоступный в большинстве реальных задач. Второй основан на применении мел-частотных кепстральных коэффициентов и меры спектральной плоскостности в качестве признаков входного сигнала и моделей гауссовых смесей в качестве классификатора.
На стадии обучения классификаторов рассматривается три сценария: отсутствие предобработки, выделение неречевых фрагментов с использованием порогового детектора голосовой активности, выделение неречевых фрагментов с использованием детектора голосовой активности на основе моделей гауссовых смесей. Аналогичные сценарии осуществляются и на стадии тестирования. Приводятся результаты сравнения работы двух классификаторов типа шума для предложенных сценариев. Демонстрируется возможность повышения точности классификации шумов за счёт детектирования голосовой активности.

Ключевые слова: речевой сигнал, акустический шум, детектор голосовой активности, машинное обучение.

Введение

Задача определения типа акустического шума (автоматической классификации типов акустических шумов) имеет большое значение для современной цифровой обработки речевых сигналов, а также для развития аудиоаналитики [1–4]. Суть задачи состоит в том, чтобы по фрагменту с записью шума определить тип шума или более широко – акустическую среду, в которой сделана запись. Таким образом, классификация типов акустических шумов является важной составляющей распознавания звуковых сцен [5–7]. Информация о шумовых условиях может использоваться для выбора набора параметров алгоритмов шумоподавления, идентификации и распознавания.

Во многих реальных приложениях решение задачи осложняется тем, что акустический шум находится в аддитивной смеси с сигналами, в первую очередь, речевыми. Это негативно сказывается на точности определения типа шума. Логично, что применение детектора голосовой активности (ДГА) способно повысить точность работы классификатора типов шума в описанных условиях [8].

Целью работы является продолжение исследований положительного эффекта от применения детектора голосовой активности в задаче классификации шумов.

База звуковых сигналов

В работе использовалась речевая база, составленная на основе базы «NOIZEUS» [9]. Текстовый материал, основанный на Гарвардских предложениях, прочитан 6 дикторами: 3 мужчинами и 3 женщинами. Общее количество незашумленных файлов в базе равно 30. Запись производилась с частотой дискретизации 25 кГц и последующим понижением до 8 кГц. Средняя длительность каждого сигнала равна 2,5 с.

Также база содержит зашумленные версии исходных сигналов с отношениями сигнал/шум (ОСШ) 0, 5, 10 и 15 дБ. Записи шумов взяты из базы данных AURORA. Шум искусственно добавлялся к речевому сигналу [9]. Для этого случайным образом выделялся сегмент шума той же длины, что и речевой сигнал, масштабировался по амплитуде для достижения определенного ОСШ и добавлялся к исходному сигналу. Набор шумов включает в себя записи, сделанные в условиях, характерных для современного города: шум толпы (babble, crowd of people), автомобиль (car), выставочный зал (exhibition hall), ресторан (restaurant), улица (street), аэропорт (airport), железнодорожная станция (train station), поезд (train).

Классификация типов шумов

Для проведения исследований реализовано два алгоритма определения типа шума: на основе моделей гауссовых смесей (МГС) и на основе метода опорных векторов (МОВ). В обоих случаях в качестве признаков речевого сигнала выступают мел-частотные кепстральные коэффициенты (МЧКК) и мера спектральной плоскостности (МСП). Для вычисления этих параметров входной сигнал разбивался на непересекающиеся окна (фреймы) длиной 128 отсчетов.

На этапе обучения модели и классификации сигналов рассматривались три сценария: без использования детектора голосовой активности, с использованием порогового ДГА и ДГА, основанного на модели гауссовых смесей. Далее более подробно рассмотрим отдельные составные части исследуемых алгоритмов.

Оценка важности признаков

Положительный эффект от использования наиболее важных признаков вместо полного их набора в задачах анализа и обработки речевых сигналов продемонстрирован в [10, 11]. Поэтому в данном исследовании наряду с полным набором признаков предлагается рассмотреть использование их сокращённого набора.

Для оценки важности признаков использовался метод на основе решающих деревьев, который относится к классу логических методов. Их основная идея состоит в объединении простых решающих правил. Одна из особенностей решающих деревьев заключается в том, что они позволяют определить важность всех используемых признаков. Важность признака можно оценить на основе того, как сильно улучшился критерий Джини благодаря использованию этого признака в вершине дерева.

