

## ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ФИЗИОЛОГИЧЕСКИХ СИГНАЛОВ

Сенкевич Ю.И.

## Введение

Функциональная диагностика без преувеличения может быть названа главным направлением медицинской практики и научного поиска неинвазивных методов анализа и выявления сердечнососудистых и психоневрологических заболеваний. По данным аналитических обзоров Всемирной организации здравоохранения на долю этих нозологических форм приходится от 40 до 45% смертности [1,2]. Методы функциональной диагностики основаны на определении функционального состояния (ФС) человека по данным измерений параметров физиологических сигналов. Цифровая обработка и выделение информационных признаков из физиологических сигналов определяют успех диагностики, терапии и профилактики заболеваний. Поэтому поиск эффективных и корректных способов обработки физиологических сигналов в интересах формирования пространства признаков ФС есть ключевой процесс разработки автоматизированной системы диагностики заболеваний. Известные алгоритмы цифровой обработки сигналов (ЦОС) функционально не ориентированы на сигналы естественных систем, представляющих некоторые закодированные сообщения в символах разнообразных языковых форм.

**Задачей исследований** является разработка инструмента (метода) распознавания ФС биологической системы через параметры физиологических сигналов, представляемых в форме текстовых сообщений.

Среди наиболее применяемых инструментов ЦОС в медицинской диагностике выделяются корреляционный, регрессионный и спектральный (в том числе вейвлет) анализы. Перечисленные решения оптимизированы для обработки стационарных или квазистационарных сигналов, порождаемых искусственными системами. Сигналы, получаемые от биологических объектов, являются сугубо нестационарными. Поэтому используемые алгоритмы и методы ЦОС дают лишь оценки статистических показателей на коротких интервалах наблюдения, которые сложно интерпретируются в диагностике. В некоторых случаях указанные методы принципиально неприменимы, поскольку в силу накладываемых ограничений математического аппарата обработки, порождают фантомные значения [3], что, в конечном счете, приводит к ложным диагностическим заключениям.

Биологические системы, объединяющие множество нелинейных взаимодействующих процессов, наилучшим образом описываются моделями нелинейной динамики [1,2]. В контексте синергетического подхода к живой природе наиболее общая модель функционального состояния биологического объекта может быть представлена в виде нелинейной динамической системы с  $N$  возможными состояниями потенциальной поверхности

*Разработаны теоретические положения лингвистического анализа физиологических сигналов биологических систем. Предложен метод решения задачи распознавания функционального состояния биосистемы с использованием алгоритмов лингвистического анализа физиологических сигналов.*

(поля аттракторов)  $D$ -мерного пространства, где  $N$  и  $D$  - счетные множества. Устойчивое состояние биосистемы не означает стационарность генерируемых ею физиологических сигналов в статистическом смысле. Динамика биосистемы в устойчивом состоянии активно проявляется разнообразием траекторий, порождаемых внутренними процессами, но в рамках границ аттрактора соответствующего ФС, в который биосистема была затянута. С позиции синергетики ограничения, накладываемые процессом затягивания системы в аттрактор, обеспечивают ее способность к генерации совокупности колебаний сигнала с конечным разнообразием форм. Следовательно, операция распознавания ФС в этом случае может быть сведена:

1. К выделению параметров траекторий фазового пространства устойчивых состояний, в которые затягивается биосистема.

2. К контролю приращений выделенных параметров для фиксации выхода биосистемы из устойчивого состояния.

## Модель описания физиологических сигналов

Будем рассматривать физиологический сигнал с позиции теории информации как кодовую последовательность, которую можно интерпретировать как символическое сообщение, записанное на некотором условном языке, что дает возможность проводить манипуляции с кодом и анализировать сообщение подобно грамматическому разбору обычного текста. Выбор позиции требует преобразовать физиологический сигнал в кодовую последовательность без потери переносимой сигналом информации, и выделить значимую информацию из полученного кода.

