

АНАЛИЗ И КЛАССИФИКАЦИЯ ЭХОМЕТРИЧЕСКОГО СИГНАЛА В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ ЖИДКОСТИ В НЕФТЕДОБЫВАЮЩИХ СКВАЖИНАХ

Заикин А.А., к.ф.-м.н., научный сотрудник НИЛ изучения состояния и эволюции подземных резервуаров, Институт геологии и нефтегазовых технологий, Казанский Федеральный Университет, e-mail: AAZaikin@kpfu.ru
Миннуллина Р.А., лаборант НИЛ изучения состояния и эволюции подземных резервуаров Институт геологии и нефтегазовых технологий, Казанский Федеральный Университет, e-mail: raminnullina@gmail.com

ANALYSIS AND CLASSIFICATION OF AN ECHOMETRIC SIGNAL FOR FLUID LEVEL DETECTION IN OIL-PRODUCING WELLS

Zaikin A.A., Minnullina R.A.

Echometry is one of the most widely used methods of fluid level detection in oil-producing wells. On the observed signal (echogram), position of the acoustic wave reflection from the surface is displayed as diminishing peaks with the same distance between them. Peak position detection depends on the quality of the echogram because external sounds and foreign objects also affect it. This study proposes a method for the signal modeling through the solution of the quadratic programming problem (for denoising) and a peak detection algorithm, based on the objective function maximization, which describes the total impact of peaks. Obtained results are used for feature extraction to distinguish the normal echograms from the defective echograms. According to the results, proposed methods accurately determine positions of peaks and classifier correctly selects undamaged echograms.

Key words: noise suppression, peak detection, classification.

Ключевые слова: шумоподавление, поиск пиков, классификация.

Введение

Одним из широко распространенных методов для определения забойного давления по уровню жидкости в скважине является эхометрия, основанная на измерении времени прохождения звуковой волны в межтрубном пространстве скважины. Суть процесса эхометрии заключается в следующем: в трубное пространство с помощью датчика импульса звуковой волны посылается звуковой импульс, звуковая волна, пройдя по стволу скважины, отражается от уровня жидкости, возвращается к устью скважины и улавливается чувствительным микрофоном. По полученному сигналу фиксируется половинное время пробега акустического импульса по межтрубному пространству скважины от момента его посылки до прихода значимого отклика, которое затем умножают на значение скорости звука в затрубном газе. Полученная величина принимается за искомый уровень жидкости в скважине. Современные приборы автоматически определяют отражение от уровня жидкости, измерение времени прохождения сигнала и оценку скорости. Однако распознавание положения отражения от уровня жидкости на эхограмме все еще остается проблемой.

Для корректного определения уровня жидкости с помощью эхометрии необходима точная фиксация времени отклика сигнала, которое зависит от термобарических условий в скважине (давление, температура), наличия пены, часто образующейся в результате газосепарации. Отражение звуковой волны может происходить не только от зеркала жидкости, но и от образующейся пены, от любой границы раздела сред, где имеется существенное изменение плотности, или от любого

Эхометрия один из самых широко используемых методов определения уровня жидкости в скважине. Положение отражения звуковой волны от жидкости на сигнале (эхограмме) отображается в виде убывающих пиков, причем расстояние между ними всегда одинаково. Определение положения этих пиков напрямую зависит от качества полученного сигнала, так как на него влияют внешние звуки и посторонние объекты в скважине. В данном исследовании предложен метод моделирования сигнала через решение задачи квадратичного программирования, чтобы убрать шум, и алгоритм поиска пиков, основанный на максимизации целевой функции, описывающей сумму влияния от всех пиков. При помощи полученных результатов были определены признаки, отличающие нормальные сигналы от поврежденных. Эти признаки были использованы в задаче классификации эхограмм, с упором на точное определение именно нормальных сигналов. Результаты показали, что предложенные методы точно определяют положение пиков и классификатор корректно отбирает неповрежденные эхограммы.

постороннего объекта в скважине. Также на волну могут влиять внешние звуки или колебания прибора. Это напрямую влияет на точность полученных данных.

В так называемом нормальном случае, если измерению ничего не помешало, сигнал имеет характерные пики, которые со временем убывают, так как эхо всегда только ослабевает, и находятся на одинаковом расстоянии друг от друга. В противном случае, сигнал считают бракованным. Время между появлением этих пиков и есть время отклика. Для автоматического анализа эхограмм требуется разработка алгоритмов для универсального шумоподавления сигнала и точного обнаружения откликов.

В диссертации [1] подобная задача решалась путем увеличения отношения сигнала к шуму усовершенствованием самого метода эхометрии и алгоритма, работающего на фазо-частотных характеристиках сигнала. В других же источниках задачи автоматической обработки эхограмм не решалось.

Все существующие алгоритмы для уменьшения шума и поиска пиков основаны на представлении сигнала во временной или частотной области. Так в других источниках были использованы различные подходы: гибридная линеаризация и анализ основных компонентов [2], методы кратномасштабного анализа и сплайн-вейвлеты [3], модифицированный алгоритм Пана-Томпкинса и метод двойной огибающей Гилберта [4], фильтр Савицкого-Голая и определение пороговых значений [5], адаптивный метод сегментации [6]. Приведенные алгоритмы подстроены под определенные сигналы, например, биометрические, и работают с их конкретными особенностями: различные ограничения и

