

УДК 004.934.2+534.784

ДЕТЕКТИРОВАНИЕ БУЛЬБАРНЫХ НАРУШЕНИЙ ПРИ БОКОВОМ АМИОТРОФИЧЕСКОМ СКЛЕРОЗЕ НА ОСНОВЕ АНАЛИЗА РЕЧЕВОГО СИГНАЛА

А.Д. ГВОЗДОВИЧ¹, Ю.Н. РУШКЕВИЧ², М.И. ВАШКЕВИЧ¹

¹Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

²Республиканский научно-практический центр неврологии и нейрохирургии, Республика Беларусь

Поступила в редакцию 27 июня 2018

Аннотация. В работе предложен подход к анализу речевого сигнала для выявления случаев бульбарных нарушений при боковом амиотрофическом склерозе. В качестве признаков наличия патологических отклонений в голосе использовались результаты формантного анализа гласных звуков /а/ и /и/, расстояние между огибающими данных гласных звуков, а также информация о паузах между словами в тестовом речевом сигнале. Для выполнения классификации использовался алгоритм линейного дискриминантного анализа.

Ключевые слова: речевой сигнал, боковой амиотрофический склероз, диагностика, классификация.

Abstract. The approach to analysis of speech signal in order to amyotrophic lateral sclerosis diagnosis is proposed. As a features of pathological changes in the voice the results of a format analysis of vowel /aa/ and /iy/, the distance between the envelopes of these vowel, and information about pauses between words in the test speech signal were used. To perform the classification, the linear discriminant analysis was used.

Keywords: speech signal, amyotrophic lateral sclerosis, diagnosis, classification.

Doklady BGUIR. 2018, Vol. 116, No. 6, pp. 52-58
Detection of bulbar amyotrophic lateral sclerosis based on speech analysis
A.D. Gvozdovich, Y.N. Rushkevich, M.I. Vashkevich

Введение

Ощутимые изменения в речи присущи многим неврологическим заболеваниям. В частности, бульбарные нарушения (т. е. трудности с глотанием и речью) являются первым симптомом в приблизительно 30 % случаев заболевания боковым амиотрофическим склерозом (БАС) [1]. В большинстве случаев диагностика отклонений в речи выполняется медицинским работником, однако такой способ оценки чувствителен к различным источникам ошибок и не может быть автоматизирован. Более того, некоторые изменения в речи, наблюдаемые при БАС, выявляются только с привлечением инструментальных средств [2]. В настоящее время существует необходимость в улучшении систем оценки изменений в речи для повышения эффективности ранней диагностики БАС и дальнейшего терапевтического лечения [3]. Трудности в достоверной диагностике БАС и других неврологических заболеваний побудили исследователей к разработке средств поддержки принятия решений, основанных на алгоритмах, которые позволяют разделить образцы речевых сигналов здоровых людей от имеющих то или иное неврологическое заболевание. Данные средства не позволяют выполнить дифференциальную диагностику (средство дифференциальной диагностики предполагает определение конкретного типа неврологического заболевания), но являются важным этапом на пути к этой долгосрочной цели.

Особенности построения систем диагностики по речевому сигналу

Одной из проблем в выявлении БАС является отсутствие стандартизированной процедуры речевой диагностики [3]. В большинстве случаев системы диагностики построены на основе речевого теста с протяжным произнесением гласных звуков. Исходя из этого основными признаками, извлекаемыми из речевого сигнала, являются частота основного тона (F_0) и производные от нее – джиттер (степень вариации F_0 между соседними периодами основного тона) и шиммер (степень вариации амплитуды между соседними периодами основного тона) [4]. Часто в качестве параметров используются мел-кепстральные коэффициенты [5], а также параметр отношения гармоник-шум (англ. HNR – harmonic-to-noise ratio), характеризующий наличие аэроакустического шума, возникающего в результате неполного смыкания голосовых связок у неврологических больных. Однако данные параметры малоэффективны, если их применять в случае, когда тестовый сигнал представляет собой запись беглой речи. В данной работе для диагностики неврологических заболеваний предлагаются новые признаки, которые извлекаются из тестового сигнала беглой речи.