На рис. 1 приведены усреднённые значения важности используемых признаков для ОСШ от 0 до 15 дБ, где первый признак – МСП, со 2-го по 25-й – 24 мел-частостных кепстральных коэффициента.

Далее в работе использовались признаки как с важностью более 0,031 в задаче определения шума (мера спектральной плоскостности и первые 13 мел-частотных кепстральных коэффициентов), так и весь рассчитанный вектор признаков.

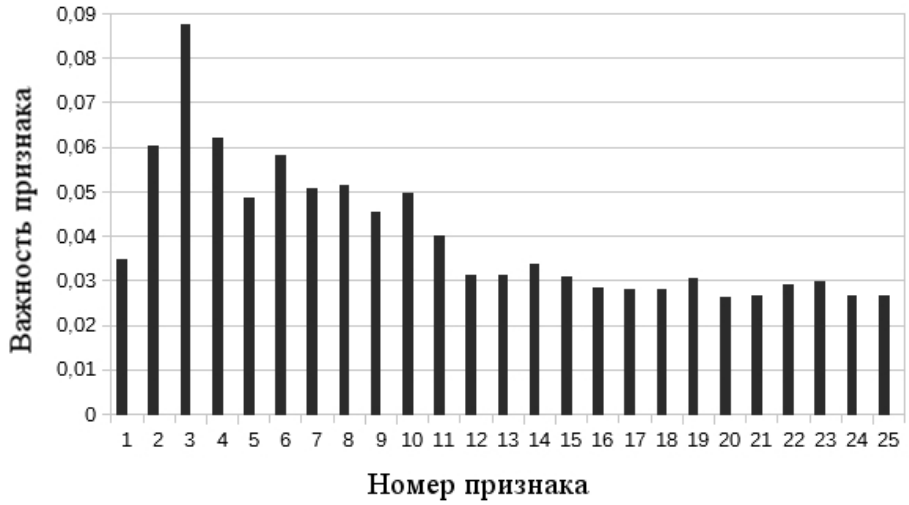

Рис. 1. Важность признаков для задачи определения типа акустического шума

Пороговый алгоритм детектирования

Для разметки тестовой базы реализован пороговый детектор голосовой активности. Он выступает в качестве идеализированного алгоритма, так как в рамках исследования ему всегда известен незашумленный сигнал, недоступный в большинстве реальных задач. Этот детектор голосовой активности, описанный в работе [12], принимает решение на основе сравнения энергии фрейма с заданным пороговым значением. С его помощью речевая база предварительно (до зашумления) размечается на речесодержащие фреймы и паузы следующим образом: для каждого фрейма рассчитывается логарифм энергии:

$$E_t = 10\log_{10}\left(\frac{1}{N-1}\sum_{n=1}^{N}\left(x_t[n]-\mu_t\right)^2+\varepsilon\right),$$

где $\mu_t = \frac{1}{N}\sum_{n=1}^{N}x[n]$, $x_t[n]$ – отсчёт входного сигнала с номером n во фрейме с номером t, N – размер фрейма, $\varepsilon = 10^{-16}$ вводится для избегания вычисления логарифма ноля.

Далее вычисленное значение энергии фрейма сравнивается с пороговым значением:

$$\text{flag}_t = \begin{cases} 1, & \text{если } \left(E_t > E_{max} - \theta_{main}\right) \wedge \left(E_t > \theta_{min}\right) \\ 0, & \text{в ином случае} \end{cases},$$

где θ_{main} и θ_{min} обозначают заданные начальные и минимальные пороговые значения энергии соответственно.

В [12] пороговые значения выбраны следующим образом: $\theta_{min} = -55$ дБ и $\theta_{main} = 30$ дБ. При реализации алгоритма замечено, что более точное определение границ речи достигается при $\theta_{main} = 40$ дБ, данное значение и используется в работе. Полученное значение flag используется в качестве метки. Значение flag = 1 соответствует речесодержащему фрейму, 0 – паузе.

Пример работы порогового ДГА для фразы «Hats are worn to tea and not to dinner» из речевой базы изображен на рис. 2.