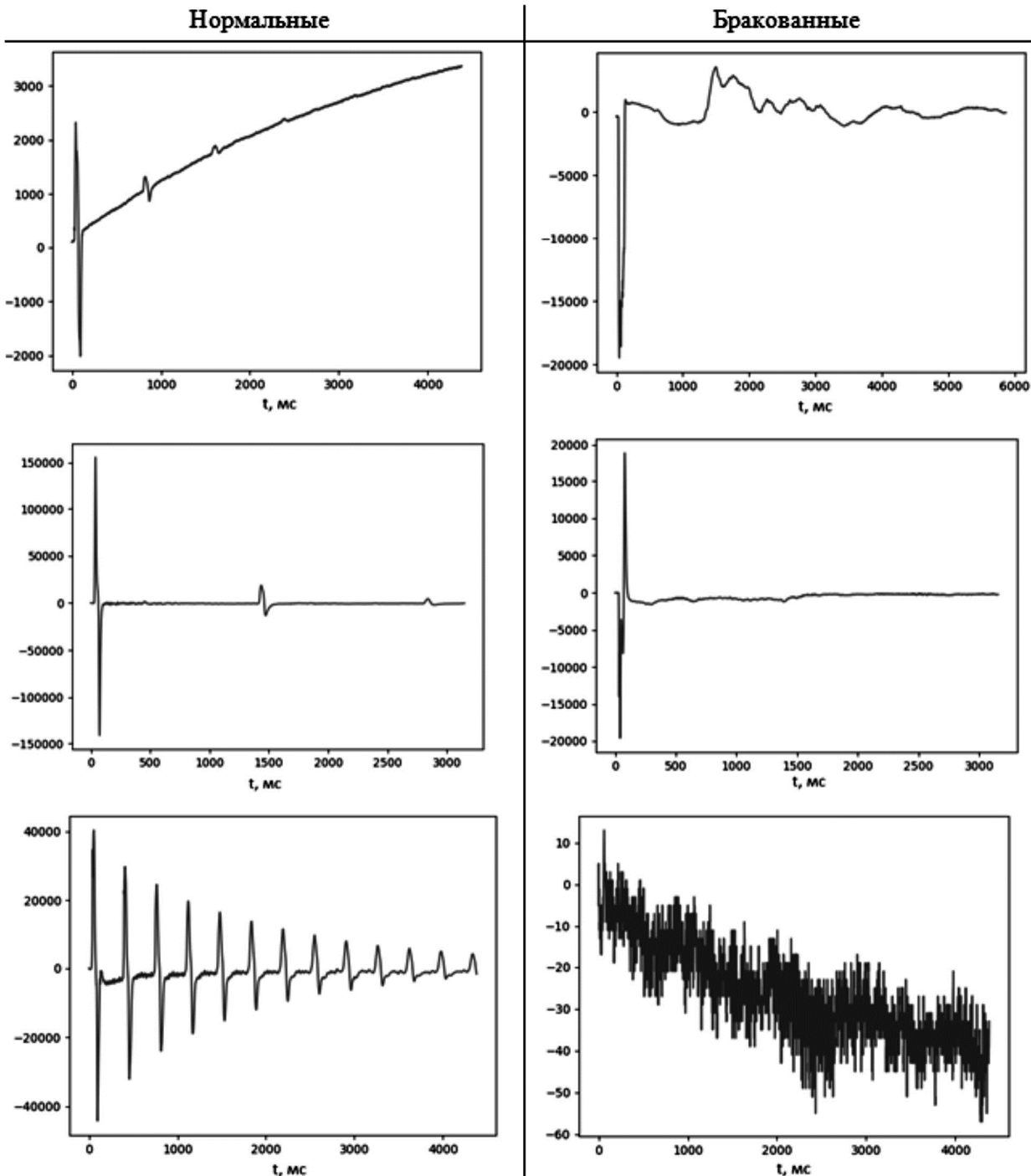
допущения на высоту пиков, расстояние между ними и их количество. Такой подход не гарантирует автоматизированную обработку сигналов.

Цель работы – разработка алгоритма для поиска пиков на предварительно смоделированном сигнале и классификация эхограмм на нормальные и бракованные.

Данные

В исследовании использовалось 12630 эхограмм, которые можно условно разделить на нормальные и бракованные. В табл. 1 представлены примеры таких сигналов.

Табл. 1. Примеры нормальных и бракованных эхограмм



Можно заметить, что в начале каждого нормального сигнала есть самый высокий пик, так называемое начальное смещение (положение нулевого пика) a_0 . Расстояние между пиками будет обозначаться d , их количество – k , то есть положение каждого пика определяется как $a_0 + i \cdot d, i = 1, \dots, k$. Расстояние между этими пиками считается одинаковым. На рис. 1 проиллюстрированы описанные выше термины.

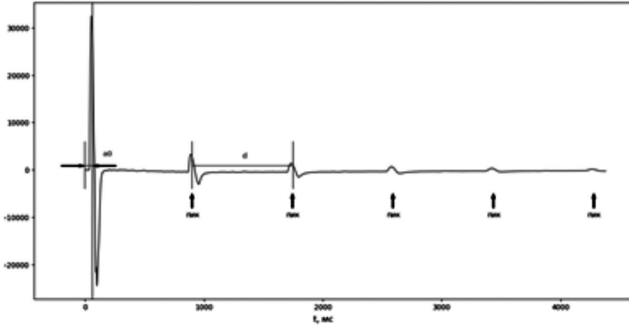


Рис. 1. Схематическое изображение начального смещения и расстояния между пиками

Также можно выделить отдельную категорию эхограмм, изображенную на рис. 2. Ее особенностью является крайне малое расстояние между пиками и нельзя однозначно сказать, нормальный это сигнал или нет. Сигнал выглядит таким образом из-за образования пены в скважине.

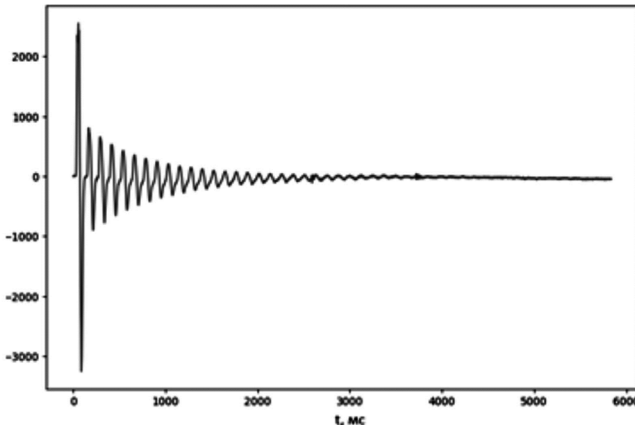


Рис. 2. Вид сигнала при образовавшейся в скважине пене

Задача поиска расстояния между пиками Моделирование сигнала

Для поиска оптимального расстояния между пиками нужно смоделировать сигнал на некоторой сетке $\{\mu_i\}_{i=1}^M$, сетка задается как

$$\mu_k = \mu_1 + \Delta \cdot k, k = 1, \dots, M, \text{ где } M = \left\lceil \frac{N}{\Delta} \right\rceil, \quad (1)$$

где N – продолжительность сигнала, μ_1 и Δ задаются заранее. Каждый пик можно описать функцией вида

$$f(t, \mu) = -\frac{t - \mu}{\sigma} \exp \left\{ -\left(\frac{t - \mu}{\sigma} \right)^2 \right\}. \quad (2)$$

Исходный сигнал можно представить в виде

$$y(t) = \alpha(t) + \sum_{j=1}^M \beta_j f(t, \mu_j) + \varepsilon(t), \quad (3)$$

где $\alpha(t)$ – линейный тренд, f – базисные функции вида (2), $\{\beta_i\}_{i=1}^M$ – некоторые коэффициенты, $\varepsilon(t)$ – шум.

