

ИССЛЕДОВАНИЕ РАБОТЫ ДЕТЕКТОРА РЕЧЕВОЙ АКТИВНОСТИ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ ДИКТОРА

Кравцов Сергей Андреевич

аспирант кафедры динамики электронных систем
Ярославского государственного университета им. П.Г. Демидова.
E-mail: sk860@outlook.com.

Тупицин Геннадий Сергеевич

аспирант кафедры динамики электронных систем
Ярославского государственного университета им. П.Г. Демидова.
E-mail: genichyar@genichyar.com.

Топников Артем Игоревич

кандидат технических наук, ассистент кафедры динамики электронных систем
Ярославского государственного университета им. П.Г. Демидова.
E-mail: topnikov@gmail.com.

Сагациян Максим Владимирович

аспирант кафедры динамики электронных систем
Ярославского государственного университета им. П.Г. Демидова.
E-mail: smw.pantheon@bk.ru.

Приоров Андрей Леонидович

доктор технических наук, доцент кафедры динамики электронных систем
Ярославского государственного университета им. П.Г. Демидова.
E-mail: andcat@yandex.ru.

Адрес: 150000, г. Ярославль, ул. Советская, д. 14.

Аннотация: Исследован алгоритм детектирования речевой активности на основе голосующих групп в задаче идентификации диктора, особенностью которого является построение нескольких моделей для отдельных диапазонов отношения сигнал/шум и использование процесса голосования. Модель речи и модель паузы, рассчитанные для определенного диапазона зашумления, образуют группу. Наличие или отсутствие речевой активности в исследуемом фрагменте определяется каждой группой отдельно на этапе сравнения значений логарифмической функции правдоподобия. Итоговое решение принимается на основе сравнения суммы голосов всех групп с пороговой величиной. Получены значения точности идентификации диктора для различных ОСШ с применением алгоритма шумоподавления, использующего функцию коррекции спектра Винера, и детектирования речевой активности.

Ключевые слова: речевой сигнал, шумоподавление, идентификация диктора, детектор речевой активности.

Введение

Создание помехоустойчивых систем обработки речи [1, 2] является актуальной задачей. Так, задача идентификации диктора может быть решена достаточно надежно в чистых акустических условиях, однако присутствие фонового шума способно ухудшить точность идентификации. Это происходит из-за несоответствия признаков диктора, полученных во время обучения в чистых акустических условиях, и при-

знаков, наблюдаемых в присутствии шума [3]. Кроме того, при низких значениях отношения сигнал/шум (ОСШ, Signal to Noise Ratio, SNR) значительно снижается эффективность работы детектора речевой активности (ДРА, Voice Activity Detector, VAD), задачей которого является определение границ речесодержащих фрагментов входного звукового сигнала.

Несмотря на множество работ, направленных на совершенствование алгоритмов шумо-

подавления и детектирования речевой активности, исследования в этой области не теряют научной и практической актуальности.

В области подавления шума в речевых сигналах наибольшее распространение получили методы, которые работают в частотной области и используют различные функции коррекции спектра (ФКС), зависящие от апостериорного отношения сигнал/шум и/или оценки априорного ОСШ [4–10].

В области детектирования речевой активности преобладают решения, основанные на методах машинного обучения, например, использующие в качестве классификатора аппарат моделей гауссовых смесей (МГС, Gaussian Mixture Models, GMM) [2, 11–13]. Эффективность таких методов значительно превосходит простые пороговые детекторы [14, 15].

Целью данного исследования является повышение качества работы системы идентификации диктора с использованием алгоритма шумоподавления и оригинального алгоритма детектирования речевой активности.