Данный алгоритм выбран для разметки тестовых незашумленных сигналов, так как прост в реализации, не требует обучения и имеет высокую точность выделения голосовой активности в незашумлённых речевых сигналах.

Детектор голосовой активности на основе МГС

Также для проведения исследования выбран детектор голосовой активности на основе МГС, хорошо зарекомендовавший себя для решения разных задач в области речевой обработки [13–15]. Для его обучения и тестирования речевая база размечена с помощью идеализированного порогового ДГА. Далее рассчитанные векторы признаков разделялись на две группы: речесодержащие фреймы и фреймы с паузами. На основе полученных входных

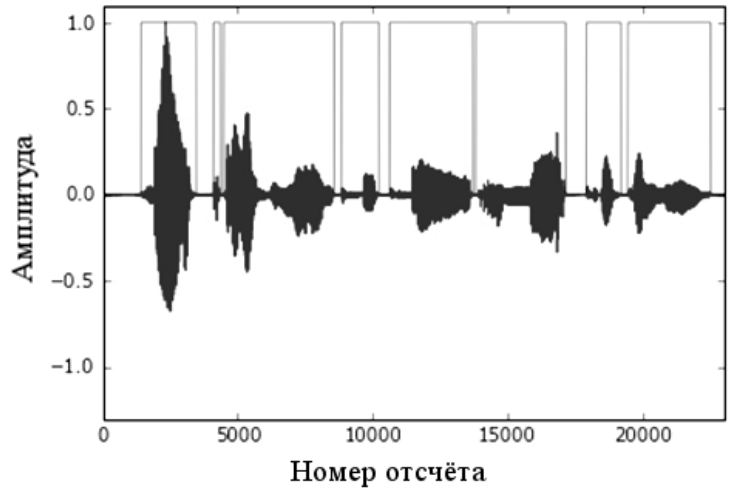

Рис. 2. Детектирование голосовой активности с помощью порогового ДГА

векторов и с использованием алгоритма максимизации ожидания (expectation maximization) построено две модели: для речи и для паузы. Количество гауссиан равно 128.

Значения равновероятной ошибки (Equal Error Rate – EER, %) для ДГА при различных значениях ОСШ приведены в таблице 1.

Таблица 1. Зависимость равновероятной ошибки от ОСШ

ОСШ, дБ	EER, %
0	12,5
5	10,0
10	7,5
15	5,0

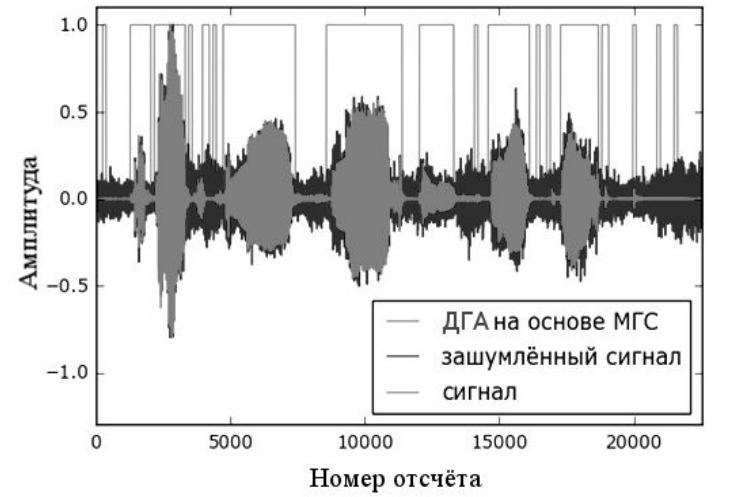

Рис. 3. Детектирование голосовой активности с помощью ДГА на основе МГС

На рис. 3 изображён пример работы обученного ДГА на зашумленном сигнале для фразы «The birch canoe slid on the smooth planks» с шумом поезда и ОСШ 5 дБ.

На рис. 4 отображены основные этапы обучения детектора голосовой активности на основе моделей гауссовых смесей.

Обучение модели и определение типа шума

На этапе обучения для каждого шума построено 3 модели. Их различие вызвано содержанием обучающей выборки, которая в зависимости от сценария может включать:

1. Все речевые сигналы.
2. Фрагменты сигналов, классифицированных как не содержащие речь пороговым ДГА.
3. Фрагменты сигналов, классифицированных как не содержащие речь ДГА на основе МГС.