Выберем из сигнала некоторую ограниченную совокупность его экстремумов, которые позволяют схематично описать форму сигнала. Если в ходе анализа сообщения рассматривать только значения амплитуд в экстремумах и значения интервалов между ними, то можно утверждать, что сохраняется наиболее весомая информация об амплитудно-фазовых характеристиках исходного сигнала. Подтверждением этому утверждению является существование эффективных

алгоритмов восстановления сигналов на основании выбранных измерений, которые известны как адаптивные методы аппроксимации [6,7,8]. В результате для дальнейшей обработки будет выделена совокупность параметров, состоящая из пар чисел (4):

$$x_1, \tau_1 \quad x_2, \tau_2 \quad \dots \quad x_N, \tau_N \quad (1)$$

При составлении правил кодирования последовательности (1) ставится задача сохранить информацию о форме сигнала. Для этого получим данные о положении каждого экстремума относительно ряда соседних ему экстремумов, рассчитав их отдельно для амплитудных значений и временных интервалов между  $M$  соседними точками:

$$r_{i,i+m} = \begin{cases} 1, & x_i > x_{i+m} \\ 0, & x_i \leq x_{i+m} \end{cases} \quad \omega_{i,i+m} = \begin{cases} 1, & \tau_i > \tau_{i+m} \\ 0, & \tau_i \leq \tau_{i+m} \end{cases} \\ m = 1 \dots M (M \leq N - i) \quad (2)$$

Упорядочим результаты в виде бинарных матриц вида (3) – для сравнения амплитудных значений экстремумов

$$\mathbf{R} = \begin{pmatrix} r_{i,i} & r_{i,i+1} & \dots & r_{i,i+(M-2)} & r_{i,i+(M-1)} \\ r_{i+1,i} & r_{i+1,i+1} & & & r_{i+1,i+(M-1)} \\ \vdots & & \ddots & & \vdots \\ r_{i+(M-2),i} & & r_{i+(M-2),i+(M-2)} & r_{i+(M-2),i+(M-1)} \\ r_{i+(M-1),i} & r_{i+(M-1),i+1} & \dots & r_{i+(M-1),i+(M-2)} & r_{i+(M-1),i+(M-1)} \end{pmatrix} \quad (3)$$

и вида (4) – для сравнения значений интервалов между экстремумами

$$\mathbf{W} = \begin{pmatrix} \omega_{i,i} & \omega_{i,i+1} & \dots & \omega_{i,i+(M-2)} & \omega_{i,i+(M-1)} \\ \omega_{i+1,i} & \omega_{i+1,i+1} & & & \omega_{i+1,i+(M-1)} \\ \vdots & & \ddots & & \vdots \\ \omega_{i+(M-2),i} & & \omega_{i+(M-2),i+(M-2)} & \omega_{i+(M-2),i+(M-1)} \\ \omega_{i+(M-1),i} & \omega_{i+(M-1),i+1} & \dots & \omega_{i+(M-1),i+(M-2)} & \omega_{i+(M-1),i+(M-1)} \end{pmatrix} \quad (4)$$

Матрицы (3) и (4) симметричные. Поэтому без потери информации о характеристиках сигнала можно выполнить преобразование (3) в нижнетреугольную матрицу  $\mathbf{R}'_i$ , а (4) – в верхнетреугольную матрицу  $\mathbf{W}'_i$ . Исключая избыточность полученных матриц, составим единую матрицу, объединяющую в себе отношения амплитуд и интервалов. Для этого выполним операцию алгебраического сложения матриц  $\mathbf{R}'_i$  и  $\mathbf{W}'_i$ . Учитывая, что, в соответствии с правилом (2), показатели на главной диагонали матриц (3) и (4) равны нулю, получим

$$\mathbf{D} = \mathbf{R}' + \mathbf{W}' = \begin{pmatrix} 0 & \omega_{i,i+1} & \dots & \omega_{i,i+(M-2)} & \omega_{i,i+(M-1)} \\ r_{i+1,i} & 0 & & & \omega_{i+1,i+(M-1)} \\ \vdots & & \ddots & & \vdots \\ r_{i+(M-2),i} & & & 0 & \omega_{i+(M-2),i+(M-2)} \\ r_{i+(M-2),i} & r_{i+(M-2),i+1} & \dots & r_{i+(M-2),i+(M-2)} & 0 \end{pmatrix} \quad (5)$$

Матрица (5) представляет определенный образ выбранного ( $i$ -го) экстремума в сигнале, характеризующий его амплитудное и временное положение по отношению к соседним экстремумам на глубину  $M$  вправо. Как следствие применения правил отношения (2), преобразование фрагмента сигнала из  $M$  экстремумов в матри-