Подавление шума в речевых сигналах

Рассмотрим речевой сигнал, искаженный аддитивным шумом. Обозначим модуль спектра зашумленной речи как R_k , а незашумленной – A_k . Задачу подавления шума можно решить путем нахождения \hat{A}_k – оценки модуля спектра незашумленного сигнала. Это можно сделать с помощью поэлементного умножения модуля спектра зашумленного сигнала R_k на некоторую ФКС G_k [5]:

$$\hat{A}_k = G_k \cdot R_k,$$

Известно, что ФКС часто является функцией от априорного ξ_k и/или апостериорного γ_k ОСШ, которые определяются следующим образом:

$$\xi_{p,k} = \frac{\lambda_{p,k}^A}{\lambda_k^D},$$

где $\lambda_{p,k}^A = E\{A_{p,k}^2\}$ – спектральная плотность незашумленного сигнала в окне p и частотной

полосе k , $\lambda_k^D = E\{D_{p,k}^2\}$ – спектральная плотность шума в частотной полосе k .

В данной работе будет использоваться ФКС Винера [9]:

$$G_{Wiener}(\xi_{p,k}) = \frac{\xi_{p,k}}{1 + \xi_{p,k}}.$$

Функция коррекции спектра является оптимальной по критерию минимума среднеквадратичной ошибки оценки комплексного спектра чистого сигнала [9].

Для оценки априорного ОСШ при известной спектральной плотности шума λ_k^D и неизвестной спектральной плотности незашумленного сигнала $\lambda_{p,k}^A$ наиболее часто применяют подход прямого принятия решения [5]. При его использовании априорное ОСШ оценивается на основании апостериорного ОСШ в текущем окне и оценки незашумленного сигнала в предыдущем окне:

$$\hat{\xi}_{p,k} = \alpha \frac{\hat{A}_{p-1,k}^2}{\lambda_k^D} + (1 - \alpha) \max(0, \gamma_{p,k} - 1),$$

где $\max(x, y)$ – функция, определяющая большее из двух аргументов; α – параметр алгоритма, часто принимаемый равным 0,98.

Детектирование речевой активности

В работе ДРА используется взвешенное коллективное решение бинарных классификаторов на основе МГС. В качестве информативных признаков, рассчитываемых в каждом сегменте (окне) сигнала, выбраны мел-частотные кепстральные коэффициенты (МЧКК, Mel-Frequency Cepstral Coefficients, MFCC) [16] и мера спектральной плоскостности (МСП, Spectral Flatness Measure, SFM). Входной сигнал разбивается на непересекающиеся фрагменты длиной 15 мс, далее для каждого фрагмента рассчитывается 12-мерный вектор МЧКК и значение МСП:

$$SFM = 10 \log_{10} \left(\frac{\bar{G}}{\bar{A}} \right),$$

где \bar{A} – среднее арифметическое, \bar{G} – среднее геометрическое амплитудного спектра фрагмента сигнала [17].

На этапе обучения весь диапазон исследуемых отношений сигнал/шум делится на интервалы шириной 5 дБ. Таким образом, в задаче детектирования речевой активности при диапазоне ОСШ от -5 до 20 дБ выделяются 6 интервалов. Далее обучаются модель речи и модель пауз для каждого отношения сигнал/шум, количество гауссиан – 16. По результатам обучения строятся 12 моделей. Каждую пару моделей (одна модель – для речи, другая – для паузы), рассчитанных при определенном отношении сигнал/шум, будем называть группой. Таким образом, получаем 6 групп моделей.

Описанная схема обучения используется в случае отсутствия предварительного шумоподавления. При использовании шумоподавления входной сигнал с известным ОСШ обрабатывается алгоритмом шумоподавления, после чего обучаются модели паузы и речи.

Наличие или отсутствие речевой активности в исследуемом фрагменте определяется каждой группой отдельно на основе сравнения значений логарифмической функции правдоподобия. Если в i -й группе более вероятна принадлежность к модели паузы, то голос группы $V_i = 0$, иначе – $V_i = 1$.

В данном случае количество голосов $Vote$ может варьироваться от 0 (ни одна из групп не классифицировала фрагмент как содержащий

речь) до 6 (во всех группах была вероятнее модель речи).

