

Поиск скрытых закономерностей в потоках импульсных сигналов акустической и электромагнитной эмиссии

Ю. И. Сенкевич

Институт космофизических исследований и распространения радиоволн
дальневосточного отделения Российской Академии Наук (ИКИР ДВО РАН)
с. Паратунка, Камчатский край, Россия
senkevich@ikir.ru

Аннотация. Своевременное предупреждение катастроф, вызываемых землетрясениями, обеспечивает безопасность жизнедеятельности. Поэтому поиск маркеров предсейсмических событий, предшествующих землетрясениям, остается важнейшей исследовательской задачей. В статье представлены экспериментальные методы оценки сейсмической активности в Камчатском регионе по результатам обработки и анализа сигналов геоакустической и электромагнитной эмиссии. Проведенные исследования направлены на обнаружение аномалий количественных и качественных показателей, характеризующих потоки импульсов акустической эмиссии приповерхностных пород и электромагнитной эмиссии в приземном слое атмосферы. Обработка и анализ сигналов осуществляются с применением специальных алгоритмов, учитывающих структурные особенности разнообразия форм импульсов и их распределения во времени.

Ключевые слова: акустическая эмиссия; электромагнитная эмиссия; структурно-лингвистический анализ; паттерн импульса; сейсмическое событие

I. ВВЕДЕНИЕ

Вопросы поиска устойчивых маркеров сейсмических событий на основе анализа сейсмических и электромагнитных сигналов, не смотря на достаточно долгую историю изучения, постоянно находятся в поле внимания исследователей. В этой области знаний были созданы наиболее значимые модели предсказаний землетрясений, например [1, 2]. В основном рассматривался диапазон очень низких частот (до 100 Гц) акустических колебаний. В последние годы стали появляться работы по изучению геофизических сигналов (ГФС) более высокого диапазона частот (от 20 Гц до 10–16 кГц и выше), связанных с физическими процессами пластической деформации. Это сигналы геоакустической эмиссии (ГАЭ), наблюдаемые в приповерхностном слое Земли, а также сигналы электромагнитного излучения (ЭМЭ), зарегистрированные в приземном слое атмосферы [3, 4]. На протяжении последнего десятилетия в ИКИР ДВО РАН ведутся исследования возможности создания

детекторов аномалий для таких геофизических сигналов (ГФС) и связи этих аномалий с сейсмическими событиями. Несмотря на различную физическую природу, акустические и электромагнитные эмиссионные сигналы при ближайшем рассмотрении имеют много общих свойств, связанных с разнообразием форм импульсов и групповым поведением импульсов в потоке. Авторами [4, 5] показано, что изменение импульсного разнообразия в различные моменты времени связано с сейсмическими событиями. Поэтому возникает особый интерес к изучению динамических характеристик эмиссионных ГФС. В настоящем материале представлено одно из направлений исследований, базирующееся на информационном подходе к обработке ГФС [6], идея которого состоит в трансформации импульсного потока эмиссионного ГФС в кодовое сообщение, что позволит применять к нему методы и алгоритмы лингвистического анализа.

II. АНАЛИЗ ИНФОРМАТИВНОСТИ ИМПУЛЬСНЫХ ПОТОКОВ ГЕОФИЗИЧЕСКИХ СИГНАЛОВ

Для получения нормальных и аномальных характеристик эмиссионных ГФС необходимо классифицировать импульсы в потоке, а затем охарактеризовать их распределение по классам на временной оси.

На первом этапе осуществляется кластеризация импульсов по форме по значениям «особых» точек, которые в значительной степени характеризуют функцию сигнала. Такими точками являются локальные экстремумы амплитудно-фазового паттерна импульса, как это имеет место в ряде методов математической морфологии при распознавании и описании графических объектов [7, 8]. Фиксируются значения амплитуды локальных экстремумов и временные промежутки между соседними экстремумами. Таким образом, формируются два ряда значений, которые попадают под определение упорядоченных множеств: $\{a_i\} = a_0, a_1, \dots, a_N$ и $\{\tau_i\} = \tau_0, \tau_1, \dots, \tau_{N-1}$. Строятся две квадратные матрицы ранговых отношений значений амплитуд локальных экстремумов и значений интервалов между ними по правилу:

Работа выполнена при финансовой поддержке РФН, проект № 18-11-00087).

$$r_{i,j} = \begin{cases} 1, & a_i > a_j \\ 0, & a_i \leq a_j \end{cases} \quad i, j = \overline{1, N}$$

$$\omega_{i,j} = \begin{cases} 1, & \tau_i > \tau_j \\ 0, & \tau_i \leq \tau_j \end{cases} \quad i, j = \overline{1, N-1}$$

