

УДК 534.4

В. Е. Гай

Нижегородский государственный технический университет им. Р.Е. Алексеева

Информационный подход к описанию звукового сигнала

Рассматривается подход к решению задачи раскрытия априорной неопределённости звукового сигнала. Предлагаемый подход основан на последовательном применении к сигналу операций интегрирования и дифференцирования. Это позволяет избавиться от некорректности, возникающей при предварительной обработке сигнала. Метод идентификации звуковых записей, предлагаемый в работе, основан на использовании данного подхода.

Ключевые слова: цифровая обработка сигналов, звуковой сигнал, активное восприятие, идентификация звуковых сигналов.

Введение

Следуя теории активного восприятия [1], введём понятие звукового сигнала. Звуковым сигналом называется множество M , каждый элемент которого в фиксированный момент времени t есть неотрицательная действительная функция действительного аргумента:

$$M = \begin{cases} f(t), & \text{если } t \in \bar{T} \subset R, \\ 0 & \text{— в противном случае,} \end{cases}$$

определённая на конечном множестве точек замкнутого одномерного интервала \bar{T} , суммируемая: $\int_{t \in \bar{T}} f(t) dt < \infty$ и квадратично интегрируемая: $\int_{t \in \bar{T}} f(t)^2 dt < \infty$.

Разработка оптимальных методов цифровой обработки звуковых сигналов невозможна без создания подхода к описанию сигнала. Звуковой сигнал представляет собой физический процесс, принадлежащий окружающему миру, вследствие этого он обладает свойствами и законами окружающего мира. К таким свойствам относятся упорядоченность и структурированность. Условие упорядоченности обеспечивает существование бинарного отношения частичного порядка типа

$$(a \leq b) = \begin{cases} a < b & \text{— строгий порядок,} \\ a = b & \text{— тождество.} \end{cases}$$

Условие структурированности, с учётом упорядоченности, указывает на наличие для любых пар элементов (a, b) существование границ — максимальной и минимальной. Учитывая, что о наблюдаемом звуковом сигнале в заданном интервале отсутствуют какие-либо априорные данные, последующая обработка такого сигнала выполняется в условиях априорной неопределённости. Рассматривая звуковой сигнал как системное образование, задача раскрытия априорной неопределённости сигнала заключается в выделении структурных элементов сигнала и определении связей между ними.

1. Предварительная обработка сигнала

Система обработки данных (СОД) с точки зрения системного анализа выполняет три этапа обработки данных (см. рис. 1): подготовку и анализ данных, принятие решения. Система распознавания образов является одним из примеров СОД.

На этапе создания исходного описания в классических системах распознавания решается задача предварительной обработки анализируемого сигнала, которая заключается в

фильтрации сигнала. Данную операцию можно описать уравнением Фредгольма первого рода:

$$g(x) = \int_{t \in \bar{T}} K(x, t) f(t) dt,$$

где $g(x)$ – наблюдаемая функция, $f(t)$ – неизвестная функция (априори неизвестное изменение во времени амплитуды регистрируемого сигнала на входе устройства), $K(x, t)$ – ядро преобразования.

Фильтрация сигнала выполняется с целью компенсации помех, которые искажают сигнал в процессе его передачи от источника к приёмнику.