Итоговое решение принимается на основе сравнения суммы голосов всех групп со значением порога VT . Если значение $Vote$ больше или равно VT , то анализируемый фрагмент отмечается как содержащий речь, иначе – как пауза.

Также получены результаты работы алгоритма идентификации диктора для 9 групп в ДРА. Для обучения 9 голосующих групп использовались тестовые сигналы речевой базы с ОСШ от -15 до 25 дБ.

Система идентификации диктора

Данная система имеет два режима работы: режим регистрации пользователей в системе (режим обучения) и режим идентификации [16]. Блок-схема работы системы идентификации диктора в режиме обучения представлена на рис. 1. Каждый пользователь предварительно регистрируется в системе, записав образец своего голоса. В блоке предобработки производится очистка входного речевого сигнала от шума и выделение речевой активности с помощью описанных ранее алгоритмов. Далее производится извлечение из речевого сигнала релевантных (для задачи распознавания диктора) признаков. На последнем этапе обучения строится модель диктора.

Блок-схема работы системы в режиме идентификации представлена на рис. 2. Извлеченные из входного речевого сигнала признаки

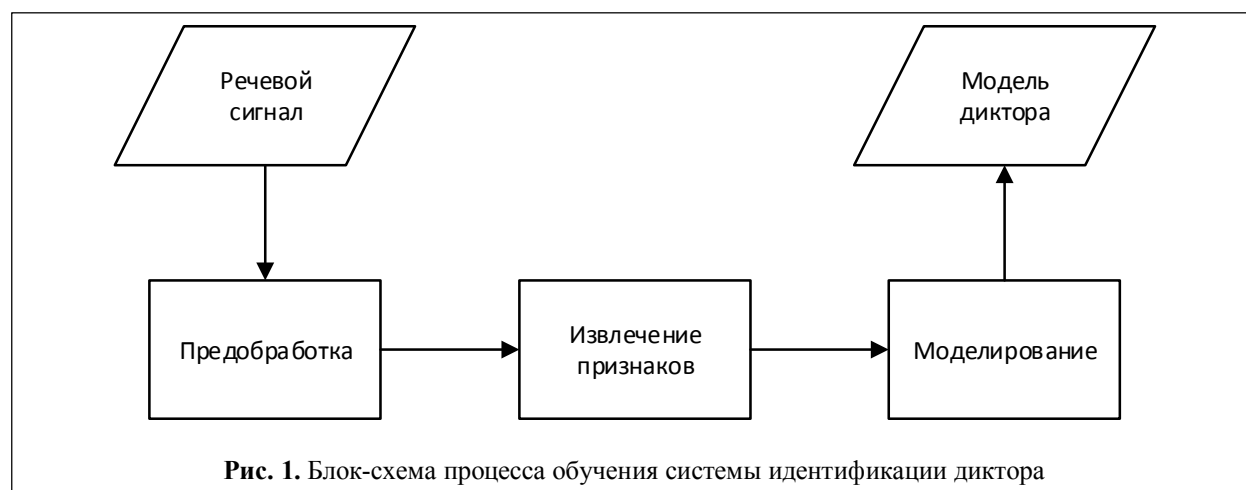


Рис. 1. Блок-схема процесса обучения системы идентификации диктора

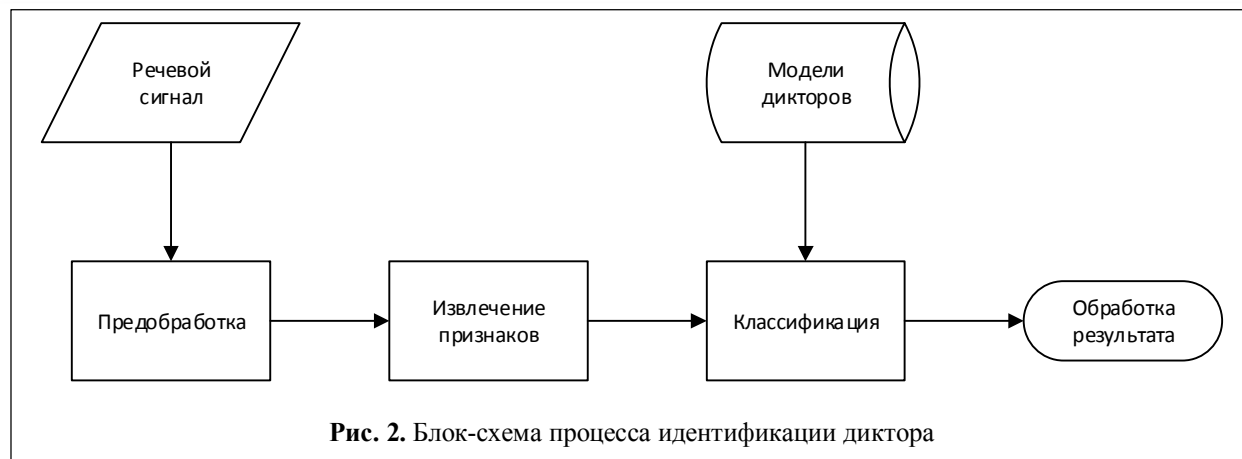


Рис. 2. Блок-схема процесса идентификации диктора

поступают на вход блока классификации, где определяется, какому диктору соответствует представленный речевым сигналом образец голоса.

Для анализа работы модифицированного алгоритма шумоподавления и сравнения его с исходным в среде Matlab моделировалась система идентификация диктора. Во всех тестах речевой сигнал разбивался на перекрывающиеся окна длиной 25 мс, что при стандартной в цифровой телефонии частоте дискретизации в 8000 Гц соответствует 200 отсчетам. Величина перекрытия окон составляла 60 % от длины окна. При получении спектра сигнала использовалось окно Хэмминга.