Признаки, используемые для диагностики неврологических заболеваний по голосу

В результате неврологических заболеваний происходят изменения в стимуляции мускулатуры вообще и мышц языка в частности. С точки зрения звукообразования важным является положение языка по горизонтали («вперед» или «сзади») и по вертикали («высокий» подъем или «низкий» подъем), см. рис. 1 [6].

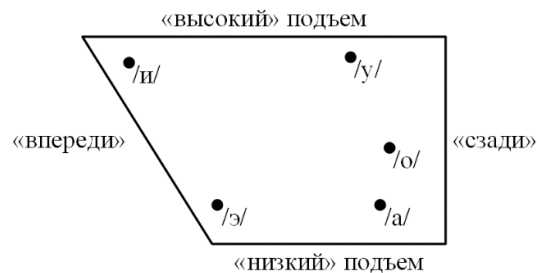


Рис. 1. Относительное положение языка при произнесении гласных звуков /и/, /у/, /о/, /а/, /э/

Для выявления признаков неврологических заболеваний из речевого сигнала целесообразно выбрать звуки /а/ и /и/, поскольку для их формирования требуется значительная работа мышц языка. Для практического сравнения из тестового речевого сигнала с записью счета от 1 до 10 выбирались близкие по времени участки, содержащие звуки /а/ и /и/, после чего выполнялось сравнение взаимной формантной структуры этих звуков (как правило, звуки выбирались из слов «один», «два», «три»). Пример формантной структуры звуков /а/ и /и/, произнесенных здоровым человеком, показан на рис. 2, а.

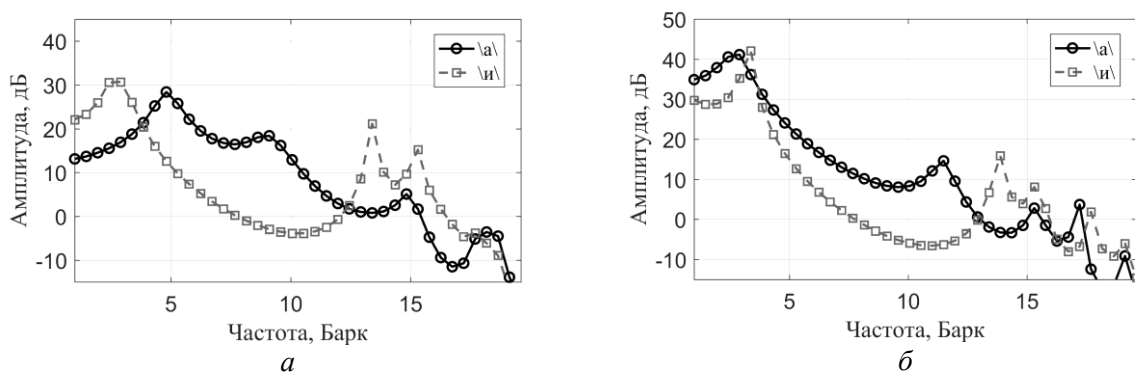


Рис. 2. Взаимное расположение формантных частот звуков /а/ и /и/ :
а – здоровый человека; б – пациент с БАС

Анализ рис. 2, а показывает, что форманты имеют высокую добротность, значительно разнесены на частотной оси и расположены в следующей последовательности $F_i(1) < F_a(1) < F_a(2) < F_i(2)$. Правильный порядок формант у больных БАС может быть нарушен (рис. 2, б). В случаях, когда правильный порядок расположения формант не нарушен, наблюдается существенное сближение формантных частот. Замечено, что у здоровых людей расстояние между первыми и вторыми формантами звуков /а/ и /и/ составляет 2 Барка и более, у больных с БАС это расстояние меньше (см. рис. 3).