Для разложения сигнала на коэффициенты β_1, \dots, β_M решается задача квадратичного программирования

$$\min_{\beta} \sum_{i=1}^N \left(y(t) - \alpha(t) - \sum_{j=1}^M \beta_j f(t, \mu_j) \right)^2 + \sum_{j=1}^M \lambda_j \beta_j^2, \beta > 0, \quad (4)$$

штрафные значения определяются как

$$\lambda_j = \frac{l_0}{j + (m_0)^2} + l_1, j = 1, \dots, M, \quad (5)$$

где l_0, l_1, m_0 задаются заранее.

Проводя аналогию с вейвлет-преобразованиями, можно сказать, что в качестве базиса (материнского вейвлета) берется эрмитов вейвлет. В данном случае используется только единственный масштаб, а параметр положения задается на заранее заданной сетке (1), в базисных функциях $f(t, \mu_j)$ отвечает за смещение функции вправо вдоль оси x (рис. 3). Можно рассматривать разложение как сглаживание высоких частот: из оригинального сигнала $y(t)$ получаем отфильтрованный «сигнал», выраженный коэффициентами β_1, \dots, β_M . В свою очередь использование тренда $\alpha(t)$ можно представить как фильтр низких частот.

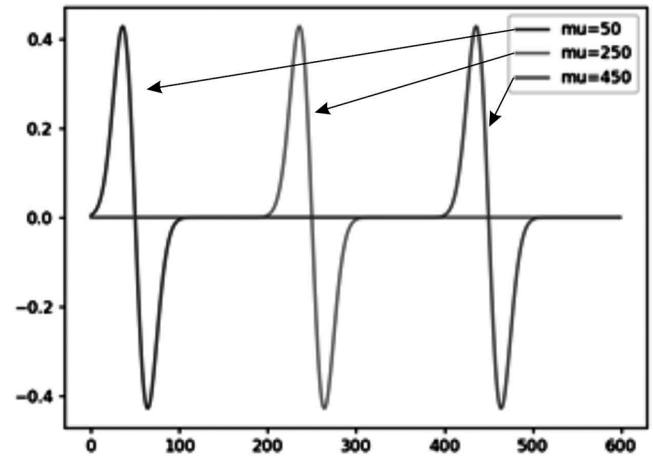


Рис. 3. Вид функции $f(t, \mu)$ при различных μ

Условие $\beta > 0$ позволяет сконцентрировать влияние именно на целевых пиках, исключая слишком сильное влияние шума. В противном случае коэффициенты разложения будут чередовать знаки в области шума.

Штрафная функция λ задается таким образом, чтобы она убывает подобно тому, как убывают пики. Таким образом в модель сигнала не попадает шум и сохраняется отношение высот пиков к друг другу. На рис. 4 изображены модели с штрафом вида $\lambda = const$ и с штрафом в виде убывающей функции λ (5) для исходного сигнала.

Сначала сигнал моделируется с фиксированными параметрами $l_0 = 150, m_0 = 0,1, \sigma = 15, \Delta = 16, l_1$ выбирается в зависимости от высоты нулевого пика h , эти значения были выбраны экспериментально (табл. 2).

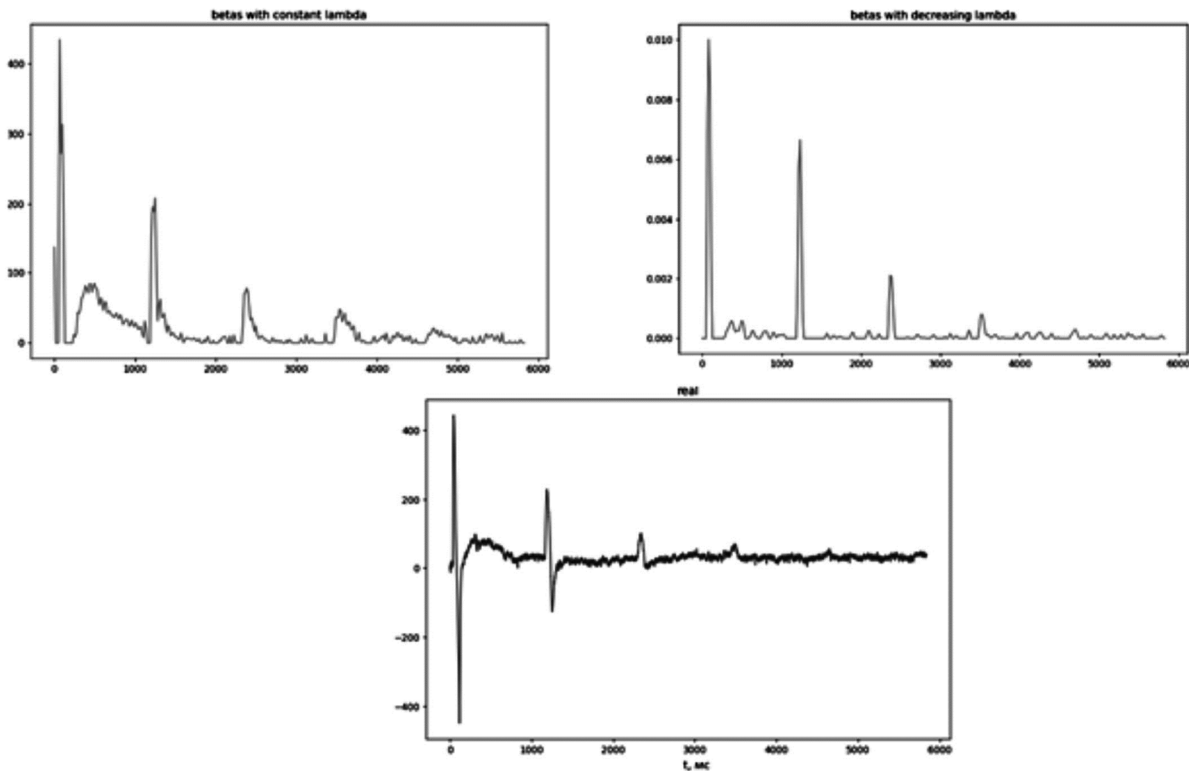


Рис. 4. Исходный сигнал и примеры его разложения при $\lambda = const$ и убывающих λ