Для эффективного выделения уникальных признаков голоса из речевого сигнала использовались МЧКК. Банк треугольных фильтров построен для диапазона частот от 0 до 3700 Гц. Для каждого окна рассчитывались 24 коэффициента и брались коэффициенты со 2-го по 16-й.

Для классификации тестовых сигналов применялись модели гауссовых смесей [16] со 128 компонентами. Отметим, что для расчета модели не применялся стандартный EM-алгоритм: вместо него использовалась универсальная фоновая модель (Universal Background Model – UBM), которая обучалась на 110 дикторах (по 30 с речи на каждого). Все дикторы – мужчины.

Результаты исследования

Для тестирования использовалась база дикторов со следующими параметрами: язык – рус-

ский; число мужчин – 33; число женщин – 0; длительность обучающего сигнала для каждого диктора с учетом пауз – 120 с; число тестовых сигналов для каждого диктора – 5; длительность каждого тестового сигнала с учетом пауз – 4 с. Речевая база составлена авторами самостоятельно.

В качестве критерия качества системы идентификации диктора в работе используется такая величина, как точность идентификации диктора (ТИД). Она определяется следующим образом:

$$\text{ТИД} = \frac{\text{Число правильно идентифицированных тестов}}{\text{Общее число тестовых сигналов}}$$

Обычно ТИД выражается в процентах.

Перед проведением исследования тестовые сигналы могут обрабатываться с помощью алгоритма шумоподавления и детектора речевой активности. Для зашумления тестовых сигналов используется аддитивный белый гауссовский шум (АБГШ). В первом эксперименте применялся ДРА, использующий 6 групп голосующих моделей. Исследовался случай как с предварительным шумоподавлением, так и без него (таблица 1).