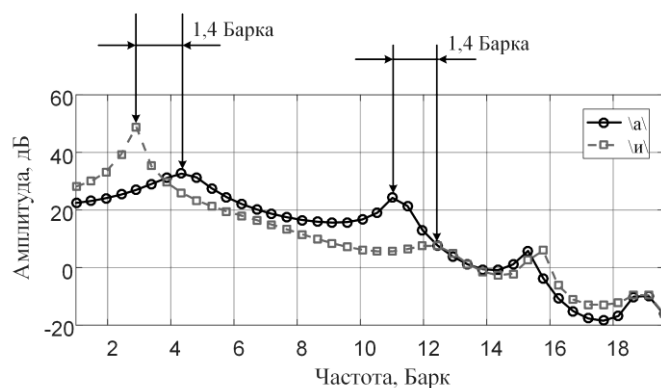


Рис. 3. Пример сближения формантных частот звуков /а/ и /и/ у пациента с БАС

Для количественной оценки степени нарушения взаимной формантной структуры звуков /а/ и /и/ в [7] предложено использовать следующую меру:

$$fnt_{err}(F_i, F_a) = \begin{cases} 2, & \text{если } F_i(1) > F_a(1) \text{ или } F_a(2) > F_i(2) \\ 2 - \frac{F_a(1) - F_i(1)}{2} - \frac{F_i(2) - F_a(2)}{2}, & \text{если } F_i(1) - F_a(1) < 2 \text{ и } F_i(2) - F_a(2) < 2 \\ 1 - \frac{F_a(1) - F_i(1)}{2}, & \text{если } F_a(1) - F_i(1) < 2 \\ 1 - \frac{F_i(2) - F_a(2)}{2}, & \text{если } F_i(2) - F_a(2) < 2 \\ 0, & \text{иначе.} \end{cases} \quad (1)$$

Помимо изменений во взаимной формантной структуре звуков /а/ и /и/, у пациентов наблюдалось увеличение сходства между формами огибающих этих звуков (см. рис. 4).

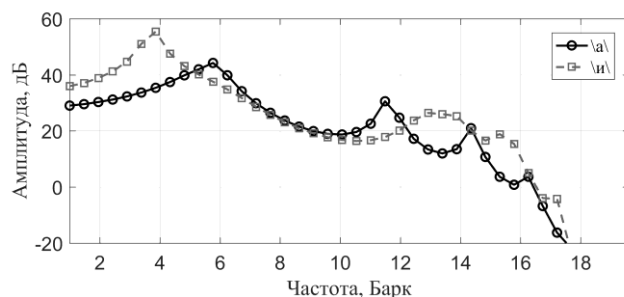


Рис. 4. Пример схожести огибающих звуков /а/ и /и/ у пациента с БАС

Для количественной оценки различия между огибающими звуков /а/ (E_a) и /и/ (E_i) предлагается использовать d_1 -норму расстояния:

$$d_1(E_i, E_a) = \sum_{k=1}^K |E_i \langle \rangle - E_a \langle \rangle| \quad (2)$$

где K – число точек в частотной области барков.

Так как неврологические заболевания влияют также на темп речи, то для их выявления можно анализировать и паузы между словами. Для оценки пауз предлагается анализировать среднюю паузу и сумму между максимальной (P_{\max}) и минимальной паузой (P_{\min}):

$$P_{cp} = \frac{\sum_{z=1}^N P(z)}{N}, \quad P_{mm} = \frac{P_{\max} + P_{\min}}{P_{cp}}. \quad (3)$$

Так как здоровый человек может выдержать паузу при медленном счете, то P_{mm} у здорового будет стремиться к 2, а у больного оно не обязательно будет близко к 2.

Схема анализа речевого сигнала для выявления БАС

Для выявления бульбарных нарушений при БАС предложена схема анализа речевого сигнала, показанная на рис. 5.