цу (5) обладает свойством инвариантности к операциям амплитудного и временного транспонирования исходного сигнала. Полученное свойство вытекает из основного свойства неравенств: если  $a > b$ , то  $a + c > b + c$  – при любом  $c$ , – для операции смещения сигнала во времени, и если  $a > b$  и  $c > 0$ , то  $ac > bc$  – для операции усиления сигнала. Поэтому каждой полученной матрице (4) можно сопоставить графический инвариант формы сигнала. Можно утверждать, что матрица (5), как образ, несет в себе некоторую конструктивную нагрузку, соответствующую амплитудному и фазовому значениям для каждой особой точки, которую она описывает. На самом деле описывается не только точка экстремума, но и ее сопровождение в виде отношений амплитуд и интервалов  $M$  ближайших экстремумов.

Сформулируем правило для поиска подобных матриц образов экстремумов (5) в сигнале. Если матрицы образы (5), составленные для различных экстремумов, встречаются в сигнальном сообщении более чем  $N \geq 2$  раз, то такая матрица образ описывает устойчивую форму сообщения, где  $N$  – порог устойчивости. Правило описывает известный в теории динамических систем факт, что в стационарных состояниях система описывает подобные по форме трассы в фазовом пространстве (свойство самоподобия), порождаемые подобными фрагментами сигнала. В соответствии с показанным свойством инвариантности матриц-образов, таким временным фрагментам будет соответствовать одна и та же устойчивая форма сигнала.

Применяя созданное правило к физиологическому сигналу, вырабатывается некоторый код. В принятой языковой парадигме такой код будет представлять определенный набор *символов*. Преобразуя сигнал с использованием созданного правила, в конечном счете, получим текстовое сообщение.

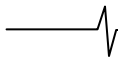
### Лингвистическая модель физиологических сигналов

Построим лингвистическую модель ФС биосистемы. Для этого разработаем ряд критериев, которые позволили бы при анализе текстовых сообщений выявлять структуру языка на разных уровнях объединения символов. В лингвистическом анализе формализм языка известен как грамматика. Придерживаясь единой терминологии, обнаруживаемые признаки состояния биосистемы будем называть соответствующими лингвистическими понятиями.

Назовем *алфавитом*  $A$  сообщения множество выделенных из сигнала символов. *Размером алфавита*  $D = |A|$  назовем общее число обнаруженных символов алфавита в сообщении. Частота появления того или иного символа в сообщении различна и указывает на события, которые наиболее часто повторяются в генерирующей сигнал биосистеме. Для оценки частоты появления символа в сообщении удобно использовать *вероятность появления символа*  $a_i$  алфавита  $A$ , равную отношению числа появлений символа  $m_i$  в сообщении к общему числу символов в этом сообщении  $D$ :

$$P(a_i) = \frac{m_i}{D} \quad (6)$$

Параметр (6) с точки зрения теории информации является еще и оценкой информационной нагрузки, приходящейся на каждый символ алфавита.



Если принять, что в некотором устойчивом состоянии набор событий характеризующих состояние системы не изменяется на протяжении длительного времени, то логично предположить, что состав алфавита  $A$  воспринимаемого сообщения, отражающий этот набор событий, и его размерность  $N$  также не должны изменяться. Меняться может только порядок появления каждого из символов сообщения или порядок последовательности появления символов, косвенно отражающих некоторые события в устойчивом состоянии биосистемы.

Выдвигается гипотеза о том, что в рамках каждого устойчивого состояния система способна генерировать только фиксированный набор символов алфавита определенной размерности. Если произойдет изменение состояния системы, это должно повлечь за собой изменение набора событий, что, в свою очередь, должно вызвать качественные и количественные изменения в составе алфавита, которые можно обнаружить, применяя аппарат теории распознавания образов.