Указанные значения ТИД при использовании ДРА для каждого ОСШ рассчитаны как среднее значение ТИД для данного ОСШ (таблицы 2, 4). Применение описанного ДРА в поставленной задаче увеличивает ТИД на 1,1–12 процентных пункта (п. п.) для различных ОСШ, а в случае использования алгоритма предварительного шумоподавления ТИД выше на 1,3–4 п. п. для ОСШ от -5 до 15 дБ.

Таблица 1. Зависимость ТИД (%) от ОСШ

ОСШ, дБ	без шумоподавления		с шумоподавлением	
	без ДРА	с ДРА	без ДРА	с ДРА
-5	5,5	6,6	14,6	15,9
0	10,3	14,6	32,7	34,0
5	29,7	39,5	46,7	50,7
10	46,1	58,1	66,1	67,7
15	67,9	74,1	77,0	78,7
20	79,4	84,4	87,3	84,6

Для более детального анализа исследовалась зависимость ТИД от величины порога VT , задаваемого для алгоритма детектирования речевой активности. Рассматривались два случая: без предварительного шумоподавления (таблицы 2, 3) и с ним (таблицы 4, 5). Результаты приводятся для двух вариантов ДРА: с использованием 6-ти (таблицы 2, 4) и 9-ти моделей голосующих групп (таблицы 3,5). Жирным шрифтом выделены наилучшие результаты для конкретных значений ОСШ.

Таблица 2. Зависимость ТИД (%) от величины порога VT для случая без предварительного шумоподавления и ДРА с 6 голосующими группами

ОСШ, дБ	Порог VT					
	1	2	3	4	5	6
-5	4,8	4,9	4,9	6,1	9,7	9,1
0	10,3	10,3	12,7	14,6	17,0	22,4
5	29,7	32,1	34,6	41,8	47,9	50,9
10	48,5	50,3	58,2	64,2	64,2	63,0
15	71,5	72,7	73,3	75,8	75,2	76,4
20	82,4	84,2	85,5	84,9	84,9	84,8

Установлено, что для отношения сигнал/шум -5, 0 дБ и 9 голосов невозможно оценить точность идентификации диктора (табл. 3), так как сигнал, полученный после

обработки ДРА, пустой (не содержит речь). То есть, речевых фрагментов, получивших 9 голосов ($Vote=9$), нет.

Видно, что при использовании алгоритма шумоподавления до ДРА ТИД выше на 0,1-19,5 п. п. (в зависимости от ОСШ) с 6 голосующими группами и 0,5–19,6 п. п. с 9 голосующими группами по сравнению с идентификацией диктора без фильтрации входного сигнала.

Так как при увеличении порога VT уменьшается ошибка I-го рода (False positive, пауза

Таблица 3. Зависимость ТИД (%) от величины порога VT для случая без предварительного шумоподавления и ДРА с 9 голосующими группами

ОСШ, дБ	Порог VT								
	1	2	3	4	5	6	7	8	9
-5	5,5	5,5	5,5	5,5	5,5	7,9	8,5	10,9	–
0	10,3	10,3	10,3	12,1	12,7	15,2	21,2	24,9	–
5	29,7	29,7	29,7	33,3	40,0	46,7	49,1	49,7	40,0
10	46,7	47,3	49,1	58,2	63,0	61,8	63,0	60,6	55,8
15	69,3	69,7	71,5	74,6	76,4	75,2	75,8	76,4	75,8
20	79,4	80,0	82,4	86,1	85,5	84,9	84,9	84,9	86,1

распознается как речевой сегмент), и увеличивается ошибка II-го рода (False negative, речевой сегмент оценен как пауза), можно сделать вывод, что, чем ниже ОСШ, тем больше влияние неречевых сегментов на точность идентификации диктора.