Табл. 2. параметры выбора l_1

Высота нулевого пика h	l_1
$h < 50000$	250
$50000 \leq h < 500000$	150
$h \geq 500000$	25

На рис. 5 изображена разница в коэффициентах разложения при заниженном значении l_1 относительно истинного сигнала. Для большей наглядности вычисленные коэффициенты β и сам сигнал были нормированы. Можно заметить, что при $l_1 = 25$ первый пик больше нулевого, то есть таким образом результаты поиска пиков будут заведомо неверными.

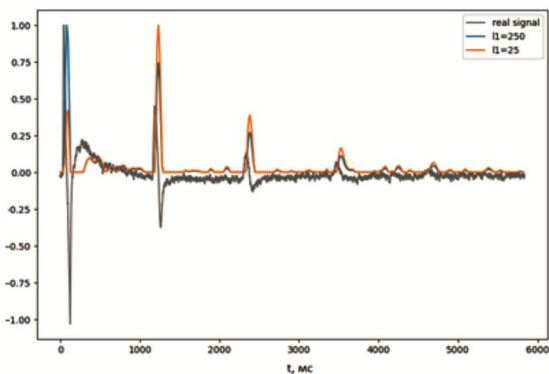


Рис. 5. Коэффициенты β при различных l_1

После первоначального разложения на коэффициенты β_1, \dots, β_M оценивается предположительное расстояние между пиками d_0 . Эта оценка есть среднее значение расстояния между соседними локальными максимумами. Стоит отметить, что такой подход не эффективен для поиска самих пиков, так как не каждый локальный максимум соответствует самому пику. Если

$d_0 < 10$, очевидно, что данный сигнал описывает сигнал подобный рис. 2, и нужно получить новые коэффициенты с меньшим шагом $\Delta = 8$, чтобы пики корректно определялись в коэффициентах β .

В зависимости от высоты нулевого пика h и величины d_0 , выбираются значения σ , которые описывают ширину пика (табл. 3).

Табл. 3. Параметры выбора σ

h, d_0	σ
$d_0 \geq 50, h > 150000$	35
$40 < d_0 \leq 75, 50000 < h \leq 150000$	40
$d_0 \geq 90, 50000 < h \leq 150000$	50
$75 \leq d_0 < 90, 100000 < h \leq 150000$	70

Если значение σ слишком мало или велико, в разложении искажается истинное положение пика и отношение их высот (рис. 6).

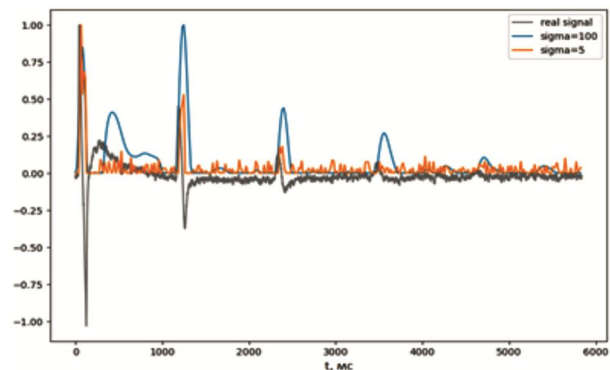
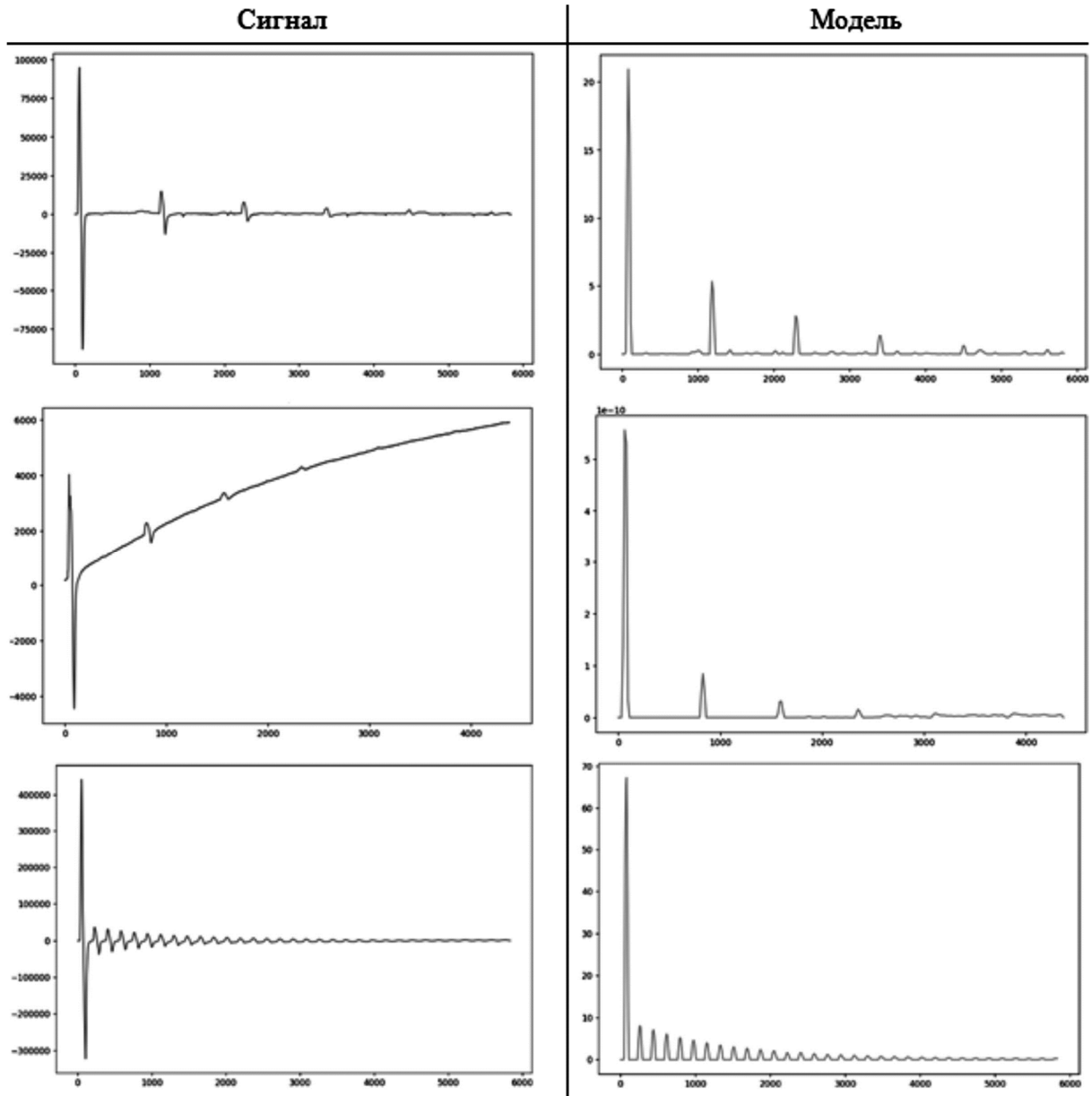