Для оценки связи двух алфавитов сообщения используем коэффициент перекрытия алфавитов -  $k$ , представляющий собой отношение числа символов, которые совпадают для сравниваемых алфавитов, к суммарному числу символов в этих алфавитах. Пусть  $A$  и  $B$  – алфавиты двух сообщений с длинами  $N$  и  $M$  соответственно. Тогда из теории множеств мощность пересечения  $A$  и  $B$  будет представлять количество совместных элементов множества, полученного от их пересечения, а коэффициент перекрытия алфавитов  $A$  и  $B$  выразится как

$$k = 2 \cdot |A \cap B| / (N + M). \quad (7)$$

Использование (7) позволяет построить алгоритм идентификации ФС. Так, в соответствии с выдвинутой гипотезой, одинаковые состояния биосистемы будут иметь одинаковые алфавиты ( $k=1$ ). Чем больше состояния отличаются друг от друга, тем меньшее количество символов в них будет совпадать ( $k=0$ ).

Следующий этап анализа текстового сообщения направлен на поиск связей совокупностей символов алфавита, то есть слов. Слово – это устойчивая последовательность символов алфавита. Для поиска слов  $w$  применим следующий алгоритм, включающий перебор возможных сочетаний символов выделенного алфавита по правилу:

Если  $a_i$  и  $a_{i+1}$  – пара рядом стоящих символов сообщения встречается в этом же сообщении более чем  $L \geq 2$  раз, то эта пара представляет устойчивую последовательность символов  $w$ . Назовем значение  $L$  статистическим порогом существования связи символов. В свою очередь, возможно, что сочетание  $w$  и  $a_{i+2}$ , также окажется устойчивым. Проверка будет продолжаться до тех пор, пока статистика связи не станет ниже порога  $L$ . В последнем случае принимается решение, что последняя из обнаруженных связанная цепочка символов  $w$  есть более высокая грамматическая форма - слово.

Применяя правило поиска слов ко всему сообщению, выделяется весь набор слов  $W = \{w\}$ , составляющий словарь сообщения объемом  $V = |W|$ .

Таким образом, удалось разработать набор параметров, описывающих устойчивое состояние био-

системы, - это состав алфавита, его размерность, словарь и его размерность, выделенные из физиологического сигнала.

Основным свойством сложной динамической системы является постоянная смена ее состояний. Как известно из теории нелинейных динамических систем, под действием внешних сил и потоков энергии и массы система может выходить из устойчивых состояний, менять их, переходя от одного состояния к другому с определенной вероятностью, а в некоторых случаях переходить в состояние хаоса. Контроль динамики системы, выход системы из устойчивого состояния в синергетике осуществляется расчетом производства информационной энтропии [1]. В контексте лингвистического подхода свяжем динамику поведения биосистемы с вероятностью (6) генерации ею символов алфавита. Тогда энтропия производства символа алфавита будет выглядеть следующим образом

$$S^{(j)} = - \sum_{j=0}^N (P_j) \cdot \ln(P_j) \quad (8)$$

где

$$P_j = \frac{z_j}{\sum_{i=0}^N z_i} \quad (9)$$

есть вероятность появления  $j$ -го символа в сообщении. Критерий (8) является каноническим, но на практике при анализе текстовых сообщений биосистем удобнее пользоваться более простым в вычислении и легко интерпретируемым критерием скорости производства новых символов ( $v$ ), которую определим как количество новых символов ( $\Delta N$ ) обнаруживаемых (генерируемых системой) в единицу времени ( $\Delta t$ ):

$$v = \frac{\Delta N}{\Delta t} \quad (10)$$

Вычисление (10) дает возможность построить относительно простой алгоритм слежения за скоростью производства символов. Как отмечалось ранее, постоянство размерности алфавита указывает на устойчивое состояние системы. В таком состоянии системы критерий (10) будет поддерживаться в рамках узких фиксированных границ. Выход за рамки границ с любым знаком указывает на процесс потери устойчивости системы. Контроль изменения легко осуществлять по второй производной по времени от показателя (10), который явно указывает на смену состояния системы, поскольку меняется количественный (10) и качественный (11) показатель, который назовем тенденцией производства символов:

$$\varphi = \begin{cases} 1, & \frac{\Delta v}{\Delta t} > 0 \\ 0, & \frac{\Delta v}{\Delta t} = 0 \\ -1, & \frac{\Delta v}{\Delta t} < 0 \end{cases} \quad (11)$$

Критерий (10) подсчитывает изменение интенсивности генерации символов, изменяя знак (11) и указывая на смену состояния системы.