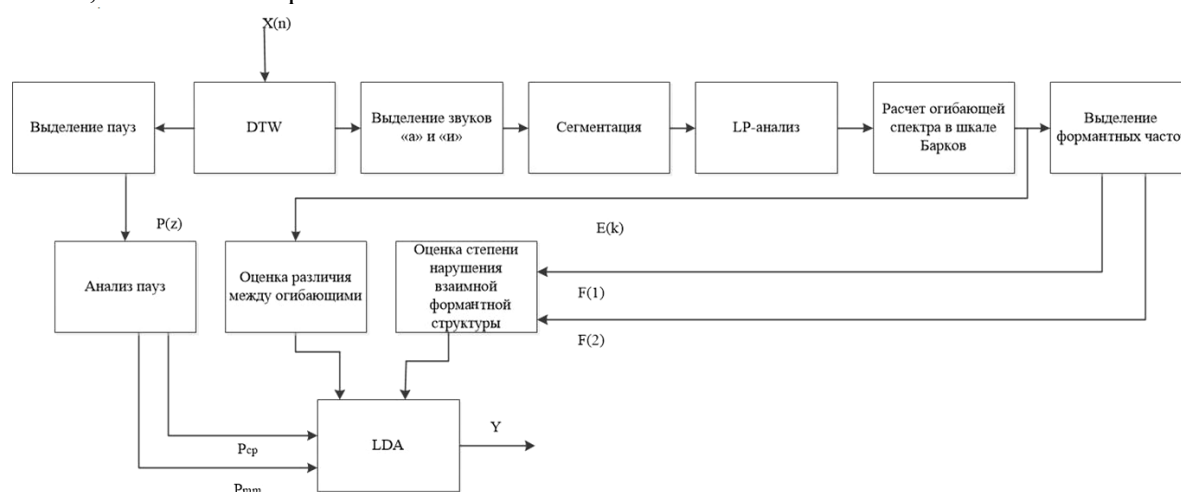


Рис. 5. Схема анализа речевого сигнала для выявления бульбарных нарушений при БАС

На вход системе подается речевой сигнал с записью счета от одного до десяти. Так как для выявления БАС планируется анализировать спектральные огибающие гласных звуков и паузы, то исходной задачей является выделение гласных звуков /а/ и /и/ и пауз из речевого сигнала. Звук /а/ берется из слов «один» (произносится как «а») и/или «два». Звук /и/ из слов «один» и/или «три».

Для решения указанной задачи использовался алгоритм динамической трансформации временной шкалы (англ. DTW – dynamic time warping). Исходными данными для блока DTW является анализируемый входной сигнал и размеченный эталонный сигнал. DTW-алгоритм выполняет «выравнивание» двух временных последовательностей путем трансформации временной шкалы на основании сходства между элементами двух последовательностей. Это означает, что, имея разметку слов и пауз в одной эталонной последовательности, и выполнив выравнивание по ней анализируемой последовательности можно получить разметку слов и пауз в анализируемом сигнале.

Таким образом, входом для DTW-алгоритма служат две последовательности векторов – $S = (s_1, s_2, \dots, s_n)$ и $T = (t_1, t_2, \dots, t_m)$, где каждый вектор является огибающей спектра. В общем случае, длины последовательностей не равны, поскольку каждый диктор произносит тестовую фразу с различной скоростью. На первом этапе выполняется расчет локальных отклонений между элементами двух последовательностей:

$$D_{i,j} = \sum_{k=1}^p |s_{i,k} - t_{j,k}|, \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

На втором этапе вычисляется матрица трансформаций исходя из соотношения:

$$Q_{i,j} = D_{i,j} + \min(Q_{i-1,j}, Q_{i-1,j-1}, Q_{i,j-1}).$$

На третьем этапе строится оптимальный путь деформации, который минимизирует расстояние между S и T и устанавливает соответствие между ними [8] (см. рис. 6).