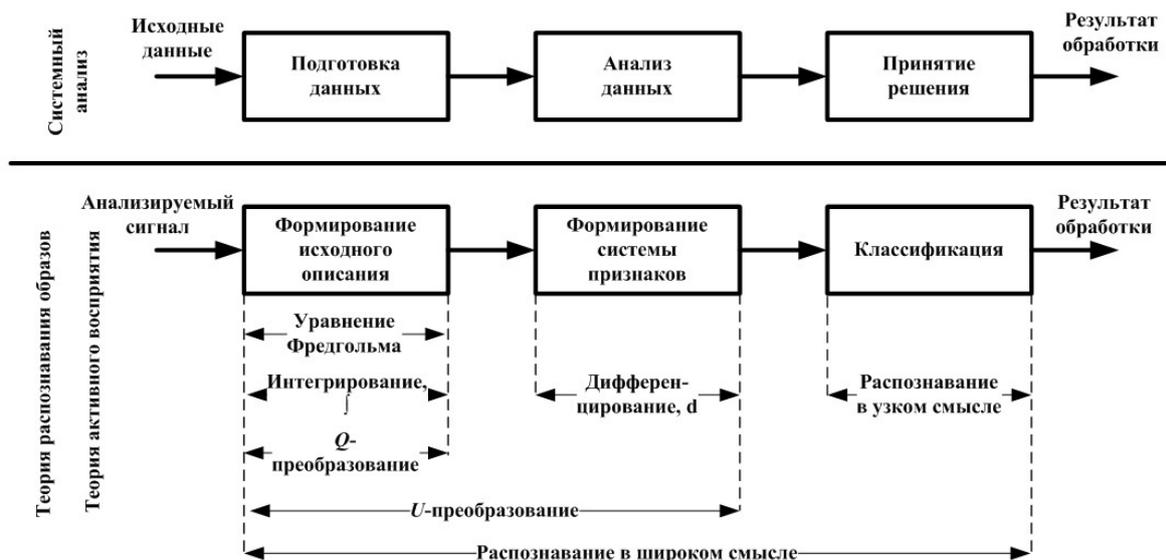


Рис. 1. Этапы обработки информации

Учитывая, что незначительные изменения в наблюдаемом сигнале $f(t)$ могут привести к недопустимо большим изменениям в решении, результат, получаемый после выполнения фильтрации, в рамках использования уравнения Фредгольма является некорректным. Таким образом, сама задача предварительной обработки, описываемая с помощью уравнения Фредгольма, также некорректна [2]. Следовательно, основной задачей на шаге предварительной обработки является избавление от некорректности, возникающей в результате вычисления преобразования Фредгольма.

2. U -преобразование

Методам решения некорректных задач посвящено большое количество работ. К численным методам решения некорректных задач относятся методы подбора и методы регуляризации, к формализованным – метод Галёркина и проекционные методы [2]. Один из способов приведения некорректной задачи к корректной описан в Теории активного восприятия (ТАВ) [1].

В данной теории этап формирования исходного описания сигнала соответствует Q -преобразованию. Q -преобразование выполняется в соответствии с уравнением Фредгольма и в качестве ядра преобразования $K(t, x)$ использует единичную весовую матрицу. В результате уравнение Фредгольма приводится к следующему виду:

$$g(x) = \int_{t \in \bar{T}} f(t) dt.$$

Уравнение данного вида является корректным, вследствие того, что из него исключена операция дифференцирования, а операция интегрирования является корректной [3].

Применение операции интегрирования к сигналу $f(t)$:

$$m(\bar{T}) = \int_{t \in \bar{T}} f(t) dt$$

позволяет получить представление о сигнале в целом:

1) если результат интегрирования ($m(\bar{T})$) отличен от нуля, то рассматриваемый сигнал содержит информацию и можно продолжить исследование структуры сигнала;

2) если результат интегрирования равен нулю, дальнейший анализ сигнала не имеет смысла.

С позиций физики Q -преобразование выявляет массу в области определения, для которой $f(t)$ – это плотность вещества. Таким образом, Q -преобразование за одно измерение позволяет получить об анализируемом сигнале один бит информации и выявить структурный элемент сигнала.

Поскольку все отсчёты сигнала находятся в отношении эквивалентности, то множество отсчётов можно разбить на любое число подобластей без пересечения этих областей между собой. Последовательное применение операции дихотомии и Q -преобразования к сигналу позволяет сгенерировать пирамидальную структуру (см. рис. 2).

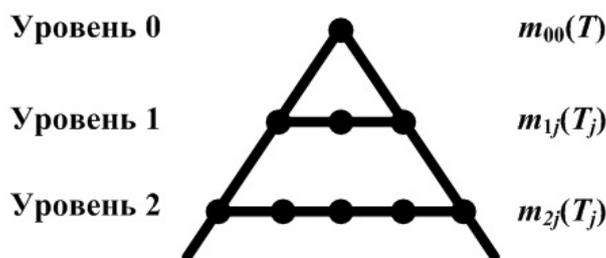


Рис. 2. Пирамида описания сигнала

Результатом формирования пирамидального описания сигнала является множество структурных точек $\{m_{ij}\}$, где i – уровень анализа, j – номер подобласти сигнала. Для выявления бинарных отношений (связей) на множестве структурных элементов необходима и достаточна операция вычитания. Таким образом, в рамках теории активного восприятия звуковой сигнал рассматривается как системное образование. Для обнаружения системных элементов используется интегральное преобразование, а для выявления связей между элементами – пространственное дифференцирование. Результатом выявления дифференциальной структуры является спектральное описание сигнала. Преобразования интегрирования и дифференцирования образуют композицию, которая называется U -преобразованием: $U = d \cdot \int$.

