

## ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ В ДИАГНОСТИКЕ БУЛЬБАРНЫХ НАРУШЕНИЙ



**А.Н. Осипов**

Проректор по научной работе БГУИР, кандидат технических наук, доцент



**С.А. Лихачев<sup>1</sup>**

Заведующий неврологическим отделом РНПЦ Неврологии и нейрохирургии, доктор медицинских наук, профессор



**Ю.Н. Рушкевич<sup>1</sup>**

Ведущий научный сотрудник неврологического отдела РНПЦ Неврологии и нейрохирургии, кандидат медицинских наук, доцент



**М.М. Меженная**

Доцент кафедры инженерной психологии и эргономики БГУИР, кандидат технических наук



**А.А. Борискевич**

Профессор кафедры сетей и устройств телекоммуникаций БГУИР, доктор технических наук, доцент



**Т.П. Куль**

Магистрантка кафедры инженерной психологии и эргономики БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

<sup>1</sup> РНПЦ Неврологии и нейрохирургии МЗ, Республика Беларусь

E-mail: mezhennaya@bsuir.by, rushkevich@tut.by

*Abstract.* The method of qualitative and quantitative differential diagnosis of bulbar palsy has been offered on the basis of digital processing of speech signals. The software with the graphic user interface has been developed by authors for implementation of this method which allows to increase the accuracy and speed of diagnosis.

*Введение.* Бульбарные нарушения представляют собой симптомокомплекс, который включает изменение звучности голоса (дисфонию), затруднения при глотании - дисфагию и замедленность речи, нарушение артикуляции - дизартрию, т. е. симптомы, связанные с вовлечением мускулатуры языка, глотки, гортани и мягкого нёба. Причинами бульбарных нарушений является непосредственное поражение ядер языкоглоточного, блуждающего и подъязычного черепных нервов, расположенных в каудальных отделах ствола головного мозга (бульбарный синдром), а также поражение вышеописанных мышц, нервно-мышечного аппарата, патологические процессы в области ствола мозга и задней черепной ямки [1-3].

К ранним проявлениям бульбарного синдрома относится дисфония: голос больных становится слабым, глухим, истощающимся вплоть до полной афонии. Возникает гнусавость. Звуки при этом произносятся невнятно, «смазанно». Гласные звуки становятся трудноотличи-

мыми друг от друга, согласные звуки, разные по способу образования (твёрдые, мягкие, смычные, щелевые) и месту артикуляции (губные, переднеязычные, заднеязычные), произносятся однократно с неопределённым местом артикуляции. Речь оказывается резко замедленной и утомляет больных. Развивающаяся дисфагия из-за невозможности сглатывать слюну и приводит к слюнотечению. При бульбарном параличе наступает атрофия мышц языка и выпадают глоточный и нёбный рефлексы. У тяжелобольных с бульбарным синдромом, как правило, развиваются расстройства ритма дыхания и сердечной деятельности, что нередко приводит к смерти [2,3].

Эффективная дифференциальная диагностика позволяет своевременно оказать медицинскую помощь пациентам с бульбарными нарушениями. К достоверным методам диагностики бульбарного синдрома относятся данные электромиографии и прямого осмотра ротоглотки. Однако в настоящее время имеются сложности постановки диагноза на ранней стадии, сопровождающейся, как отмечалось выше, нарушениями речевой функции. Для проведения своевременной и объективной диагностики бульбарных нарушений предлагается использовать методы цифровой обработки речевых сигналов.

*Методика регистрации и обработки речевых сигналов.* Диагностические исследования бульбарных нарушений были проведены на базе РНПЦ неврологии и нейрохирургии МЗ РБ. В группе пациентов с боковым амиотрофическим склерозом с бульбарным синдромом, а также в контрольной группе здоровых лиц были зарегистрированы тестовые речевые сигналы. Тест представлял собой счет от одного до десяти.

