

УДК 004.934.2+534.784

АКУСТИЧЕСКИЙ АНАЛИЗ ГОЛОСА ДЛЯ ВЫЯВЛЕНИЯ РЕЧЕВЫХ НАРУШЕНИЙ ПРИ БОКОВОМ АМИОТРОФИЧЕСКОМ СКЛЕРОЗЕ

М.И. ВАШКЕВИЧ¹, А.Д. ГВОЗДОВИЧ¹, Ю.Н. РУШКЕВИЧ², А.А. ПЕТРОВСКИЙ¹¹Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь²Республиканский научно-практический центр неврологии и нейрохирургии, Республика Беларусь

Поступила в редакцию 9 ноября 2018

Аннотация. Рассматривается способ акустического анализа голосового сигнала, содержащего протяжные гласные звуки, для построения системы детектирования речевых нарушений при боковом амиотрофическом склерозе (БАС), являющимся неврологическим заболеванием. Предложен способ сегментации голосового сигнала на периоды основного тона, который используется при расчете параметров джиттер и шиммер. Выполнено сравнение двух систем детектирования речевых нарушений при БАС, в одной из которых исходными данными являлись параметры голоса, полученные предлагаемым способом, а во второй – параметры, полученные в распространенной системе PRAAT. Результаты экспериментов показали, что применение прилагаемого способа анализа значительно улучшает (на 20 %) точность детектирования.

Ключевые слова: акустический анализ сигнала, джиттер, шиммер, боковой амиотрофический склероз.

Abstract. A method of acoustic signal analysis with sustain vowel phonation for detection of amyotrophic lateral sclerosis (ALS) is considered. A method for segmentation of the voice signal into periods of the fundamental tone, which is used for evaluation of the jitter and shimmer parameters, is proposed. A comparison of two ALS detectors was performed. The first detector was trained using voice features extracted by the proposed method, while the second detector was trained using features obtained with PRAAT toolkit. The result showed a significant improvement (by 20 %) in the accuracy of detecting ALS disease using the proposed method.

Keywords: acoustic analysis, jitter, shimmer, amyotrophic lateral sclerosis.

Doklady BGUIR. 2018, Vol. 117, No. 7, pp. 64-68

Acoustic analysis of voice for detection of speech disorder for amyotrophic lateral sclerosis

M.I. Vashkevich, A.D. Gvozdovich, Y.N. Rushkevich, A.A. Petrovsky

Введение

Боковой амиотрофический склероз (БАС) относится к быстро развивающимся, нейродегенеративным, не поддающимся лечению заболеваниям, ведущим к гибели. В ходе заболевания у пациента наблюдается постепенное ослабление в работе мышечной системы, ухудшение работы двигательного аппарата, проблемы с глотанием и речью [1]. Особенностью БАС является трудность его диагностирования. Одним из перспективных подходов к детектированию и отслеживанию динамики заболевания является акустический анализ голоса [2]. Среди различных видов речевых тестов, которые используются для диагностирования неврологических заболеваний, самым распространенным является тест на протяжное произнесение гласного звука («а» или «и»). Полученный в результате теста сигнал анализируется с использованием методов обработки во временной, частотной и кепстральной областях [3]. На основе выделенных из голоса признаков строится система детектирования речевых нарушений при БАС, основанная на методах машинного обучения.

Параметры голосового сигнала

Для голосового сигнала, полученного в процессе теста на протяжное произнесение гласных звуков, важнейшими параметрами являются джиттер и шиммер [4]. Оба эти параметра характеризуют работу голосовых складок.

Джиттер – мера пертурбации (возмущений) частоты основного тона, показывающая произвольные изменения в частоте смежных вибрационных циклов голосовых складок [4].

Шиммер – мера аналогичная джиттеру, только характеризующая пертурбации амплитуд сигнала на смежных циклах колебаний основного тона (рис. 1).

Причина широкого применения параметров джиттер и шиммер в задачах диагностики патологии в голосе заключается в том, что структура вибраций голосовых складок имеет периодический характер для здоровых голосов, в то время как эта периодичность в структуре вибраций значительно нарушена в случае наличия патологических изменений [4]. Таким образом, высокое значение джиттера и шиммера характерны для голосов с патологиями.