Таблица 4. Зависимость ТИД (%) от величины порога VT для случая с предварительным шумоподавлением и ДРА с 6 голосующими группами

ОСШ, дБ	Порог VT					
	1	2	3	4	5	6
-5	14,6	14,5	15,2	15,2	17,6	18,2
0	33,9	34,6	33,9	32,7	35,2	33,9
5	51,5	50,3	50,9	50,9	50,9	49,7
10	66,1	67,3	69,1	68,5	67,3	67,9
15	78,8	79,4	78,2	78,8	78,8	78,2
20	85,5	84,9	85,5	84,9	83,6	83,0

Таблица 5. Зависимость ТИД (%) от величины порога VT для случая с предварительным шумоподавлением и ДРА с 9 голосующими группами

ОСШ, дБ	Порог VT								
	1	2	3	4	5	6	7	8	9
-5	15,6	13,9	15,2	14,6	13,3	15,2	18,2	17,0	18,8
0	33,9	34,6	34,6	32,7	33,9	33,3	35,2	33,3	36,4
5	50,3	50,9	50,3	50,3	50,9	51,5	52,1	52,1	50,9
10	67,9	67,3	66,7	67,3	66,7	67,9	68,5	67,9	67,3
15	78,2	80,0	80,0	78,2	78,2	79,4	78,8	77,6	75,2
20	86,7	85,5	85,5	84,2	83,6	84,2	83,6	83,6	81,8

При этом использование большего количества голосующих групп позволило увеличить значение ТИД в среднем на 1 п. п. Также можно сделать вывод о том, что для широкого диапазона значений ОСШ невозможно подобрать фиксированный порог VT , максимизирующий значение ТИД.

Заключение

Таким образом, полученные экспериментальные результаты показывают преимущество использования детектора речевой активности в поставленной задаче идентификации диктора, особенно в условиях сильного фонового шума. Применение ДРА увеличивает ТИД на 1,1-12 п. п. без шумоподавления, а в случае использования алгоритма предварительного шумоподавления ТИД выше на 1,3-4 п. п. для ОСШ от -5 до 15 дБ. Замечено, что чем ниже ОСШ, тем больше влияние неречевых сегментов на ТИД.

Литература

1. 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.
2. Кропотов Ю.А., Бейлекчи Д.В., Белов А.А., Ермолаев В.А., Карасев О.Е., Колпаков А.А., Коноплев А.Н., Проскуряков А.Ю. Информационно-управляющие телекоммуникационные системы аудиообмена и автоматизированного мониторинга// Радиотехнические и телекоммуникационные системы. 2014. № 1. С. 99-114.

3. May T., van de Par S., Kohlrausch A. Noise-robust speaker recognition combining missing data techniques and universal background modeling // IEEE Trans. Audio, Speech, Lang. Process. 2012. V. 20, N. 1. P. 108-121.

4. Boll S.F. Suppression of acoustic noise in speech using spectral subtraction // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1979. V. 27. Is. 2. P. 113-120.

5. Ephraim Y., Malah D. Speech enhancement using a minimum mean-square error

- short-time spectral amplitude estimator // IEEE Trans. Acoustics, Speech, and Signal Processing. 1984. V. 32. N. 6. P. 1109-1121.

6. Scalart P., Filho J.V. Speech enhancement based on a priori signal to noise estimation // IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96). 1996. V. 2. P. 629-632.

7. Plapous C., Marro C., Scalart P., Mauuary L. A Two-Step Noise Reduction Technique // IEEE Intl. Conf. Acoust., Speech, Signal Processing. 2004. V. 1. P. 289-292.

8. Plapous C., Marro C., Scalart P. Speech Enhancement Using Harmonic Regeneration // IEEE Intl. Conf. Acoust., Speech, Signal Processing. 2005. V. 1. P. 157-160.

9. Lim J.S., Oppenheim A.V. Enhancement and bandwidth compression of noisy speech // Proceedings of the IEEE. 1979. 12. P. 197-210.

10. Сагацян М.В., Тупицин Г.С., Кравцов С.А., Приоров А.Л. Повышение эффективности коллективного нейросетевого алгоритма на основе обучения SCG в задаче дикторонезависимого распознавания речевых команд в условиях шумов // Информационные системы и технологии. 2015. № 4. С. 39-46.