Для предотвращения переобучения применялась перекрестная проверка (cross-validation). Для этого имеющаяся выборка данных разбивалась на 3 части. Затем на одной части данных производилось обучение моделей классификатора, а оставшиеся две использовались для тестирования. Таким образом, в имеющейся базе зашумленных речевых сигналов выделялось 40 зашумленных сигналов для каждого источника шума, по 10 для 4 уровней ОСШ. Оставшиеся 80 сигналов использовались в качестве тестового набора сигналов. Стоит отметить, что при подобном разбиении в тестовую выборку не попадали дикторы, чьи речевые сигналы использовались на этапе обучения.

На этапе тестирования также исследовалось 3 сценария. Согласно

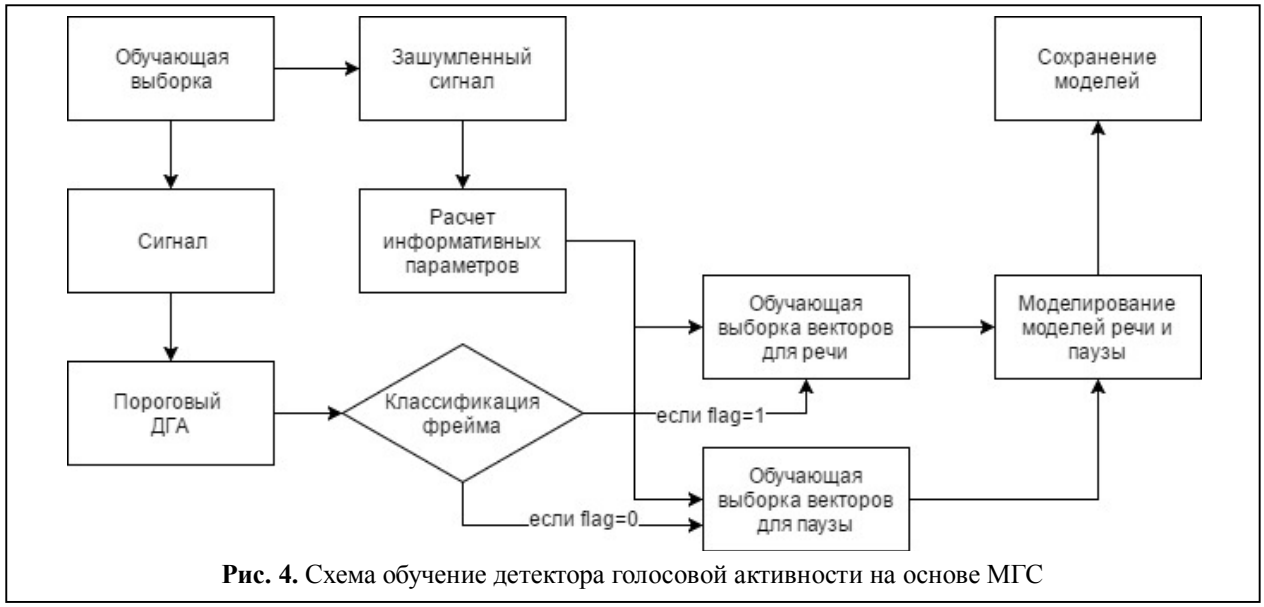

Рис. 4. Схема обучение детектора голосовой активности на основе МГС

им на вход алгоритма классификации шума поступали:

1. Все речевые сигналы.
2. Фрагменты сигналов, классифицированные как не содержащие речь пороговым ДГА.
3. Фрагменты сигналов, классифицированные как не содержащие речь ДГА на основе МГС.

Методика исследования

В качестве показателя качества работы классификатора использовалась точность определения типа шума, задаваемая следующим образом:

$$I = \frac{a}{\Sigma},$$

где I – точность определения типа шума, a – число правильно классифицированных сигналов, Σ – общее число тестовых сигналов. Для удобства этот показатель выражался в процентах.

Число правильно классифицированных тестовых сигналов определяется следующим образом:

1. Производится построение модели шума, используя обучающие сигналы.
2. Каждый тестовый сигнал подается на вход системы классификации шума. Если система правильно определяет тип шума, то число правильно классифицированных тестовых сигналов увеличивается на единицу.