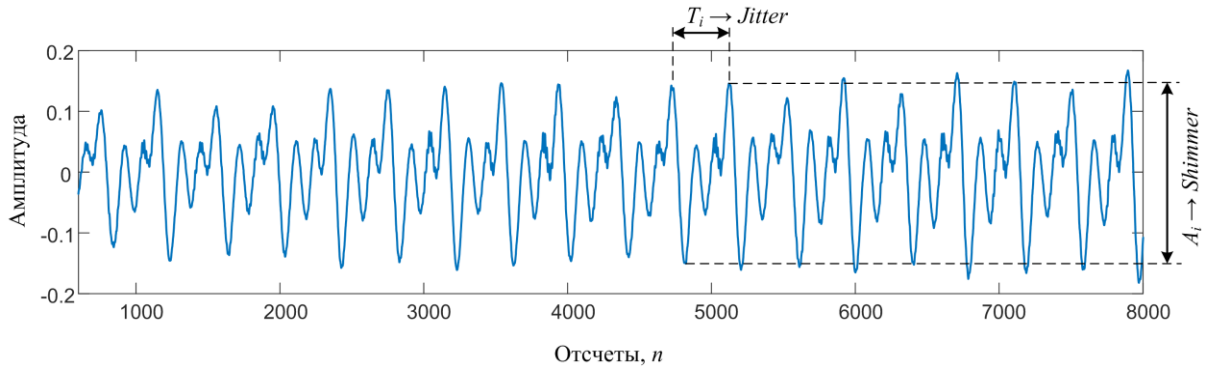


Рис. 1. Представление мер джиттера и шиммера в речевом сигнале

Существует несколько вариантов оценок параметров джиттер и шиммер [4]. Так, локальный джиттер определяется выражением

$$Jitter_{loc} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \Big/ \frac{1}{N} \sum_{i=1}^N T_i, \quad (1)$$

где T_i – длительность i -го периода основного тона, а N – число периодов основного тона исследуемого сигнала. Часто для оценки джиттера используют отклонение текущей длительности периода не от предыдущей, а от локально усредненного значения, которое рассчитывается на окне в 3 или 5 выборок:

$$Jitter_{rap} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right| \Big/ \frac{1}{N} \sum_{i=1}^N T_i, \quad (2)$$

$$Jitter_{ppq5} = \frac{1}{N-4} \sum_{i=3}^{N-2} \left| T_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right| \Big/ \frac{1}{N} \sum_{i=1}^N T_i. \quad (3)$$

Значение параметра шиммер определяется аналогично джиттеру, с тем отличием, что вместо значений длины периодов T_i берутся значения амплитуды колебания на данном периоде A_i . Таким образом, формула для локального шиммера $Shimmer_{loc}$ получается из (1) путем замены T_i на A_i . По аналогии с формулами (2) и (3) существуют варианты, определяющие шиммер, как относительное отклонение амплитуды от локально-усредненного значения на интервале в L выборок:

$$Shimmer_{apqL} = \frac{1}{N-L+1} \sum_{i=1+(L-1)/2}^{N-(L-1)/2} \left| A_i - \left(\frac{1}{L} \sum_{n=i-(L-1)/2}^{i+(L-1)/2} A_n \right) \right| \Big/ \frac{1}{N} \sum_{i=1}^N A_i. \quad (4)$$

Чаще всего L принимает значения 3, 5 и 11.

Сегментация параметров голосового сигнала

Для расчета параметров джиттер и шиммер требуется первоначально сегментировать голосовой сигнал на периоды основного тона. В данной работе предлагается следующий способ сегментации. На первом этапе, для входного сигнала $s(n)$ выполняется оценка контура частоты основного тона оценщиком IRAPT [5] с шагом по времени в 5 мс. На втором этапе производится линейная интерполяция контура таким образом, чтобы получить значение частоты основного тона $F_0(n)$ в каждый дискретный момент времени n . На третьем этапе вычисляется функция полной фазы:

$$\varphi(n) = \sum_{i=1}^n 2\pi F_0(i) / f_s. \quad (5)$$

На последнем этапе $\varphi(n)$ анализируется с целью определения дискретных моментов времени, в которые функция пересекает значения $2\pi k$, где k – целое число. Найденные точки определяют границы между соседними периодами основного тона, зная которые легко можно получить значения T_i для расчета джиттера. Имея T_i , из исходного сигнала $s(n)$ можно получить значения A_i , на основании которых оценивается параметр шиммер.