Последующая обработка речевых сигналов выполнялась в среде MatLab с помощью специально разработанного авторами статьи программного обеспечения с графическим интерфейсом. Обработка включала следующие этапы:

- 1 Автоматическое выделение в зарегистрированном сигнале речевых фрагментов.
- 2 Подсчет количества выделенных речевых фрагментов.
- 3 Построение спектрограммы зарегистрированного сигнала.
- 4 Построение кепстрограмм для выделенных речевых фрагментов.
- 5 Определение с помощью кепстральной функции частоты основного тона для каждого речевого фрагмента.
- 6 Расчет средней величины частоты основного тона.
- 7 Расчет коэффициента вариации частоты основного тона.
- 8 Построение гистограммы для массива, представляющего собой результат «склейки» всех выделенных речевых фрагментов.
- 9 Вычисление средней амплитуды выделенных речевых фрагментов.
- 10 Вычисление общего времени всех выделенных речевых фрагментов.
- 11 Вычисление коэффициента асимметрии гистограммы.
- 12 Вычисление коэффициента эксцесса гистограммы.

Далее приведена подробная методика реализации вышеописанных этапов.

Исходный зарегистрированный сигнал характеризовался частотой дискретизации 44,1 кГц, разрядностью 16 бит. Предварительно производилось усреднение зарегистрированного сигнала в окне без перекрытия для снижения исходной частоты дискретизации:

$$A[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i \cdot M + j] \quad (1)$$

где  $A[i]$  – отсчеты, полученные из исходного сигнала  $x$  посредством усреднения;

$i = 0 \dots \frac{N}{M} - 1$  – номер окна;  $N$  – число отсчетов исходного речевого сигнала  $x$ ;

$M$  – длина окна (число усредняемых точек);

$j$  – номер временного отсчета внутри окна.

В результате усреднения при  $M = 5$  частота дискретизации была понижена до 8,82 кГц. Это позволило впоследствии увеличить скорость обработки данных без потери полезной информации в сигнале.

Речь человека содержит паузы между словами. Традиционно для решения задачи разделения речевого сигнала на голосовые и неголосовые участки исходный сигнал разделяется на фрагменты длиной 5-100 мс (с точки зрения динамики речи самые быстрые изменения могут происходить всего за несколько миллисекунд, в то время, как некоторые гласные звуки остаются относительно стабильными в течение 100-200 мс). Для классификации принадлежности фрагмента к сигналу или паузе рассчитывалась кратковременная энергия сигнала в данном фрагменте:

$$E_m = \sum_{j=0}^{L_{fr}-1} A[m \cdot L_{fr} + j]^2 \quad (2)$$

где  $L_{fr}$  – длина фрагмента;

$m = 0 \dots \frac{N}{M \cdot L_{fr}} - 1$  – количество фрагментов;  $j$  – номер временного отсчета усредненного

сигнала внутри фрагмента. В качестве  $L_{fr}$  авторами выбраны 400 отсчетов, что соответствует временной реализации сигнала в 45,4 мс.

На основе экспериментальных исследований речевых сигналов в норме было сформировано условие, при выполнении которого принималось решение о принадлежности  $m$ -ого фрагмента к речи:

$$E_m \geq level_E \cdot \langle E \rangle, \quad (3)$$

где  $\langle E \rangle$  – средняя кратковременная энергия всех фрагментов,  $level_E$  – пороговый уровень кратковременной энергии.

Авторами установлено, что при  $level_E = 0.2$  происходит автоматическое выделение слов и/или отдельных фонем в сигнале.

Далее выполнялось построение спектрограммы сигнала. Для этого речевой сигнал разделялся на временные отрезки, в пределах которых его можно считать стационарным (5-100 мс). Исходный сигнал  $A$  на выбранном отрезке умножался на оконную функцию  $w$  и подвергался быстрому преобразованию Фурье в соответствии с выражением:

$$STFT_A^w [f_k, \tau] = \sum_{i=0}^{L-1} A[i] \cdot w[i] \cdot e^{-\frac{j2\pi ki}{L}}, f_k = \frac{k \cdot f_d}{L}, k = 0 \dots (L-1)/2, \quad (4)$$

где  $L$  – длина окна,  $\tau$  – величина перекрытия окон,  $f_d$  – частота дискретизации.