Рис. 6. Коэффициенты β при различных σ

Примеры конечного разложения сигнала представлены в табл. 4.

Табл. 4. Вид оригинального сигнала и его модели для различных эхограмм



Алгоритм поиска расстояния между пиками

Пусть имеются коэффициенты разложения β_1, \dots, β_M . Для поиска a_0 и d используется максимизация функции

$$R(a_0, d) = \sum_{k=0}^K g_k - \lambda_2 \sum_{k=1}^K (g_k - g_{k-1}) I(g_k - g_{k-1} > 0), \quad (6)$$

где $g_k = \sum_{j=-s}^s w_j \beta_{a_0+k \cdot d+j}$, $I(A)$ – индикаторная функция события A. (7)

Если предположить, что пик находится в положении $a_0 + k \cdot d$ в коэффициентах β , то смысл (7) заключается в том, что этот пик определяется несколькими соседними для $\beta_{a_0+k \cdot d}$ коэффициентами. Таким образом g_k описывает влияние пика. Весовые коэффициенты w_{-s}, \dots, w_s определяются так, чтобы они квадратично убывали от центра и наиболее значимым был цен-

тральный коэффициент w_0 , который задается заранее. Количество точек справа и слева s фиксировано и также задается заранее. На рис. 7 изображен вид весовых коэффициентов.

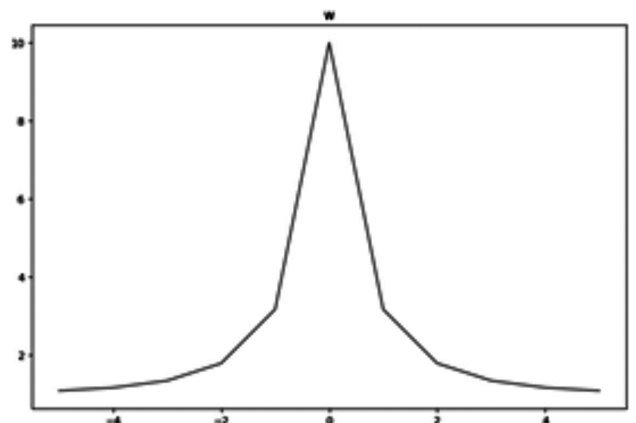


Рис. 7. Вид весовых коэффициентов

В (6) максимизируется сумма влияния всех пиков, при этом случай, когда $g_k - g_{k-1} > 0$ означает, что влияние сигнала со временем возросло, а это противоречит постановке задачи, так как это может только ослабевать. Поэтому второе слагаемое налагает штраф на несоответствующим образом выбранные шумы. Значение λ_2 должно быть достаточно большим, чтобы не пропустить лишние данные в итоговые значения функции R_0 , в данной задаче использовалось $\lambda_2 = 10$.

Для поиска нулевого смещения нужно максимизировать функцию g_0 по a_0 , перебирая значения a_0 в пределах третьей части всего времени, так как известно, что этот пик должен быть в начале сигнала. После этого нужно максимизировать функцию R_0 по некоторым значениям d . Эти значения должны быть не меньше оценки d_0 , чтобы не попасть в область влияния нулевого пика, и идти с шагом 0,5 по всем возможным значениям β . Нецелые значения d нужны для того, чтобы не упустить влияние пика между точками. В (6) суммирование

по k ведется до возможного значения количества пиков для текущего d_i , то есть до

$$K = \left\lceil \frac{M - a_0}{d_i} \right\rceil. \tag{8}$$

Важно отметить, что, если значение d не является целым, g_k определяется по следующей формуле

$$g_k = \sum_{j=-s}^s w_j \left[\frac{1}{2} \beta_{\lfloor x \rfloor + j} + \frac{1}{2} \beta_{\lceil x \rceil + j} \right], \tag{9}$$

где $x = a_0 + k \cdot d$, $\lfloor x \rfloor$ и $\lceil x \rceil$ ближайшее целое снизу и сверху для x соответственно.

На рис. 8 для исходного сигнала представлен вид функций g_0 и R_0 , при уже найденном a_0 .

Полученные значения a_0 и d нужно умножить на Δ , чтобы получить эти значения на оригинальном масштабе. В табл. 5 представлен результат работы алгоритма на коэффициентах β и самом сигнале.