Полученные бинарные квадратные матрицы с размерами N и $N-1$ соответственно, носят избыточный характер в силу свойств неравенств (если $r_{i,j} = 1$, то $r_{j,i} = 0$, аналогично для $\omega_{i,j}$), и их нижние треугольные части, включая главную диагональ, следует исключить. Матрица отношений временных интервалов между локальными экстремумами транспонируется. У матрицы отношений амплитуд локальных экстремумов удаляется первый столбец и последняя строка, содержащие незначимые нулевые элементы. Полученные матрицы складываются, в результате чего получаем квадратную бинарную матрицу вида:

$$D = \begin{pmatrix} r_{1,2} & r_{1,3} & \dots & r_{1,N-1} & r_{1,N} \\ \omega_{1,2} & r_{2,3} & \dots & r_{2,N-1} & r_{2,N} \\ \omega_{1,3} & \omega_{2,3} & \dots & r_{3,N-1} & r_{3,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{1,N-2} & \omega_{2,N-2} & \dots & r_{N-2,N-1} & r_{N-2,N} \\ \omega_{1,N-1} & \omega_{2,N-1} & \dots & \omega_{N-2,N-1} & r_{N-1,N} \end{pmatrix}$$

Показанная матрица описывает форму импульса с позиции амплитудно-фазовых отношений и является его базовой характеристикой, далее по тексту статьи такая матрица будет называться *матрицей-образом* импульса. В силу свойств неравенств матрица-образ приобретает свойство инвариантности по отношению к временной и амплитудной трансформации сигнала. Свойство инвариантности сохраняется также при наличии в сигнале монотонного тренда. Это означает, что каждая матрица-образ описывает некоторое разнообразие форм импульсов. Свойство инвариантности отражает физические процессы затухания и растяжения сигнала в плотных средах, а также дает возможность построить алгоритмы идентификации и кластеризации фиксируемого импульсного разнообразия.

Кластеры формируются из импульсов близких по форме на основе эмпирически установленного коэффициента сходства. Коэффициент сходства форм сравниваемых импульсов вычисляется по числу совпадающих элементов их матриц-образов. Каждый кластер характеризуется некоторой матрицей-образом, которой можно сопоставить символическое обозначение. Множество всех символов, ранжированное по числу соотнесенных к определенному кластеру импульсов, составляет алфавит, характеризующий импульсное разнообразие на временном интервале T . Примеры 15-минутного сигнала ГАЭ и его алфавита представлены на рис. 1. В нижней части рисунка на графике алфавита числа обозначают размеры символов. По вертикали откладывается число импульсов, включенных в кластер определенного символа.

На рис. 2 показаны результаты применения процедуры кластеризации при различных коэффициентах сходства. Процедура кластеризации позволяет снизить избыточность символического состава алфавитов, которая возникает в результате искажения паттерна импульса под влиянием помех, с одной стороны, и из-за нелинейности среды распространения, с другой стороны. Искажения паттернов импульсов изменяют матрицы-образы и порождают избыточные символы алфавита.

Сравнивая алфавиты, обнаруженные в равные последовательные эпизоды наблюдений, можно проводить количественную и качественную оценку изменений распределения импульсов и динамики разнообразия форм во времени с целью выявления аномалий в сигналах.

Запись символов в порядке появления соответствующих им импульсов реального сигнала формирует условный текст, который используется в дальнейшем анализе, как закодированное сообщение. Для «расшифровки» такого сообщения, то есть поиска скрытых связей, удобно применять, например, методы лингвистического анализа.

В данном исследовании анализ сообщений осуществляется с использованием следующих показателей и критериев.

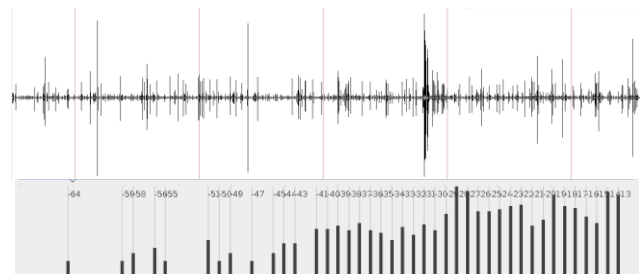


Рис. 1. Пример сигнала ГАЭ (вверху) и его алфавита (внизу). Числовые значения обозначают размер символа. Объем кластеров отображается в логарифмическом масштабе. Минимальная значащая частота повторяемости символа в файле – 2.