После данной операции путем возведения в квадрат амплитудной части оконного преобразования Фурье получали спектрограмму мощности для анализируемого окна:

$$Спектрограмма A[f, t] = |STFT_A^w [f_k, \tau]|^2. \quad (5)$$

Далее производилось смещение окна на величину  $\tau$  и процедура повторялась. Подобным образом анализировались все подинтервалы сигнала и строилась результирующая

спектрограмма, представляющая собой двумерную матрицу, строки которой соответствуют временным отсчетам  $t$  от 0 секунд до окончания времени регистрации речевого сигнала, столбцы – частотам  $f$  от 0 до 4,41 кГц, а в ячейках рассчитана амплитуда сигнала [4]. В качестве основных параметров частотно-временной обработки выбраны следующие: окно Хэмминга, размер окна  $L$  в 512 отсчетов, частота дискретизации  $f_d$  в 8,82 кГц, перекрытие окон  $\tau$  в 50%. Указанные характеристики обеспечивают качественное частотно-временное представление речевого сигнала, высокое разрешение по частоте  $\Delta f = 17,2$  Гц и по времени  $\Delta t = 29,0$  мс:

$$\Delta f = \frac{L \cdot (100 - \tau)}{f_d \cdot 100}, \quad (6)$$

$$\Delta t = \frac{f_d}{L}. \quad (7)$$

Для определения частоты основного тона сигнала использовался метод определения кепстра, заключающийся в применении к модулю спектральной плотности исследуемого сигнала обратного преобразования Фурье. При этом в кепстрограмме вокализованного отрезка звука появляется пик на расстоянии основного тона сигнала, что и является основополагающим для последующего вычисления частоты основного тона.

Частота основного тона вычислялась для каждого выделенного вокализованного фрагмента сигнала. По итогам расчетов определяли среднее значение частоты основного тона, а также коэффициент вариации данного параметра – относительную меру разброса значений признака в статистической совокупности. Значения коэффициента вариации менее 10 % свидетельствуют о малом рассеянии, от 10 % до 20 % – о среднем рассеянии, более 20 % – о сильном рассеянии вариант относительно средней арифметической величины.

Для расчета статистических показателей выполнялась «склейка» всех выделенных речевых фрагментов в единый массив. Для полученного массива рассчитывалась средняя амплитуда, а также длительность, соответствующая общей продолжительности речи.

Для визуализации данных на этапе статистической обработки выполнялось построение гистограммы для массива всех речевых фрагментов. Далее для оценки однородности распределения данных в речевых фрагментах рассчитывались показатели асимметрии и эксцесса гистограммы.

Коэффициент асимметрии может быть положительным (для правосторонней асимметрии) и отрицательным (для левосторонней асимметрии). Асимметрия выше 0,5 (независимо от знака) считается значительной, меньше 0,25 – незначительной.

Показатель эксцесса отражает, насколько резкий скачок имеет изучаемое явление. Если показатель эксцесса больше нуля, то распределение островершинное и скачок считается значительным, если коэффициент эксцесса меньше нуля, то распределение считается плосковершинным и скачок считается незначительным.

*Результаты исследований.* Результаты обработки тестовых речевых сигналов в норме, при бульбарном синдроме до лечения и после курса транскраниальной магнитной стимуляции и нейрометаболического лечения представлены на рисунках 1, 2 3 соответственно.

Анализ полученных результатов выявил следующие закономерности.

В группе здоровых лиц (рис. 1) количество распознанных речевых фрагментов составляет 10-12, что соответствует количеству произносимых слов (10) или фонем («че-тыре», «во-семь»). На спектрограмме отчетливо выделяются равностоящие друг от друга речевые фрагменты, частота основного тона, а также формантные частоты. Кепстрограммы для распознанных речевых фрагментов также имеют характерные пики в области основного тона и кратных ему формантных частот. Коэффициент вариации основного тона невысокий (равен 7 на рис.1), что свидетельствует о постоянстве данного параметра во время речи. Гистограмма речи характеризуется симметричностью (коэффициент асимметрии равен 0,58 на

рис.1) с четко выделенным пиковым значением (коэффициент эксцесса равен 12,39 на рис.1).