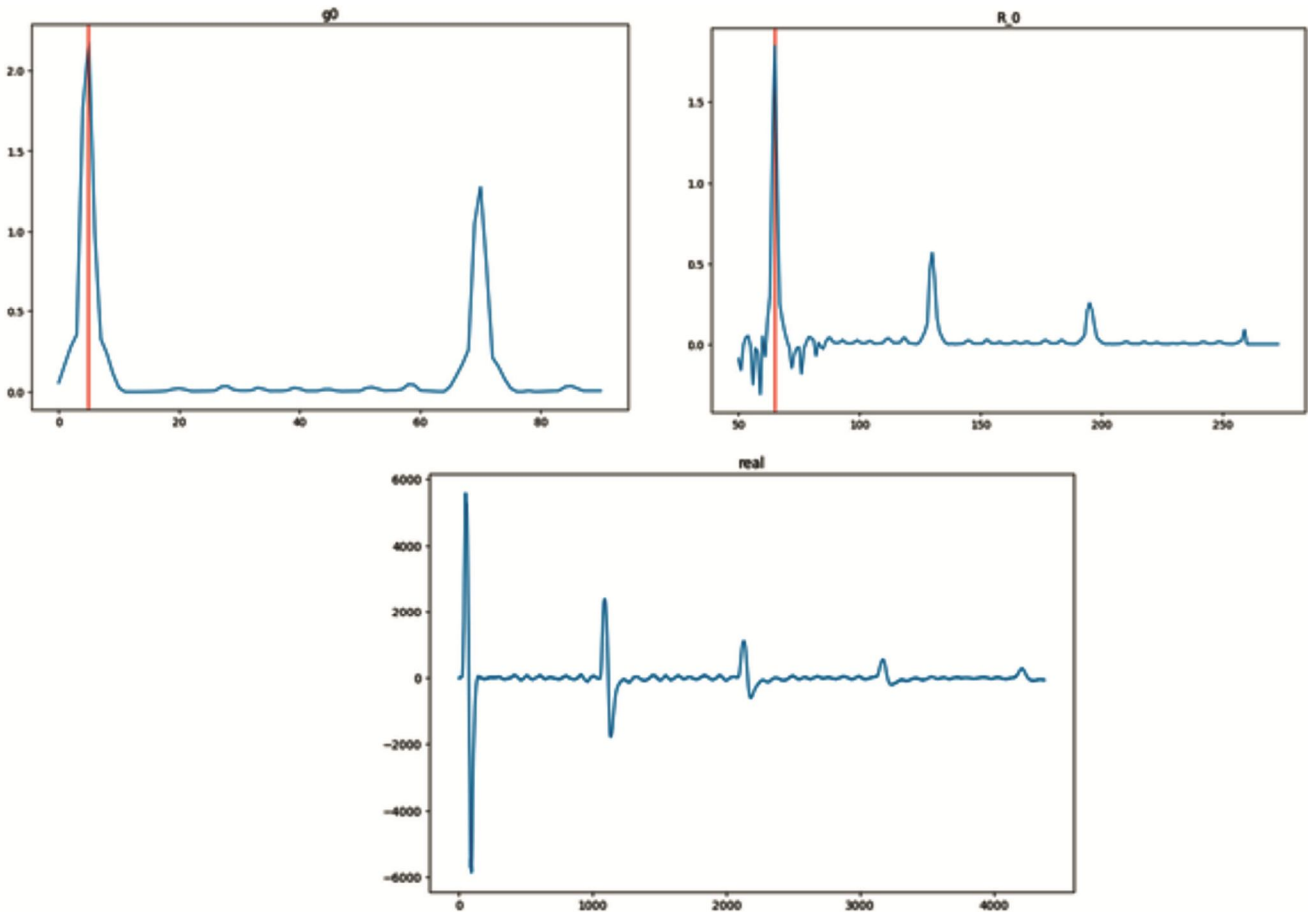
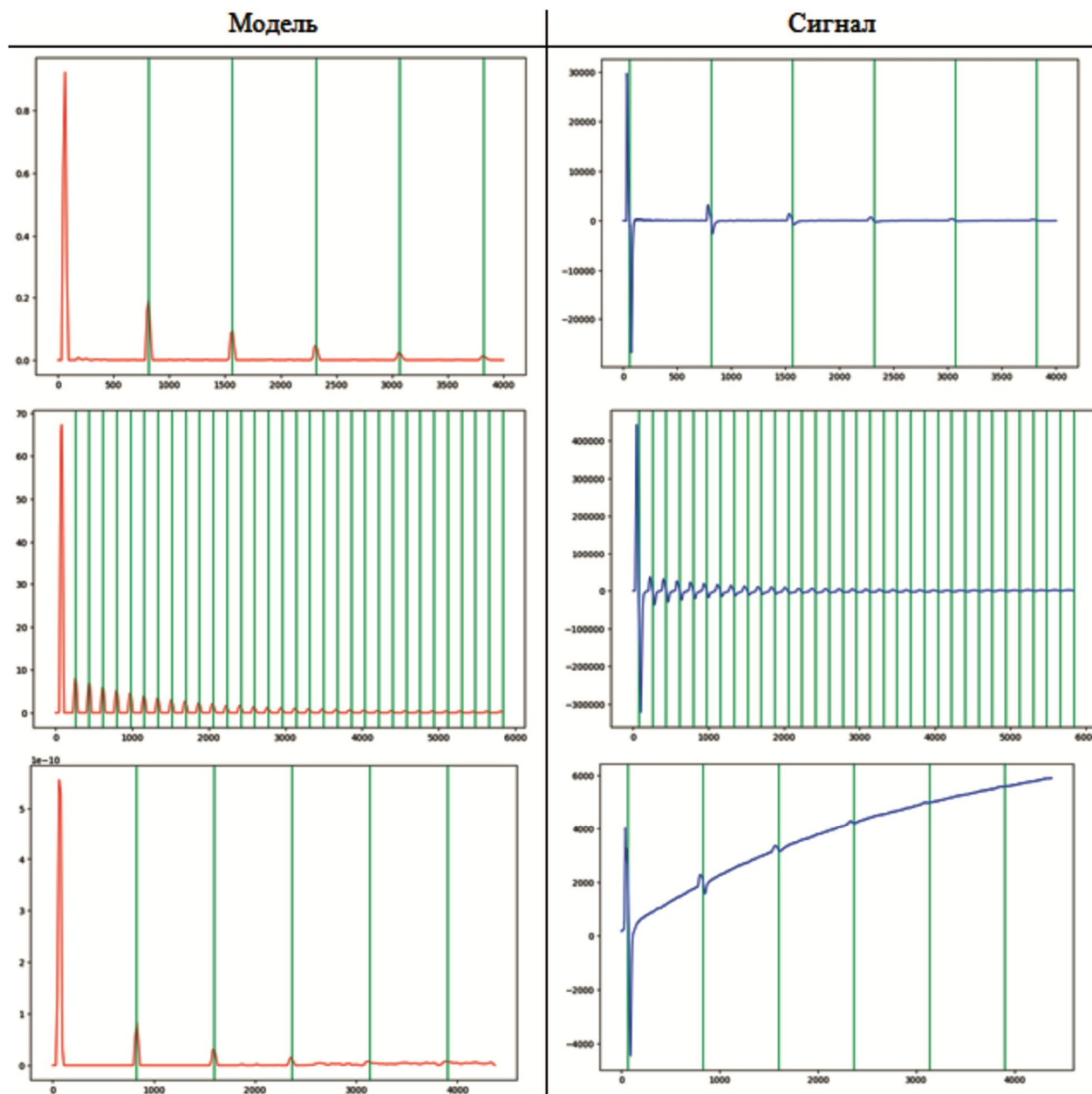