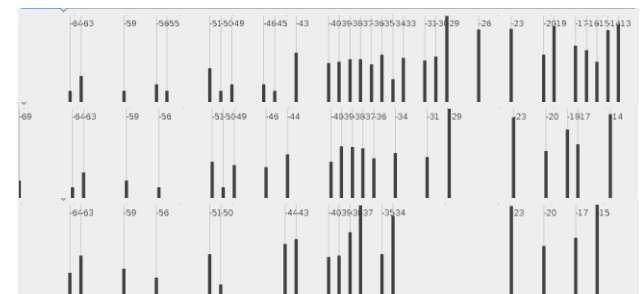


Рис. 2. Пример выполнения процедуры кластеризации импульсного потока представленного на рис.1. Сверху вниз коэффициенты сходства равны 0,9, 0,8 и 0,7 соответственно

Частота появления символа в алфавите различна и косвенно указывает на вероятное изменение поведения динамической системы. Динамика потока импульсов отражает функционирования этой системы. Поэтому при

анализе используется *вероятность появления символа* a_i алфавита \mathbf{A} , равная отношению числа появлений символа в сообщении – m к общему числу символов в этом сообщении – N : $P(a_i) = m/N$. С точки зрения теории информации этот показатель является оценкой информационной нагрузки, приходящейся на каждый символ алфавита. В соответствии с рекомендациями авторов символической динамики [9, с. 220] следует пытаться обнаружить связи на уровне цепи Маркова высоких порядков, рассчитывая условные вероятности появления сочетаний символов: $P(b/a) = P(ab)/P(a)$; $P(c/ab) = P(abc)/P(ab)$ и т.д., где a, b, c – символы выделенного из сигнала алфавита $\mathbf{A} = \{a, b, c, \dots, z\}$, z – это «нулевой символ», знак, который дополняет алфавит до полного. Этим символом обозначаются все встречающиеся импульсы в потоке, матрицы-образы которых не повторялись на эпохе анализа.

Показатель *скорости производства новых символов* (v) равен отношению количества новых символов $\Delta N = N_{i+1} - N_i$, произведенных системой за время заданного интервала измерения, к длине интервала Δt : $v = \Delta N / \Delta t$; если $v > 0$ – наблюдается возбуждение процессов в порождающей сигнал среде, выход из устойчивого состояния; если $v < 0$ – наблюдается стабилизация процессов в порождающей сигнал среде, вход в устойчивое состояние.

Критерий насыщения алфавита – скорость производства новых символов при приближении к моменту обнаружения конечного числа символов алфавита \mathbf{A} за время наблюдения $T \gg t_0$ (t_0 – установленный интервал наблюдения) стремится к нулю, $\Omega = \lim_{\Delta t \rightarrow T} \Delta N / \Delta t = 0$. При длительном наблюдении может наступить момент, когда новые символы не появляются. Критерий используется при обучении рабочих алгоритмов определения стационарных состояний порождающей среды.

Критерий связи алфавитов рассчитывается, как степень их пересечения. *Коэффициент символического перекрытия алфавитов* k_{AB} для алфавитов \mathbf{A} и \mathbf{B} размерами N и M соответственно определяется по формуле: $k_{AB} = |\mathbf{A} \cap \mathbf{B}| / (N^2 + M^2)$.

Введем следующие определения и операции для манипуляции с символами и алфавитами. Каждый элемент алфавита – это объект с атрибутами: размер символа – s , частота появления символа сообщения – p и матрица-образ – \mathbf{D} . Иначе говоря, рассматривая алфавиты, имеем дело с множествами, состоящими из объектов, каждый из которых также представляет самостоятельное неоднородное упорядоченное множество-кортеж $a = \{s, p, \mathbf{D}\}$. Алфавит: $\mathbf{A} = \{a = \{s, p, \mathbf{D}\}\}$, $s \in \mathbf{S}$, $p \in \mathbf{P}$, $\mathbf{D} \in \mathbf{D}$. Рассмотрим действия с алфавитами.

Определение 1. Объединением двух алфавитов \mathbf{A}_i и \mathbf{A}_j является алфавит \mathbf{A}_k , построенный по правилу: $\mathbf{A}_k = \mathbf{A}_i \cup \mathbf{A}_j = \{a_k: a_k \in \mathbf{A}_i \text{ или } a_k \in \mathbf{A}_j\}$, где $a_k = \{s, p, \mathbf{D}\}$, если $a_k \in \mathbf{A}_i$ и $a_k \notin \mathbf{A}_j$; $a_k = \{s, p, \mathbf{D}\}$, если $a_k \notin \mathbf{A}_i$ и $a_k \in \mathbf{A}_j$; $a_k = \{s, p_i + p_j, \mathbf{D}_i\}$, если $a_k \in \mathbf{A}_i$ и $a_k \in \mathbf{A}_j$.