11. Вознесенская Т.В., Котов М.А., Леднов Д.А. Гибридный детектор речи // Цифровая обработка сигналов. № 4. 2014. С. 53-56.

12. Rosen O., Mousazadeh S., Cohen I. Voice activity detection in presence of transient noise using spectral clustering and diffusion kernels // IEEE 28th Convention of Electrical & Electronics Engineers in Israel. 2014. P. 1-5.

13. Mamiya Y., Yamagishi J., Watts O., Clark R., King S., Stan A. Lightly supervised GMM VAD to use audiobook for speech synthesizer // IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. P. 7987-7991.

14. Brady P.T. A technique for investigating On-Off patterns of speech // Bell System Technical Journal. 1965. V. 44. N. 1. P. 1-22.

15. Rabiner L.R., Sambur M.R. An algorithm for determining the endpoints of isolated utterances // Bell Systems Technical Journal. 1975. V. 54. N. 2. P. 297–315.

16. Первушин Е.А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование. 2011. № 24. С. 41–54.

Работа выполнена при финансовой поддержке РФФИ (грант № 15-08-99639-а).

Поступила 01 октября 2015 г.

17. Moattar M.H., Homayounpour M.M. A simple but efficient real-time voice activity detection algorithm // 17th European Signal Processing Conference. 2009. P. 2549–2553.

English

Speech activity detector operation research for speaker recognition

Sergey Andreevich Kravtsov – Graduate Student Department of Electronic Systems Dynamics P.G. Demidov Yaroslavl State University.

E-mail: sk860@outlook.com.

Gennady Sergeevich Tupitsin – Graduate Student Department of Electronic Systems Dynamics P.G. Demidov Yaroslavl State University.

E-mail: genichyar@genichyar.com.

Artem Igorevich Topnikov – Candidate of Technical Sciences, Teaching Assistant Department of Electronic Systems Dynamics P.G. Demidov Yaroslavl State University.

E-mail: topnikov@gmail.com.

Maxim Vladimirovich Sagatsiyan – Graduate Student Department of Electronic Systems Dynamics P.G. Demidov Yaroslavl State University.

E-mail: smw.pantheon@bk.ru.

Andrey Leonidovich Priorov – Doctor of Engineering, Associate Professor Department of Electronic Systems Dynamics P.G. Demidov Yaroslavl State University.

E-mail: andcat@yandex.ru.

Address: 150000, Yaroslavl, Sovetskaya Str., 14.

Abstract: Problem of speaker's recognition can be solved rather accurately in pure acoustic conditions, however background noise can reduce recognition accuracy. This research goal is to improve quality of speaker's recognition system operation with use of noise suppression algorithm and speech activity detection algorithm (SAD). Method based on Wiener spectrum function correction is used in noise suppression algorithm. Speech activity detection original algorithm is applied. Its operation applies balanced collective settlement of binary classifiers on the basis of Gaussian mixture models of (GMM). Mel frequency cepstral coefficients and spectral flatness measure are taken as informative features which are evaluated in analyzed signal fraction. Development process of several models for separate investigated signal-to-noise ratio range (SNR) and voicing process application enabling to obtain more accurate results of speech availability in signal fraction is SDA algorithm feature. On-off speech patterns evaluated for a certain noise contamination range form a group; range of noise contamination range is 5 dB. Each signal fraction is evaluated by all groups. If probability logarithm for speech pattern is higher, group voice is 1, otherwise - 0.

The final decision is made on the basis of comparison of all groups' voices total number with threshold value. The obtained experimental results based on GMM with use of SAD and noise suppression algorithm for SNR 5 dB to 20 dB indicate advantage of speech activity detector application for assigned problem, especially in environment of considerable background noise. SAD improves speaker recognition accuracy (SRA) by 1,1-12 p.p. without noise suppression, and in case of preliminary noise suppression algorithm application SRA is 1,3-4 p.p. for SNR it is 5 to 15 dB.