Рис. 8. Вид исходного сигнала и функций g_0 и R_0 для него

Табл. 5. Пример работы алгоритма на сигнале и его модели



Задача классификации

В результате работы алгоритма для каждой эхограммы получены свои значения a_0 , d и k . Используя эти значения нужно определить, является ли данная эхограмма не бракованной. Для этого следует найти определяющие признаки, отличающие нормальные и бракованные сигналы. Было выбрано четыре признака:

Пусть первый пик рассматривается в окрестности равной половине найденного расстояния между пиками. Признаком, назовем его kol , считается количество локальных максимумов, высота которых в пределах 20 – 80 % высоты глобального максимума (сам пик). Очевидно, что для нормальной эхограммы это количество должно равняться нулю. Для бракованных эхограмм характерно, что значения сигнала в этой окрестности принимают отрицательные значения или, по крайней мере, глобальный максимум имеет значение, меньше 10, или количество пиков $k > 15$, в таком слу-

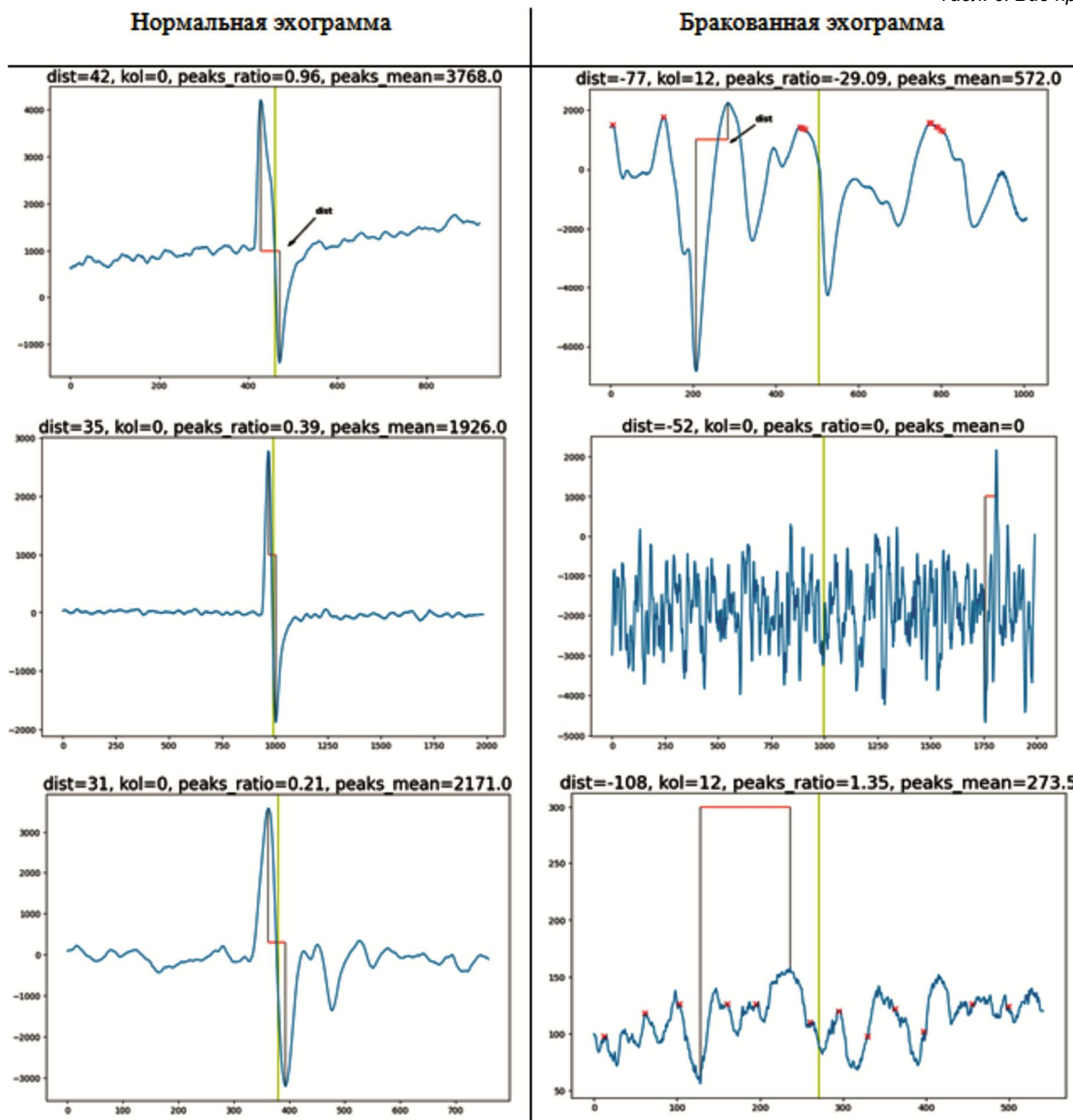
чае kol принимает значение 100. Также в качестве признака использовалось расстояние между глобальными максимумом и минимумом $dist$, что есть ширина пика.

Следующие два признака $peaks_ratio$ и $peaks_mean$ определяются как среднее значение отношения высот соседних пиков и среднее значение высот этих пиков соответственно.