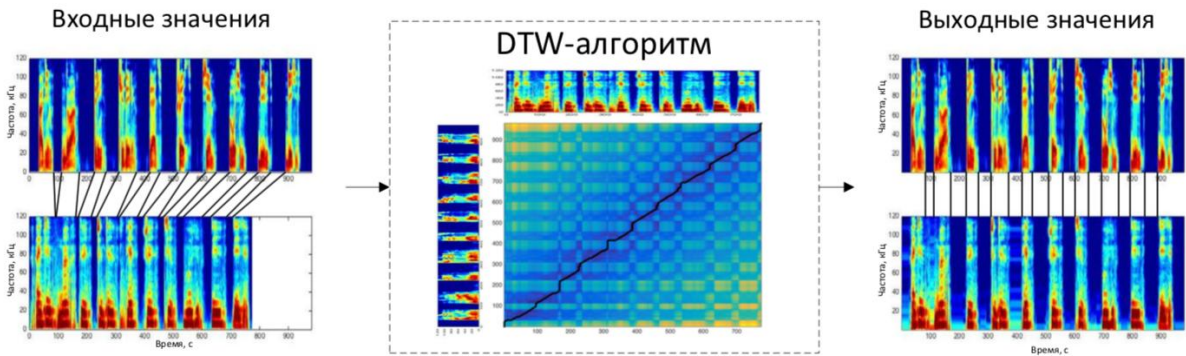


Рис. 6. Принцип работы алгоритма динамической трансформации временной шкалы

В результате работы алгоритма происходит сопоставление двух сигналов, благодаря которому можно определить длину пауз между словами в счете и выделить звуки /а/ и /и/ из входного сигнала. В дальнейшем записи гласных звуков разбивались на кадры длительностью 27 мс с перекрытием в 1/4 от длительности кадра. Для каждого кадра определяются коэффициенты линейного предсказания (англ. LP – Linear Prediction), по которым рассчитывается огибающая спектра. По совокупности всех огибающих рассчитывается усредненное их значение $E(f)$, которое затем переводится в психоакустическую частотную шкалу Барков $E(k)$. Далее из огибающих производится выделение формантных частот, которые затем используются для получения значения $fnt_{err}(F_i, F_a)$. Так же значение $E(k)$, полученное в блоке «Расчет огибающей спектра в шкале Барков», отправляется в блок «Оценка различия между огибающими» для расчета значения $d_1(E_i, E_a)$. Выделенные паузы между словами $P(z)$ обрабатываются в блоке «Анализ пауз», где высчитываются значения P_{cp} и P_{mm} . Полученные значения объединяются в вектор $\mathbf{x} = [d_1(E_i, E_a) \ fnt_{err}(F_i, F_a) \ P_{cp} \ P_{mm}]^T$ и отправляются для классификации в блок LDA (англ. linear discriminant analysis – линейный дискриминантный анализ) [9]. В результате обучения в блоке LDA формируется разделяющая гиперплоскость. Классификация выполняется следующим образом: $q = \mathbf{w}^T \cdot \mathbf{x} - b$, $c = \text{sign}(q)$, где \mathbf{w} – вектор, который описывает разделяющую гиперплоскость, b – смещение; если $c = 1$, то \mathbf{x} относится к классу здоровых, при $c = -1$ \mathbf{x} относится к классу больных.

Анализ результатов

Запись речевой базы проводилась в Республиканском научно-практическом центре неврологии и нейрохирургии (г. Минск, Республика Беларусь). Всего была выполнена запись тестовых речевых образцов у 53 лиц, 23 из которых были здоровыми, а 30 имели БАС.

Первоначально речевые сигналы обрабатывались согласно схеме, представленной на рис. 5. Ниже приводятся статистические характеристики предлагаемых признаков для выявления БАС. На рис. 7, а показано распределение плотности вероятности $d_1(E_i, E_a)$, на рис. 7, б – плотности вероятности $fnt_{err}(F_i, F_a)$.

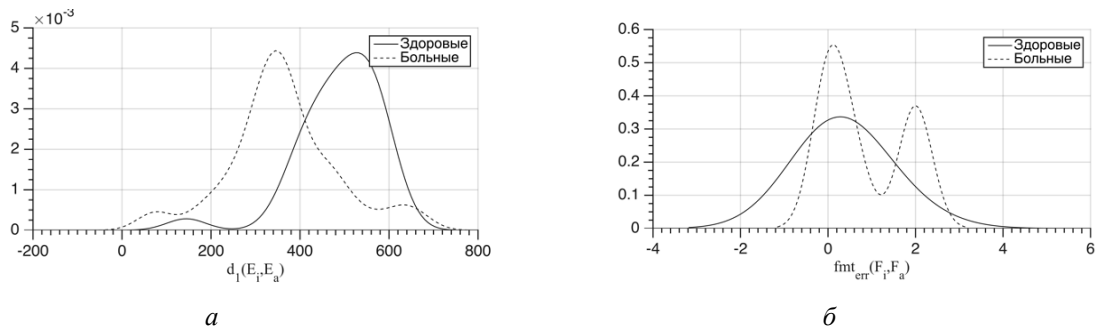


Рис. 7. Распределение плотности вероятности: а – $d_1(E_i, E_a)$; б – $fnt_{err}(F_i, F_a)$

На рис. 8, а и 8, б показаны распределения вероятности P_{cp} и P_{mm} соответственно.