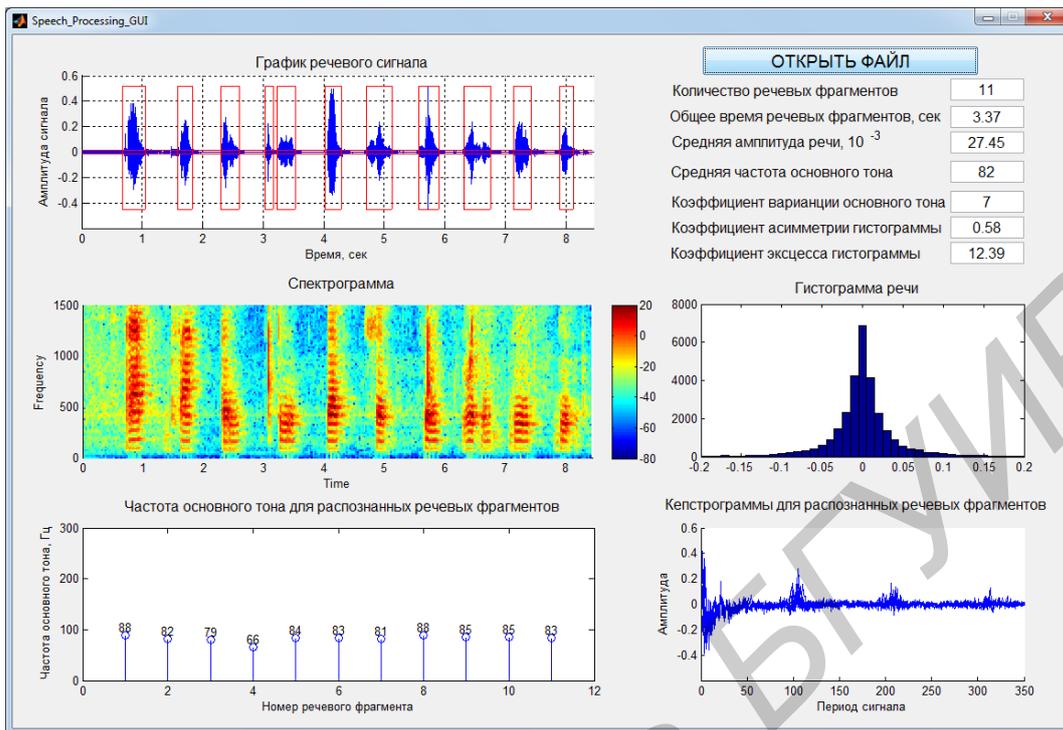


Рис. 1. Результаты обработки речевого сигнала в норме (испытуемый У)