Определение 2. Пересечением двух алфавитов \mathbf{A}_i и \mathbf{A}_j является алфавит \mathbf{A}_k , построенный по правилу: $\mathbf{A}_k = \mathbf{A}_i \cap \mathbf{A}_j = \{a_k: a_k \in \mathbf{A}_i \text{ и } a_k \in \mathbf{A}_j\}$, где $a_k = \{s, p_i + p_j, \mathbf{D}_i\}$.

Для методов выделения импульсов, выделения алфавитов сообщений, использующихся для обработки и анализа эмиссионных ГФС, были разработаны компьютерные программы. Осуществлялась оценка чувствительности и помехоустойчивости реализованных алгоритмов с применением численного эксперимента [10].

Воспользуемся введенным определением пересечения алфавитов для оценки изменений, происходящих в импульсном потоке от одного временного эпизода наблюдения к другому. Следует ожидать, что при наличии связи процессов внутри порождающей и/или влияющей на генерацию импульсов среды с количественными и качественными показателями импульсного потока, символичный состав алфавитов эпизода до возникновения некоторого события и эпизода, связанного с событием, должен заметно измениться. Пример такого изменения при снижении влажности почвы представлен на рис. 3. При высыхании почвы число импульсов уменьшается. Исчезают символы размерностью больше 100, образованные цепочками накладывающихся импульсов (рис. 4, слева), обусловленных просачиванием пузырьков почвенных газов, а также длительные высокоамплитудные импульсы, обусловленные растрескиванием почвы (рис. 4, справа).

Разработанные алгоритмы преобразования импульсных потоков в сообщения позволяют наблюдать аномалии в 3D-мерном представлении последовательно измеренных алфавитов. Пример динамики алфавитов сигнала ЭМЭ показан на рис. 5. На графике видны аномальные изменения размерности алфавитов сигналов ЭМЭ, полученных на станции ИКИР ДВО РАН «Карымшина»

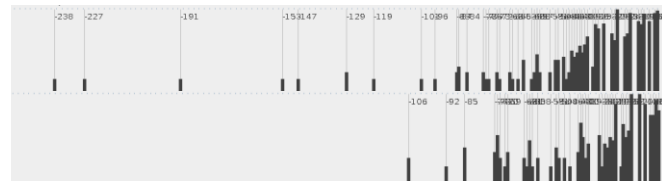


Рис. 3. Изменение состава алфавитов при снижении влажности почвы. Влажная почва в начале суток (вверху) и менее влажная почва в конце суток (внизу)

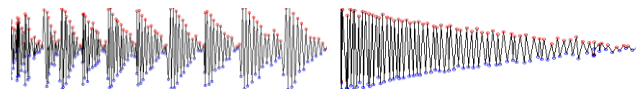


Рис. 4. Цепочка импульсов ГАЭ с наложением (слева). Импульс ГАЭ со 130 локальными экстремумами (справа)

(Камчатка) 4, 9, 16, 17, 18, 22 и 23 июля 2018 года. Аномалии наблюдаются на фоне суточных колебаний размерностей алфавитов и ассоциируются с зафиксированными землетрясениями. Для сравнения на

рис. 6 приведен график динамики дисперсии сигнала ЭМЭ. Очевидно отсутствие энергетических признаков возникновения сейсмических событий.

В ходе анализа важно было получить оценку информативности потоков эмиссионных сигналов. Данную оценку возможно получить по результатам расчета относительной информационной энтропии полученных алфавитов. Пример графика информационной энтропии для данных ЭМЭ, зарегистрированных в июле 2018 г., показан на рис. 7. Значения показателя относительной информационной энтропии рассчитываются как отношение информационной энтропии на текущем 15-минутном интервале измерения к значению энтропии, получаемому если алфавит имеет равномерное распределение частоты появления символов по всей оси размерностей алфавита на том же временном интервале. Результаты расчета информационной энтропии показывают незначительное повышение уровня в период повышенной сейсмической активности в Камчатском регионе в июле 2018 г.

III. ЗАКЛЮЧЕНИЕ

Символическое описание импульсных сигналов позволяет перевести обработку и анализ из классического численного вычисления в область обработки и анализа кодовых последовательностей, делая доступным применение методов лингвистики, поиска информации и смысла в виде выделения скрытых правил и грамматических форм метаязыка.