Результаты исследования

С использованием описанных выше алгоритмов и базы звуковых сигналов получены значения точности определения типа шума при использовании алгоритмов классификации на основе МГС и МОВ, разных наборов признаков и различных комбинаций обучающих и тестовых выборок (таблица 2).

Видно, что наибольшая точность 93,9% достигается при использовании классификатора на основе МГС и реализации сценария 1с, то есть когда обучение модели шума производилось на полной выборке, а на стадии классификации шумовые фрагменты предварительно выделялись ДГА на основе МГС. Значения точности определения типа шума для этого сценария при разных значениях ОСШ приведены в таблице 3.

В случае использования только МСП достигается минимальная усреднённая точность. Это можно объяснить недостаточностью информативности одного конкретного признака в поставленной задаче, что приводит к высоким значениям ошибки определения большинства типов шумов, кроме типов «аэропорт» (airport) и «поезд» (train).

Таблица 2. Значения точности определения типов шума, %

Признаки/ Классификатор		МСП и МЧКК			МСП			МЧКК			МСП и 13 МЧКК		
	Выборка Обучение/ классификация	1	2	3	1	2	3	1	2	3	1	2	3
МГС	a	83,3	63,8	76,6	24,5	24,7	24,5	81,6	65,9	76,7	85,8	64,5	76,1
	b	92	87,2	90,8	26	28,6	28,5	90,9	88,9	90	84,5	84,8	83,8
	c	**93,9**	90,2	91,9	25,8	28,5	27,5	**93,1**	88,6	91,6	86,1	85,2	**87,2**
МОВ	a	78	80,6	78,1	24,2	24,2	24,2	75,3	82,7	75,6	72,2	81,1	72,2
	b	80,5	90,3	80,5	23,7	27	23,7	79,5	91,4	79,7	76,9	87	77
	c	84,7	**92,8**	84,7	22,1	19,9	22,3	84,4	**92,5**	84,5	79,1	**89,1**	78,8

В целом, применение детектора голосовой активности при классификации увеличивает точность определения типа шума в среднем на 4 процентных пункта (п.п.) для порогового ДГА и 6 п.п. для ДГА на основе МГС.

Заключение

Таким образом, реализованы и исследованы классификаторы акустических шумов на основе моделей гауссовых смесей и метода опорных векторов, использующие мел-частотные кепстральные коэффициенты в качестве признаков звуковых сигналов. Также реализованы два детектора голосовой активности. Исследование показало возможность повышения точности определения типа шума за счёт применения детектора голосовой активности. Среднее увеличение точности определения типа шума при применении ДГА на основе моделей гауссовых смесей составило около 6 п.п.

Таблица 3. Значения точности определения типов шума для разных ОСШ, %

Шум	ОСШ, дБ			
	0	**5**	**10**	**15**
airport	90	95	95	95
babble	100	100	100	95
car	100	100	100	100
exhibition	100	100	100	100
restaurant	70	85	90	85
station	90	90	85	75
street	80	100	85	100
train	100	100	100	100

Литература

1. *Gaunard P.*, *Mubikangiey C.G.*, *Couvreur C.*, *Fontaine V.* Automatic classification of environmental noise events by hidden Markov models // Applied acoustics. 1998. V. 54. Is. 3. Pp. 187–206.
2. *Eamdeelerd C.*, *Songwatana K.* Audio noise classification using bark scale features and k-nn technique // International symposium on communications and information technologies (ISCIT 2008). 2008. Pp. 131–134.
3. *Ma L.*, *Smith D.*, *Milner B.* Environmental noise classification for context-aware applications // International conference on database and expert systems applications. Springer, Berlin, Heidelberg, 2003. Pp. 360–370.
4. *Uzkent B.*, *Barkana B.D.*, *Yang J.* Automatic environmental noise source classification model using fuzzy logic // Expert systems with applications. 2011. V. 38, Is. 7. Pp. 8751–8755.
5. *Barchiesi D.*, *Giannoulis D.*, *Stowell D.*, *Plumbley M.D.* Acoustic scene classification: Classifying environments from the sounds they produce // IEEE Signal processing magazine. 2015. V. 32. Is. 3. Pp. 16–34.
6. *Chu S.*, *Narayanan S.*, *Kuo C.C.*, *Mataric M.* Where am I? Scene recognition for mobile robots using audio features // 2006 IEEE International conference on multimedia and expo. IEEE. 2006. Pp. 885–888.
7. *Bisot V.*, *Serizel R.*, *Essid S.*, *Richard G.* Acoustic scene classification with matrix factorization

for unsupervised feature learning // IEEE International conference on acoustics, speech and signal processing (ICASSP). 2016. Pp. 6445–6449.