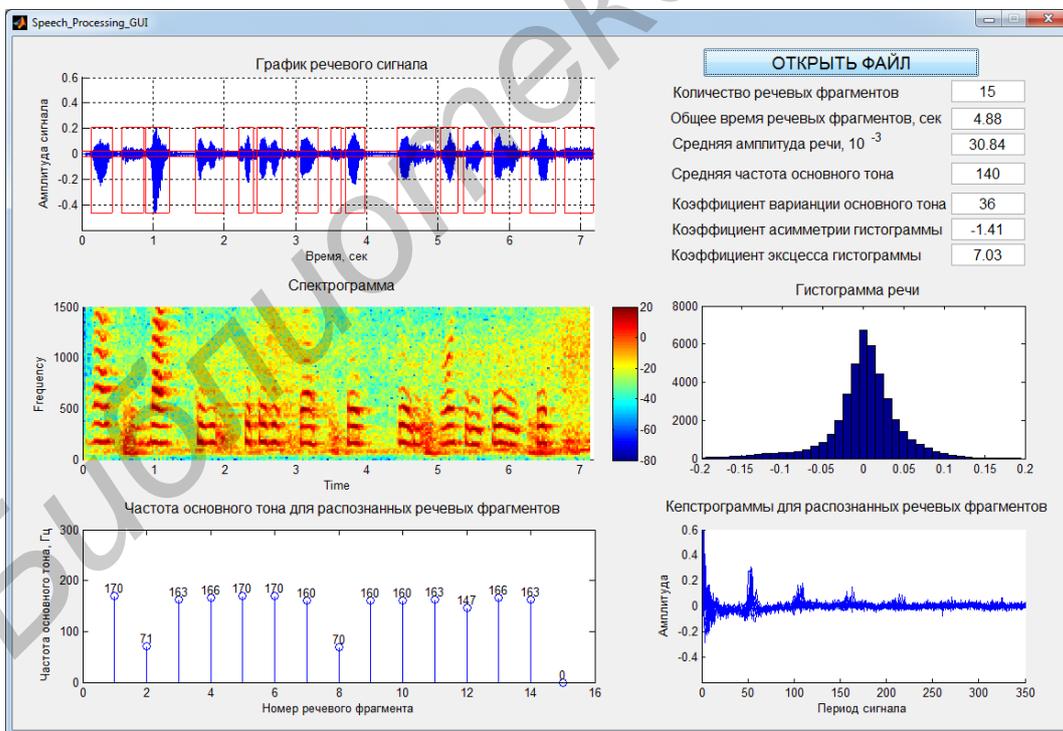


Рис. 2. Результаты обработки речевого сигнала до лечения бульбарного синдрома (пациент К.)

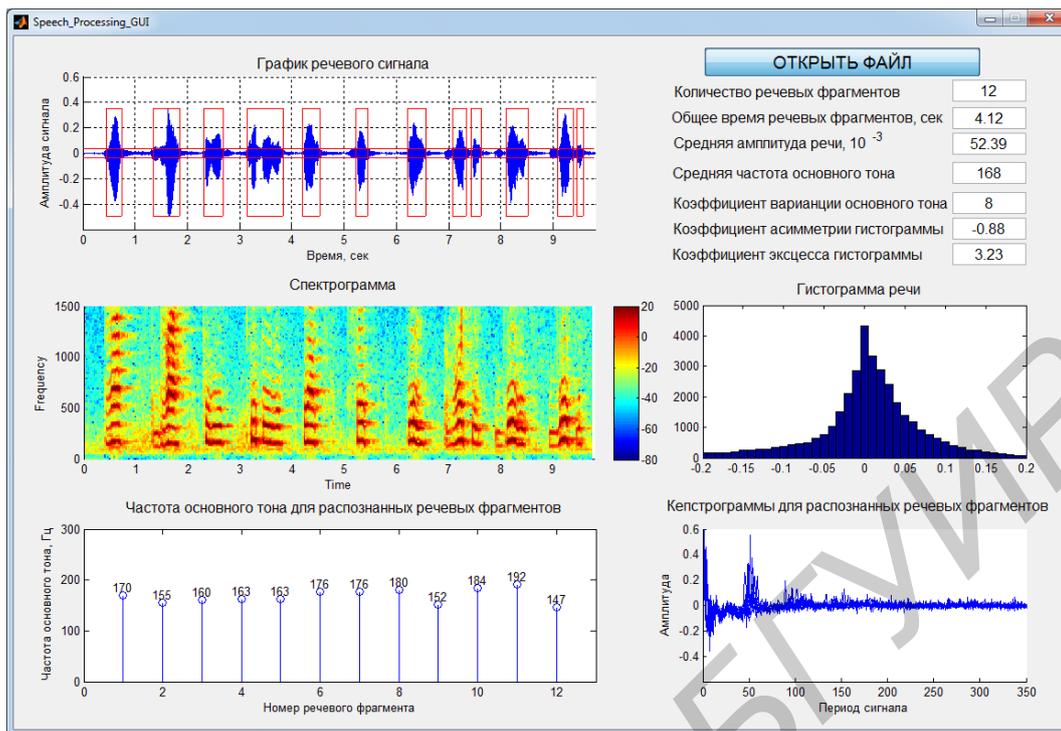


Рис. 3. Результаты обработки речевого сигнала после лечения бульбарного синдрома (пациент К.)