В табл. 6 представлен вид признаков для различных эхограмм.

В качестве модели машинного обучения выбрана логистическая регрессия. И нужно отметить, что несмотря на большое количество данных, многие эхограммы практически одинаковы, поэтому для обучения использовалось 60 бракованных и 60 нормальных эхограмм, отобранных вручную. Полученные результаты показали, что некоторые нормальные сигналы в единичных случаях могут идентифицироваться моделью как бракованные. Однако главная задача по поиску

Табл. 6. Вид признаков



только нормальных сигналов выполнена. На рис. 9 представлена ROC-кривая для результатов 5-кратной перекрестной проверки.

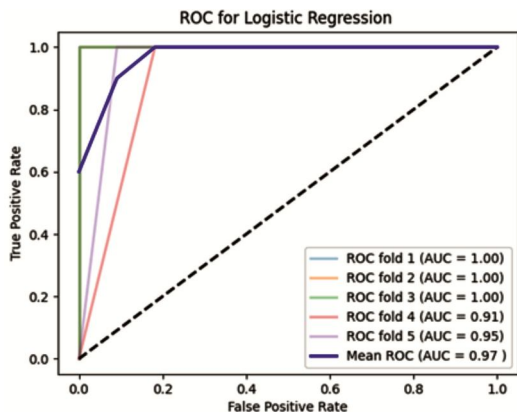


Рис. 9. ROC-кривая для результатов 5-кратной перекрестной проверки

Эхограммы с $d < 15$ в классификации не участвуют,

они образуют отдельную категорию по этому признаку.

Заключение

В статье предложены методы построения модели сигнала и поиска расстояния между пиками в сигнале и приведена их реализация. Так же была рассмотрена задача классификации для отбора нормальных сигналов. Полученные результаты и сделанные по ним выводы позволяют рекомендовать данный метод для практического применения в нефтяной промышленности.

Работа выполнена при поддержке Минобрнауки России в рамках соглашения № 075-15-2022-299 о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Рациональное освоение запасов жидких углеводородов планеты».

Литература

1. Налимов К. Г. Информационная система эхометри-

рования многоимпульсными сигналами для определения уровня жидкости в нефтедобывающих скважинах: дис. ... канд. техн. наук 05.13.01. Томский политехнический университет, Томск, 2007. 132 с.

2. Жолмагамбетова Б.Р., Мазиков Т.Ж., Букунов М.М., Изат Э.Ж. Обнаружение и шумоподавление R-пиков электрокардиограммы с гибридной линеаризацией и анализом основных компонентов. Труды университета №3 (80) 2020; с. 157-162.

3. Suyi Li, Shanqing Jiang, Shan Jiang, Jiang Wu, Wenji Xiong, Shu Diao. A Hybrid Wavelet-Based Method for the Peak Detection of Photoplethysmography Signals. Hindawi Computational and Mathematical Methods in Medicine Vol. 2017. 9468503, 8 с. <https://doi.org/10.1155/2017/9468503>

4. Filipa Esgalhado, Arnaldo Batista, Valentina Vassilenko,

Sara Russo, Manuel Ortigueira. Peak Detection and HRV Feature Evaluation on ECG and PPG Signals. Symmetry 2022, 14, 1139 с. <https://www.mdpi.com/2073-8994/14/6/1139>

5. Дахва М.С., Леухин А.Н. Сравнение пяти алгоритмов для обнаружения R-пиков в ЭКГ-сигнале. Вестник Поволжского государственного технологического университета. Сер.: Радиотехнические и инфокоммуникационные системы. 2018. № 3 (39). С. 39-49. DOI: 10.15350/2306-2819.2018.3.39

6. Kavsaoğlu, Ahmet & Polat, Kemal & Bozkurt, Mehmet. (2016). An innovative peak detection algorithm for photoplethysmography signals: An adaptive segmentation method. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES. 24. 1782-1796. 10.3906/elk-1310-177.

НОВЫЕ КНИГИ

Воскресенский Д.И., Добычина Е.М.

Цифровые антенные решетки: Монография – М.: Изд-во Радиотехника, 2020 г. – 240 с.: ил.

Рассмотрен новый класс антенных систем – цифровых антенных решеток бортовых радиолокационных комплексов, позволяющих повысить энергетический потенциал за счет использования новых методов оптимизации режима работы высокоэффективных активных устройств и обеспечения высокой точности формирования амплитудно-фазового распределения с помощью предложенной системы автоматической калибровки. Представлены результаты экспериментальных исследований возможностей цифрового диаграммообразования и точностных характеристик калибровки макета цифровой решетки.

Для научных работников и инженеров, занимающихся исследованиями в области разработки, создания и применения цифровых решеток в современных радиоэлектронных системах. Может быть рекомендована в качестве учебного пособия студентам радиотехнических специальностей, а также аспирантам и магистрантам по направлениям «Радиотехника, «Радиофизика и электроника».

Фильтрация и спектральный анализ радиосигналов.

Алгоритмы. Структуры. Устройства. Под ред. Ю.В. Гуляева: Монография – М.: Изд-во Радиотехника, 2020 г. – 504 с.: ил.

Рассмотрены устройства на поверхностных и объемных акустических волнах. Приведены принципы построения акустооптических Фурье-процессоров, даны методики их описания и характеристики. Синтезированы алгоритмы многоканальных частотных дискриминаторов, имеющих широкую дискриминационную характеристику с большой зоной линейности, что повышает точность измерения частоты радиосигналов, а также стабильность работы следящих измерителей частоты при интенсивных воздействиях помех. Обобщены результаты математического моделирования и экспериментальных исследований волноводных СВЧ-фильтров и мультиплексоров X-диапазона частот на основе прямоугольных и круглых волноводов. Показано, что устройства предназначены для работы в составе негерметизированных радиоэлектронных комплексов спутниковых систем космической связи. Исследованы двумерные периодические структуры в виде перфорированных тонких металлических экранов, малые размеры и масса которых

делают их технологически привлекательными в технике миллиметрового и субмиллиметрового диапазонов. Представленные известные и оригинальные авторские решения по широкому спектру вопросов проектирования устройств селекции могут быть полезны широкому кругу научных работников и инженеров, специализирующихся в области проектирования фильтров и аналоговых Фурье процессоров.