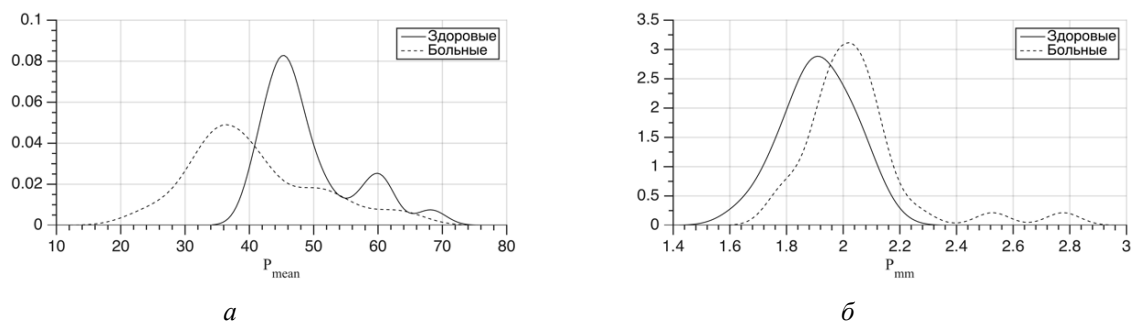


Рис. 8. Распределение вероятности: $a - P_{\text{ср}}$; $b - P_{\text{min}}$

На рис. 9 показано распределение проекций характеристических векторов на разделяющую гиперплоскость в блоке LDA.

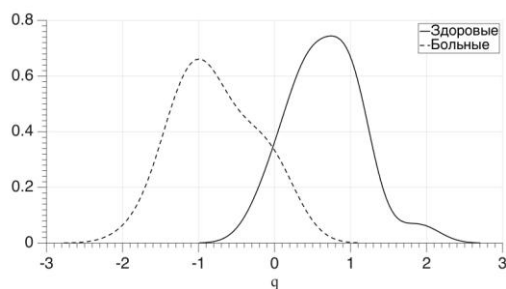


Рис. 9. Результаты работы алгоритма LDA

В результате работы точность классификации составила 92,6 % (9,7 % – не выявлено больных, 4,3 % – ложные срабатывания). Результаты классификации приведены в табл. 1.

Таблица 1. Результат классификации

	Классифицированные больные, %	Классифицированные здоровые, %
Больные	90,3	9,7
Здоровые	4,3	95,7

Заключение

В работе рассмотрена задача анализа речевого сигнала для выявления признаков бульбарных нарушений при БАС. Предложено для решения данной задачи выполнять анализ огибающих гласных звуков /а/ и /и/, извлеченных из беглой речи, а также паузы между словами. Анализ экспериментальных результатов подтвердил, что при наличии БАС огибающие звуков /а/ и /и/ имеют высокую степень схожести между собой, а взаимное расположение их формантных частот может быть нарушено. Выявлено, что при наличии болезни появляется сложность в выдерживании одинаковой паузы при медленном счете. Точность предложенного алгоритма классификации составила 92,6 % (9,7 % – не выявлено больных, 4,3 % – ложные срабатывания).

Список литературы

1. Fractal features for automatic detection of dysarthria / T. Spangler [et al.] // IEEE EMBS International Conference on Biomedical Health Informatics (BHI). Orlando, USA, Feb. 16–19, 2017. P. 437–440.
2. Detection of bulbar ALS using a comprehensive speech assessment battery / Y. Yunusova [et al.] // 8th international workshop Models and analysis of vocal emissions for biomedical applications. Florence, Italy, 16–18 Dec. 2013. P. 217–220.
3. Bulbar and speech motor assessment in ALS: challenges and future directions / J.R. Green [et al.] // Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2013. Vol. 14, № 7–8. P. 494–500.
4. Suitability of dysphonia measurements for telemonitoring of Parkinsons disease / M.A. Little [et al.] // IEEE Transactions on Biomedical Engineering. 2009. Vol. 56, № 4. P. 1015–1022.
5. Меженная М.М., Рушкевич Ю.Н., Борискевич А.А. Частотно-временной анализ речевых сигналов в диагностике бульбарных нарушений // Докл. БГУИР. 2016. № 7 (101). С. 260–264.