В группе пациентов с бульбарным синдромом до лечения (рис. 2) количество распознанных речевых фрагментов, как правило, превышает количество произносимых слов (10). Это объясняется характерной для данной патологии невнятностью (гнусавостью) речи. Сигнал на спектрограмме не имеет четкой временной структуры в виде равноотстоящих речевых актов, как у здоровых лиц. Значение частоты основного тона для ряда речевых фрагментов не входит в стандартные диапазоны 70 – 450 Гц или не определяется вовсе. Показателен коэффициент вариации частоты основного тона (равен 36 на рис.2), свидетельствующий о сильной степени рассеяния данного параметра относительно среднеарифметического значения. Гистограмма характеризуется левосторонней асимметрией (коэффициент асимметрии равен -1,41 на рис.2), менее выраженным пиковым значением по сравнению с нормой (коэффициент эксцесса равен 7,03 на рис.2).

В группе пациентов с бульбарным синдромом после лечения (рис. 3) количество распознанных речевых фрагментов в целом соответствовало количеству произносимых слов (10) или фонем («че-тыре», «во-семь»). Возросла амплитуда сигнала (см. рис.2,3). На спектрограмме речевые фрагменты приобрели четкие очертания; выделяются паузы, как в группе здоровых лиц. Характерные для основного тона и формантных частот пики демонстрируют кепстрограммы. Вариабельность частоты основного тона вернулась к показателям в норме (равна 8 на рис.3). Степень асимметричности гистограммы снижается (коэффициент асимметрии равен -0,88 на рис.3).

**Заключение.** Предложен метод качественной и количественной диагностики бульбарных нарушений на основе цифровой обработки речевых сигналов. Для реализации данного метода авторами разработано программное обеспечение с графическим интерфейсом, которое позволяет повысить точность и скорость постановки диагноза.

Установлено, что речевые сигналы пациентов с бульбарным синдромом содержат число вокализованных фрагментов, превышающее количество произносимых слов (из-за невнятности речи); сигнал на спектрограмме не имеет четкой временной структуры в виде равноотстоящих речевых актов, как у здоровых лиц; значение частоты основного тона для ряда речевых

фрагментов не входит в стандартные диапазоны 70 – 450 Гц или не определяется вовсе; наблюдается высокая степень рассеяния значений частоты основного тона; гистограмма речи асимметрична.

Метод цифровой обработки речевых сигналов также целесообразно использовать для контроля эффективности лечения неврологических патологий, сопровождающихся нарушениями речевой функции. Установлено, что речевые сигналы пациентов с бульбарным синдромом после лечения характеризуются увеличенной амплитудой, снижением варибельности частоты основного тона, а также степени асимметричности гистограммы речи. На спектрограмме речевые фрагменты приобретают четкие очертания; выделяются паузы, характерные для основного тона и формантных частот пики демонстрируют кепстрограммы.

#### *Литература*

- [1]. Завалишин, И.А. Боковой амиотрофический склероз / И.А. Завалишин – М.: ГЭОТАР-Медиа, 2009: 272.
- [2]. Andersen, P. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)-revised report of an EFNS task force / P. Andersen, et al. Eur J Neurol. 2012;19(3):360–75.
- [3]. Miller, R. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology / R. Miller, et al. Neurology. 2009;73(15):1218–1226. doi: 10.1212/WNL.0b013e3181bc0141.
- [4]. Райгайян, Р.М. Анализ биомедицинских сигналов. Практический подход / Р.М. Райгайян – М.: ФИЗМАТЛИТ, 2007. – 440 с.