8. *Кравцов С.А.*, *Топников А.И.* Применение детектора голосовой активности в задаче классификации акустических шумов // 12-я международной научно-техническая конференция Перспективные технологии в средствах передачи информации (ПТСПИ-2017). В 2-х томах. Т. 2. 2017. С. 151–154.

9. *Ни Y.*, *Loizou P.* Subjective evaluation and comparison of speech enhancement algorithms // Speech communication. 2007. V. 49. Pp. 588–601.

10. *Кравцов С.А.*, *Топников А.И.*, *Приоров А.Л.* Оценка значимости акустических признаков в задаче детектирования голосовой активности // Цифровая обработка сигналов. 2016. № 2. С. 9–13.

11. *Кравцов С.А.*, *Топников А.И.* Анализ работы линейных классификаторов в задаче детектирования речевой активности // DSPA: Вопросы применения цифровой обработки сигналов. 2016. Т. 6. № 2. С. 388–392.

12. *Kinnunen T.*, *Rajan P.* A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data // International conference on acoustics, speech and signal processing (ICASSP). 2013. Pp. 7229–7233.

13. *Топников А.И.*, *Веселов И.А.*, *Новоселов С.А.*, *Приоров А.Л.* Выделение речевых команд на основе помехоустойчивых параметров и моделей гауссовых смесей // Проектирование и технология электронных средств. 2011. № 4. С. 31–35.

14. *Кравцов С.А.*, *Тупицин Г.С.*, *Топников А.И.*, *Сагациян М.В.*, *Приоров А.Л.* Исследование работы детектора речевой активности в задаче идентификации диктора // Радиотехнические и телекоммуникационные системы. 2015. № 4 (20). С. 61–68.

15. *Кравцов С.А.*, *Топников А.И.*, *Приоров А.Л.* Детектор речевой активности на основе голосующих моделей гауссовских смесей // Электромагнитные волны и электронные системы. 2015. Т. 20. № 8. С. 29–34.

Поступила 9 октября 2018 г.

English

VOICE ACTIVITY DETECTOR IN THE TASK OF DETERMINING THE TYPE OF ACOUSTIC NOISE

Sergey Andreyevich Kravtsov – Post-graduate Student of the Department of Infocommunications and Radiophysics; P. G. Demidov Yaroslavl State University.

E-mail: whiteyar@yandex.ru.

Artem Igorevich Topnikov – Candidate of Technical Sciences, Associate Professor, Department of Infocommunications and Radiophysics; P. G. Demidov Yaroslavl State University.

E-mail: topnikov@gmail.com.

Andrey Leonidovich Priorov – Doctor of Technical Sciences, Associate Professor, Department of Infocommunications and Radiophysics; P. G. Demidov Yaroslavl State University.

E-mail: andcat@yandex.ru.

Address: 150003, Yaroslavl, Sovetskaya str., 14.

Abstract: Accuracy enhancement possibility of automatic classification of acoustic noise in the additive mixture with speech signals by using the voice activity detector is investigated. Acoustic noise classifiers based on Gaussian mixture models and support vector method are examined. Mel-frequency cepstral coefficients (MFCC) are used as speech signal features in both cases. Two voice activity detectors were also implemented for the research. The first one is based on the comparison of the signal energy logarithm in the viewport with the threshold. It serves as an idealized detector, as it uses the noise-free signal during signal marking in the research which is inaccessible in most real-world tasks. The second one is based on the use of mel-frequency cepstral coefficients and the spectral flatness measure as input signal features and the Gaussian mixture models as the classifier. Three scenarios are considered at the phase of training classifiers: no preprocessing, non-speech fragments spotting by using the voice activity threshold detector, non-speech fragments spotting by using the voice activity detector based on Gaussian mixture models. Similar scenarios are implemented at the testing phase. The determination accuracy of the noise type which is equal to the ratio of the number of correctly classified signals to the total number of test signals is calculated for each run of measurements. The operation comparison results of two noise type classifiers for the considered scenarios that are associated with the use or non-use of voice activity detectors both in training and testing are presented. Accuracy enhancement possibility of acoustic noise classification through the voice activity detection is shown.