6. лов В.Г., Златоустова Л.В. Изменение параметров речи. М.: Радио и связь, 1987. 168 с.
7. Выделение признаков неврологических заболеваний из речевого сигнала / М.И. Вашкевич [и др.] // Труды 20-й междунар. конф. «Цифровая обработка сигналов и ее применение». Москва, 28–30 марта 2018 г. Т. 1. С. 179–184.
8. Киен Т.Ч. Реализация и выбор параметров при использовании алгоритма выравнивания временных масштабов для систем конверсии голоса // Докл. БГУИР. 2008. № 3 (33). С. 96–102.
9. Рылов А.С. Анализ речи в распознающих системах. Минск: Бестринт, 2003. 264 с.

References

1. Fractal features for automatic detection of dysarthria / T. Spangler [et al.] // IEEE EMBS International Conference on Biomedical Health Informatics (BHI). Orlando, USA, Feb. 16–19, 2017. P. 437–440.
2. Detection of bulbar ALS using a comprehensive speech assessment battery / Y. Yunusova [et al.] // 8th international workshop Models and analysis of vocal emissions for biomedical applications. Florence, Italy, 16–18 Dec. 2013. P. 217–220.
3. Bulbar and speech motor assessment in ALS: challenges and future directions / J.R. Green [et al.] // Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2013. Vol. 14, № 7–8. P. 494–500.
4. Suitability of dysphonia measurements for telemonitoring of Parkinsons disease / M.A. Little [et al.] // IEEE Transactions on Biomedical Engineering. 2009. Vol. 56, № 4. P. 1015–1022.
5. Mezhenaja M.M., Rushkevich Ju.N., Boriskevich A.A. Chastotno-vremennoj analiz rechevyh signalov v diagnostike bul'barnyh narushenij // Dokl. BGUIR. 2016. № 7 (101). S. 260–264.
6. Mihailov V.G., Zlatoustova L.V. Izmenenie parametrov rechi. M.: Radio i svjaz', 1987. 168 s. (in Russ.)
7. Vydelenie priznakov nevrologicheskikh zabolevanij iz rechevogo signala / M.I. Vashkevich [i dr.] // Trudy 20-j mezhdunar. konf. «Cifrovaja obrabotka signalov i ee primenenie». Moskva, 28–30 marta 2018 g. T. 1. S. 179–184. (in Russ.)
8. Kien T.Ch. Realizacija i vybor parametrov pri ispol'zovanii algoritma vyravnivanija vremennyh masshtabov dlja sistem konversii golosa // Dokl. BGUIR. 2008. № 3 (33). S. 96–102. (in Russ.)
9. Rylov A.S. Analiz rechi v raspoznajushhijh sistemah. Minsk: Bestrint, 2003. 264 s. (in Russ.)

Сведения об авторах

Гвоздович А.Д., студент Белорусского государственного университета информатики и радиоэлектроники.

Рухкевич Ю.Н., к.м.н., доцент, ведущий научный сотрудник неврологического отдела РНПЦ неврологии и нейрохирургии.

Вашкевич М.И., к.т.н., доцент кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники.

Information about the authors

Gvozdovich A.D., student of the specialty computer engineering of the Belarusian state university of informatics and radioelectronics.

Rushkevich Y.N. PhD, leading researcher of the neurology department of the scientific and technical center for neurology and neurosurgery.

Vashkevich M.I., PhD, associate professor of computer engineering department of Belarusian state university of informatics and radioelectronics.

Адрес для корреспонденции

220013, Республика Беларусь,
г. Минск, ул. П. Бровки, 6,
Белорусский государственный университет
информатики и радиоэлектроники
тел. +375-17-293-88-05;
e-mail: vashkevich@bsuir.by
Вашкевич Максим Иосифович

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovka st., 6,
Belarussian state university
of informatics and radioelectronics
tel. +375-17-293-88-05;
e-mail: vashkevich@bsuir.by
Vashkevich Maksim Iosifovich