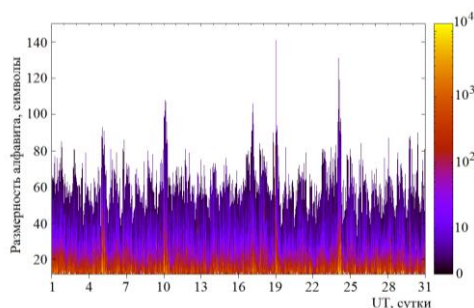


Рис. 5. 3D-представление алфавитов сигналов ЭМЭ, полученных в течение июля 2018 г. По оси x – даты измерений (шаг измерения 15 минут); по оси y – размерность алфавита; по оси z (шкала цвета) – частоты появления импульсов на 15-минутном интервале измерения

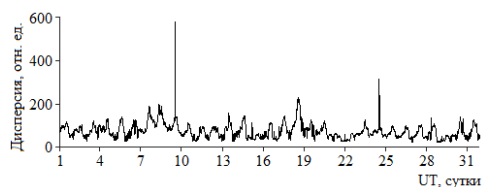


Рис. 6. Результаты расчета значений дисперсии сигнала ЭМЭ в июле 2018 г.

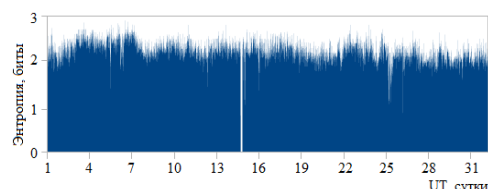


Рис. 7. Результаты расчета значений информационной энтропии алфавитов сигнала ЭМЭ в июле 2018 г.

Метод символического описания импульсов ГФС расширяет техническую базу инструментов предобработки данных для анализа состояния динамических систем на основе количественно-качественных характеристик сигналов. Как показали результаты экспериментальной проверки метода символического описания импульсных сигналов, для корректной работы алгоритмов необходимо обеспечить минимальное влияние шумовой помехи как фактора искажения первородных форм импульсов. Поэтому важно на первых этапах обработки обеспечить исключение ошибок второго рода, которые будут приводить к появлению ложных кластеров. Дальнейшие исследования предполагают разработку методов выявления аномального поведения эмиссионных ГФС, в том числе и попытку связать выявленные аномалии с сейсмическими событиями использованием в качестве базового инструмента аппаратов символического описания и энтропийного анализа с проверкой их достоверности и значимости.

СПИСОК ЛИТЕРАТУРЫ

- [1] Чебров В.Н., Салтыков В.А., Серафимова Ю.К. Прогнозирование землетрясений на Камчатке. М.: СветочПлюс, 2011. 304 с.
- [2] Соболев Г.А., Пономарев А.В. Физика землетрясений и предвестники. М.: Наука, 2003 г. 270 с.
- [3] Марапулец Ю.В., Шевцов Б.М. Мезомасштабная акустическая эмиссия. Владивосток: Дальнаука, 2012. 126 с.
- [4] Шуман В.Н., Богданов Ю.А. Электромагнитная эмиссия литосферы: пространственная структура и возможные механизмы генерации // Геофиз. журн. 2008. 30, № 6. С. 39-50.
- [5] Гордиенко В.А., Гордиенко Т.В., Краснописцев Н.В. Аномалия высокочастотных сигналов геоакустической эмиссии как оперативный предвестник землетрясения / Гордиенко В.А., Гордиенко Т.В., Краснописцев Н.В. [и др.] // Акустический журнал. 2008. Т. 54. № 1. С. 97-109.
- [6] Senkevich Yu.I , Duke V.A., Mishchenko M.A., Solodchuk A.A. Information approach to the analysis of acoustic and electromagnetic signals // E3S Web of Conferences. 2017. Vol. 20. 02012. 9 p.
- [7] Serra J. Image Analysis and Mathematical Morphology. – Academic Press, 1982.
- [8] Пытьев Ю.П. Задачи морфологического анализа изображений. В сб. «Математические методы исследования природных ресурсов Земли из Космоса». М.: Наука, 1984.
- [9] Николис Г., Пригожин И. Познание сложного. М.: Мир, 1990. 512 с.
- [10] Сенкевич, Ю.И., Марапулец, Ю.В., Луковенкова, О.О., Солодчук, А.А. Методика выделения информативных признаков в сигналах геоакустической эмиссии. // Труды СПИИРАН, 2019, 18(5), С. 1066-1092.