Экспериментальная часть

Для анализа качества выделенных параметров джиттер и шиммер с целью применения их в системе диагностики речевых нарушений при БАС выполнялся следующий эксперимент: исходная обучающая база записей голосов (норма + патология) обрабатывалась с использованием описанной выше методики, после чего для каждой записи рассчитывался характеристический вектор, состоящий из 7 компонент:

$$\mathbf{x}^T = [Jitter_{loc}, Jitter_{rap}, Jitter_{ppq5}, Shimmer_{loc}, Shimmer_{apq3}, Shimmer_{apq5}, Shimmer_{apq11}]$$

На основе полученных характеристических векторов производилось обучение детектора БАС на основе линейного дискриминантного анализа [6]. Для определения точности системы детектирования использовался следующий подход: в начале обучающее множество перемешивалось случайным образом, после чего выполнялась процедура обучения на основе 75 % данных, остальные 25 % использовались для тестирования. Данная процедура повторялась 1000 раз, после чего для точностных параметров рассчитывалось среднее значение и среднеквадратическое отклонение.

Аналогичный эксперимент выполнялся для системы, в которой характеристические вектора рассчитывались путем анализа сигнала в распространенной системе PRAAT [7].

Запись речевой базы проводилась в Республиканском научно-практическом центре неврологии и нейрохирургии, а также в Белорусском государственном университете информатики и радиоэлектроники. Всего была выполнена запись тестовых речевых образцов у 57 лиц, 29 из которых были здоровыми, а 28 имели БАС.

Анализ результата

Полученные системы детектирования БАС сравнивались по таким параметрам как точность (англ. overall accuracy – ACC), чувствительность (англ. sensitivity – Sens), позитивная предсказательная ценность (англ. positive predictive value – PPV) [4]:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, Sens = \frac{TP}{TP + FN}, PPV = \frac{TP}{TP + FP}, \quad (6)$$

где TP , TN , FP , FN – количество истинно положительных, истинно отрицательных, ложно положительных и ложно отрицательных результатов детектирования соответственно.

Полученные в результате экспериментов оценки производительности систем детектирования приведены в таблице.

Результат детектирования при использовании различных способов оценки параметров джиттер и шиммер

Параметры	Система на основе предлагаемого способа оценки параметров	Система на основе параметров полученных в PRAAT [7]
<i>ACC</i>	69,7±11,3 %	48,2±16,1 %
<i>Sens</i>	68,0±19,2 %	43,7±26,6 %
<i>PPV</i>	71,2±13,3 %	45,7±19,4 %

Таким образом, применение предлагаемого способа позволяет значительно повысить точностные характеристики системы детектирования речевых нарушений при БАС. Однако в абсолютном значении точность детектирования в 70 % нельзя считать удовлетворительной. Такой результат объясняется тем, что в работе использовался ограниченный набор параметров. Добавление других параметров характеристического вектора, описывающих тембральные и шумовые составляющие речи, прогнозируемо улучшит результат детектирования.

Дополнительно в работе были проанализированы статистические характеристики отдельных параметров голоса. В качестве примера на рис. 2 показаны диаграммы размаха («ящичковые диаграммы»), полученные для параметра *Jitter*_{ppq5}.