Keywords: speech signal, acoustic noise, voice activity detector, machine learning.

References

1. *Gaunard P.*, *Mubikangiey C.G.*, *Couvreur C.*, *Fontaine V.* Automatic classification of environmental noise events by hidden Markov models. Applied acoustics. 1998. Vol. 54. Is. 3. Pp. 187–206.

2. *Eamdeelerd C.*, *Songwatana K.* Audio noise classification using bark scale features and k-nn technique. International symposium on communications and information technologies (ISCIT 2008). 2008. Pp. 131–134.

3. *Ma L.*, *Smith D.*, *Milner B.* Environmental noise classification for context-aware applications. International conference on database and expert systems applications. Springer, Berlin, Heidelberg, 2003. Pp. 360–370.

4. *Uzkent B.*, *Barkana B.D.*, *Yang J.* Automatic environmental noise source classification model using fuzzy logic. Expert systems with applications. 2011. V. 38, Is. 7. Pp. 8751–8755.

5. *Barchiesi D.*, *Giannoulis D.*, *Stowell D.*, *Plumbley M.D.* Acoustic scene classification: Classifying environments from the sounds they produce. IEEE Signal processing magazine. 2015. V. 32. Is. 3. Pp. 16–34.

6. *Chu S.*, *Narayanan S.*, *Kuo C.C.*, *Mataric M.* Where am I? Scene recognition for mobile robots using audio features. 2006 IEEE International conference on multimedia and expo. IEEE. 2006. Pp. 885–888.

7. *Bisot V.*, *Serizel R.*, *Essid S.*, *Richard G.* Acoustic scene classification with matrix factorization for unsupervised feature learning. IEEE International conference on acoustics, speech and signal processing (ICASSP). 2016. Pp. 6445–6449.

8. *Kravtsov S.A.*, *Topnikov A.I.* Voice activity detector application in acoustic noise classification. 12-th International Scientific and Technical Conference on advanced technologies in data transmission facilities (ATDTF-2017). In 2 volumes. Vol. 2. 2017. Pp. 151–154.

9. *Hu Y.*, *Loizou P.* Subjective evaluation and comparison of speech enhancement algorithms. Speech communication. 2007. Vol. 49. Pp. 588–601.

10. *Kravtsov S.A.*, *Topnikov A.I.*, *Priorov A.L.* Significance of acoustic features in voice activity detection. Signal digital processing. 2016. No. 2. Pp. 9–13.

11. *Kravtsov S.A.*, *Topnikov A.I.* Linear classifier performance analysis in speech activity detection. DSPA: Digital Signal Processing Application . 2016. Vol.6. No. 2. Pp. 388–392.

12. *Kinnunen T.*, *Rajan P.* A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data // International conference on acoustics, speech and signal processing (ICASSP). 2013. Pp. 7229–7233.

13. *Topnikov A.I.*, *Veselov I.A.*, *Novoselov S.A.*, *Priorov A.L.* Voice command spotting through noise-free parameters and Gaussian mixture models. Proektirovanie i tekhnologiya elektronnyh sredstv. 2011. No. 4. Pp. 31–35.

14. *Kravtsov S.A.*, *Tupitsin G.S.*, *Topnikov A.I.*, *Sagatsiyan M.V.*, *Priorov A.L.* Study of voice activity detector for the speaker identification. Radiotekhnicheskiye i telekommunikatsionnye sistemy. 2015. No. 4 (20). Pp. 61–68.

15. *Kravtsov S.A.*, *Topnikov A.I.*, *Priorov A.L.* Speech activity detector based on voice models of Gaussian mixtures. Elektromagnitnye volny i elektronnye sistemy. 2015. Vol. 20. No. 8. Pp. 29-34.