**Заключение**

Собранные в совокупности показатели и критерии (2=11) составляет лингвистическую модель ФС биологической системы, позволяющую с позиций синергетики описывать ее поведение, что, в конечном счете, для специалистов может служить количественной и качественной диагностической информацией при медицинских обследованиях с использованием неинвазивных методов наблюдения за пациентами с использованием информации, скрытой в физиологических сигналах.

Представленные положения лингвистического анализа физиологических сигналов, были получены и обобщены в Санкт-Петербургском институте информатики и автоматизации РАН [1]. В этой работе изложены результаты моделирования в среде Matlab и практического использования лингвистического анализа для различных медицинских и биологических приложений.

К настоящему времени для практической апробации разработанного аппарата лингвистического анализа физиологических сигналов создано несколько компьютерных программ. Одна из них [1] с 2001 года практически используется для мониторинга процесса адаптации полярников Российских антарктических экспедиций.

Лингвистический подход к анализу сигналов открывает перспективу широкого использования известных методов обработки и анализа текстовой информации для обработки и анализа физиологических сигналов.

**Литература**

1. <http://www.euro.who.int/document/ehr/e76907rh.pdf>
2. <http://www.euro.who.int/Document/RC51/rdoc7.pdf>
3. Финк Л.М. Сигналы, помехи, ошибки... Заметки о некоторых неожиданностях, парадоксах и заблуждениях в теории связи. – 2-е изд., перераб., и доп. –М: Радио и связь, 1984. – 256 с.
4. А.Баблюянец. Молекулы, динамика и жизнь. Введение в самоорганизацию материи. М: Мир, 1990.
5. Л.Гласс, М.Мэки. От часов к хаосу. Ритмы жизни. М: Мир, 1991. 145с.
6. Макс Ж.. Методы и техника обработки сигналов при физических измерениях. Пер.с франц./ том 1 и 2. М: Мир, 1983.
7. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения. Пер.с англ.- М.: Мир, 1990.
8. Эйхофф П. Основы идентификации систем управления. Оценка параметров и состояния. Пер.с англ. -М.:Мир 1975
9. Г.Николс, И.Пригожин. Познание сложного. Серия: Синергетика от прошлого к будущему. –М: УРСС, 2003. с.186-191
10. Сенкевич Ю.И. Разработка математической модели и алгоритмов определения функционального состояния биологических объектов. Диссертация на соискание ученой степени кандидата технических наук на тему –СПб: Библиотечный фонд СПИИРАН 1998. -147с.
11. Сенкевич Ю.И.Программа энтропийно - синтаксического анализа электрофизиологических сигналов (ESAES ver. 3.0). – М.: ВНИИЦ, 2007. - №50200700165

**Краткое сообщение**

К статье М.Парфенюка и А.А.Петровского «Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение», опубликованной в №1, 2008г., с.22-36:

По технической ошибке список литературы был урезан. Ниже приводятся недостающие ссылки в порядке их упоминания в статье:

21. M. Parfieniuk, A. Petrovsky, “Implementation perspectives of quaternionic component for paraunitary filter banks”, in *Proc. of the International Workshop on Spectral methods and multirate signal processing (SMMSP'04)*, pp. 151-158, Vienna, Austria, Sep. 2004.

22. M. Parfieniuk, A. Petrovsky, “Quaternionic lattice structures for four-channel paraunitary filter banks”, *EURASIP Journal on Advances in Signal Processing*, vol 2007, Article ID 37481, 12pages, 2007.

23. Y. –J. Chen, S. Oraintara, K.S.Amaratunga, “Dyadic-based factorizations for regular paraunitary filterbanks and M-band orthogonal wavelets with structural vanishing moments”, *IEEE Trans. on Signal processing*, vol. 53, no. 1, pp. 193-207, 2005.

24. K.K. Parhi, *VLSI digital processing systems: design and implementation*, Jonh Wiley & Sons, Inc., NY, USA, 1999,-784p.

25. А.Л. Вереник, М. Парфенюк, А.А. Петровский, «FPGA-ориентированный поточный процессор параунитарного банка фильтров на распределенной арифметике», *Материалы 6-ой Междунар. конф. «Автоматизация проектирования дискретных систем (Computer-Aided Design of Discrete Devices – CAD DD'07)»*, 14-15 ноября 2007, г. Минск, т.1. – Минск: ОИПИ НАН Беларуси, с. 100-107, 2007.