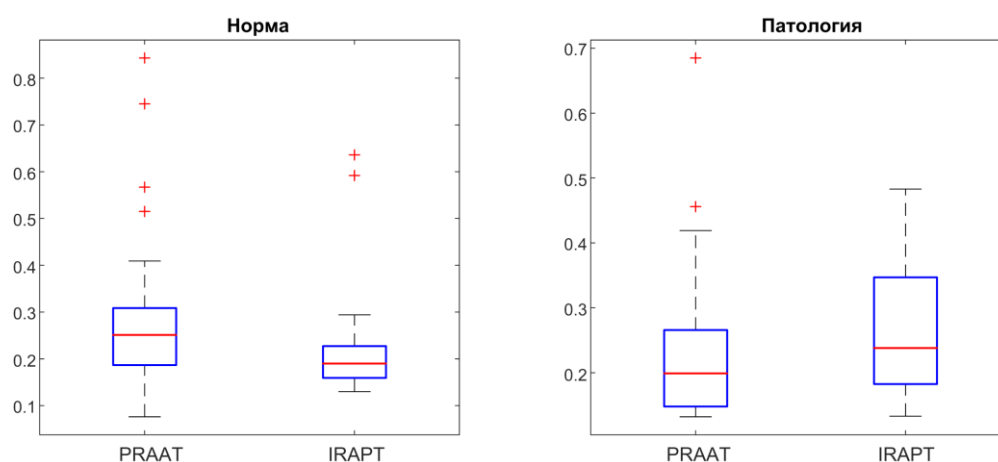


Рис. 2. Диаграммы размаха для *Jitter*_{ppq5}

Анализ показывает, что система PRAAT хуже выделяет параметр джиттер, поскольку медианное значение этого параметра в группе здоровых голосов больше, чем в группе больных. Джиттер, рассчитанный на основе предлагаемого способа, показывает компактное распределение для здоровых голосов и размытое распределение для голосов больных. Медианное значение у здоровых меньше, чем у больных.

Заключение

В работе предложен способ оценивания параметров джиттер и шиммер из сигнала голосового теста, содержащего протяжное произнесение гласных звуков. Показано, что данный способ может быть полезен для построения системы детектирования речевых нарушений при неврологическом заболевании БАС. Проведено сравнение предлагаемого способа оценки параметров джиттер и шиммер с существующей системой PRAAT. В результате сравнения было показано, что использование предлагаемого способа повышает в среднем точность детектирования на 21,5 %. В перспективе данный подход можно использовать для создания приложения для смартфона, позволяющего пациенту отслеживать степень прогрессирования болезни с целью корректировки медикаментозного лечения.

Список литературы

1. Завалишин И.А. Боковой амиотрофический склероз. М.: ГЭОТАР – Медиа, 2009. 272 с.
2. Yunusova Y., Rosenthal J.S., Green J.R. Detection of bulbar ALS using a comprehensive speech assessment battery // Mater. 8th Int. workshop «MAVEBA». Florence, Italy, 16–18 December 2013. P. 217–220.

3. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease / A. Tsanas [et al.] // IEEE Transactions on Biomedical Engineering. 2012. Vol. 59, № 5. P. 1264–1271.
4. Benda A., Jilbab A., Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA // Int. J. of Speech Technology. 2017. Vol. 20. P. 673–683.
5. Azarov E., Vashkevich M., Petrovsky A. Instantaneous pitch estimation based on RAPT framework // Proc. EUSIPCO'12. Bucharest, Romania, 27–31 August 2012. P. 2787–2791.
6. Features extraction for the automatic detection of ALS disease from acoustic speech signals / M.I. Vashkevich [et al.] // Proc. of Int. conf. «SPA'2018». Poznan, Poland, 19–21 September 2018. P. 179–184.
7. Boesma P., Weenink D. Praat, a system for doing phonestica by computer // Glot Int. 2001. Vol. 5. P. 341–345.

References

1. Zavalishin I.A. Bokovoj amiotroficheskiy skleroz. M.: GJeOTAR – Media, 2009. 272 s. (in Russ.)
2. Yunusova Y., Rosenthal J.S., Green J.R. Detection of bulbar ALS using a comprehensive speech assessment battery // Mater. 8th Int. workshop «MAVEBA». Florence, Italy, 16–18 December 2013. P. 217–220.
3. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease / A. Tsanas [et al.] // IEEE Transactions on Biomedical Engineering. 2012. Vol. 59, № 5. P. 1264–1271.
4. Benda A., Jilbab A., Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA // Int. J. of Speech Technology. 2017. Vol. 20. P. 673–683.
5. Azarov E., Vashkevich M., Petrovsky A. Instantaneous pitch estimation based on RAPT framework // Proc. EUSIPCO'12. Bucharest, Romania, 27–31 August 2012. P. 2787–2791.
6. Features extraction for the automatic detection of ALS disease from acoustic speech signals / M.I. Vashkevich [et al.] // Proc. of Int. conf. «SPA'2018». Poznan, Poland, 19–21 September 2018. P. 179–184.
7. Boesma P., Weenink D. Praat, a system for doing phonestica by computer // Glot Int. 2001. Vol. 5. P. 341–345.

Сведения об авторах

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

Гвоздович А.Д. магистрант Белорусского государственного университета информатики и радиоэлектроники.

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

Петровский А.А., д.т.н., профессор кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники.

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

Information about the authors

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

Gvozdovich A.D. master student of 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.

Petrovsky A.A., DSci, professor of computer engineering department of Belarusian state university of informatics and radioelectronics.

Address for correspondence

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