Key words: voice signal, noise suppression, speaker recognition, speech activity detector.

References

1. Topnikov A.I., Veselov I.A., Novoselov S. A., Priorov A.L. Selection of speech utterances on the basis of noise proof parameters and Gaussian mixture models. - *Proyektirovaniye i tekhnologiya elektronnykh sredstv*. 2011. No. 4. pp. 31-35.

2. Kropotov Yu.A., Beylekchi D.V., Belov A.A., Yermolaev V.A., Karasyov O. E., A.A Kolpakov., Konoplev A.N., Proskuryakov A.Yu. Information Management telecommunication systems of audio exchange and automated monitoring. - Radiotekhnicheskiye i telekommunikatsionnye sistemy. 2014. No. 1. pp. 99-114.
3. May T., van de Par S., Kohlrausch A. Noise-robust speaker recognition combining missing data techniques and universal background modeling. - IEEE Trans. Audio, Speech, Lang. Process. 2012 V. 20, N. 1. P. 108-121.
4. Boll S.F. Suppression of acoustic noise in speech using spectral subtraction. - IEEE Transactions on Acoustics, Speech, and Signal Processing. 1979. V. 27. Is. 2. P. 113-120.
5. Ephraim Y., Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. - IEEE Trans. Acoustics, Speech, and Signal Processing. 1984. V. 32. N. 6. P. 1109-1121.
6. Scalart P., Filho J.V. Speech enhancement based on a priori signal to noise estimation. - IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96). 1996. V. 2. P. 629-632.
7. Plapous C., Marro C., Scalart P., Mauuary L. A Two-Step Noise Reduction Technique. - IEEE Intl. Conf. Acoust., Speech, Signal Processing. 2004. V. 1. P. 289-292.
8. Plapous C., Marro C., Scalart P. Speech Enhancement Using Harmonic Regeneration. - IEEE Intl. Conf. Acoust., Speech, Signal Processing. 2005. V. 1. P. 157-160.
9. Lim J.S., Oppenheim A.V. Enhancement and bandwidth compression of noisy speech. - Proceedings of the IEEE. 1979. 12. P. 197-210.
10. Sagatsiyan M. V., Tupitsin G. S., Kravtsov S. A., Priorov A.L. Efficiency enhancement of collective neural network algorithm on the basis of SCG training for speech utterances recognition irrespective of speaker in noise environment. - Informatsionnye sistemy i tekhnologii. 2015. No. 4. P. 39-46.
11. Voznesenskaya T.V., Kotov M.A., Lednov D.A. Hybrid speech detector Tsifrovaya obrabotka signalov. No. 4. 2014. P. 53-56.
12. Rosen O., Mousazadeh S., Cohen I. Voice activity detection in presence of transient noise using spectral clustering and diffusion kernels. - IEEE 28th Convention of Electrical & Electronics Engineers in Israel. 2014. P. 1-5.
13. Mamiya Y., Yamagishi J., Watts O., Clark R., King S., Stan A. Lightly supervised GMM VAD to use audio-book for speech synthesizer. - IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. P. 7987-7991.
14. Brady P.T. A technique for investigating On-Off patterns of speech//Bell System Technical Journal. 1965. V. 44. N. 1. P. 1-22.
15. Rabiner L.R., Sambur M.R. An algorithm for determining the endpoints of isolated utterances. - Bell Systems Technical Journal. 1975. V. 54. N. 2. P. 297-315.
16. Pervushin E.A. Review of speaker recognition main methods. - Matematicheskiye struktury i modelirovaniye. 2011. No. 24. P. 41-54.
17. Moattar M.H., Homayounpour M.M. A simple but efficient real-time voice activity detection algorithm. - 17th European Signal Processing Conference. 2009. P. 2549-2553.