3. Формирование описания сигнала

Рассмотрим задачу определения необходимого и достаточного числа дихотомий наблюдаемого звукового сигнала с учётом построенной пирамиды исходных описаний, начиная с её вершины (см. рис. 2):

1) при нуле дихотомий вычисляется масса по всей области определения $m_{00}(T)$, если она не равна нулю ($m_{00}(T) \neq 0$), то имеем достоверное событие наличия сигнала на интервале наблюдения;

2) дихотомия интервала пополам позволяет выполнить локализацию положения экстремума на анализируемом интервале относительно центра: слева либо справа. Каждой половине интервала соответствует масса m_{11}, m_{12} (см. рис. 3а) [4]. Отношение между этими массами имеет следующий вид: $\mu = m_{11} - m_{12}$. Рассмотрим возможные варианты значения μ :

а) $\mu = 0$: сигнал симметричен относительно центра, следовательно, на данном уровне локализовать положение экстремума нельзя и следует перейти на следующий (более точный) уровень;

б) $\mu < 0$: экстремум расположен слева от центра интервала;

в) $\mu > 0$: экстремум расположен справа от центра интервала;

таким образом, дихотомия интервала пополам позволяет проверить гипотезу об однородности (симметрии) сигнала;

3) дихотомия отрезка на четыре равных интервала (уточнение положения экстремума, см. рис. 3б), поставим в соответствие каждому из интервалов массу: $m_{21}, m_{22}, m_{23}, m_{24}$. Отношение между массами имеет следующий вид: $\mu_1 = (m_{22} + m_{23}) - (m_{21} + m_{24})$ и $\mu_2 = (m_{21} + m_{23}) - (m_{22} + m_{24})$. Анализ значений μ_1 и μ_2 позволяет определить положение экстремума: в трёх случаях получаем, что на сигнале находилось два экстремума (т. е. один экстремум на нулевом уровне разделился на два экстремума на втором уровне), в одном случае (два шарика по центру) получаем, что для локализации экстремума требуется переместиться на уровень ниже (выполнить более тонкий анализ).

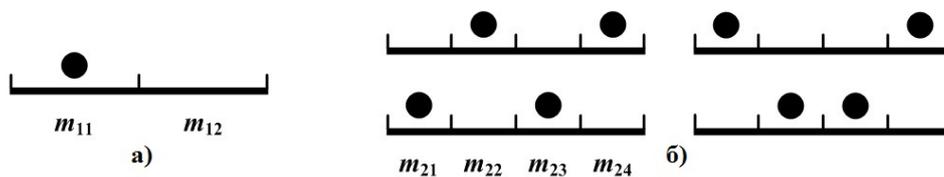


Рис. 3. Дихотомия отрезка: а) пополам; б) на четыре части

Таким образом, получаем четыре базисных одномерных фильтра-покрытия (F_0, F_1, F_2, F_3), которые позволяют выделить отношения между парой масс (см. рис. 4).

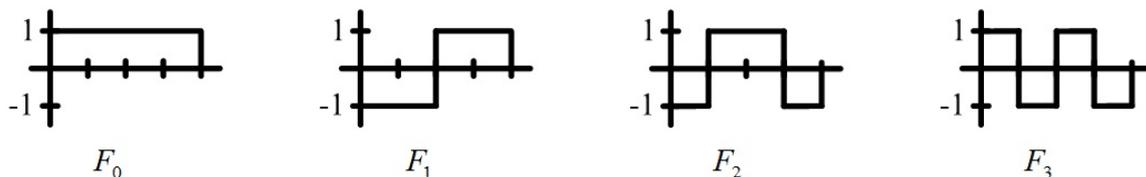


Рис. 4. Базисные функции

Исследования показали, что для целей анализа звукового сигнала необходимо использование 16 фильтров, которые можно построить по аналогии с четырьмя фильтрами. Результатом имеем систему базисных одномерных булевых функций, которая внешне похожа на систему базисных функций Уолша системы Хармута. Специфика использования данной системы заключается в том, что она применяется после реализации Q -преобразования и сразу по всей области определения.

4. Алгебра групп

Теория активного восприятия не ограничивается только формированием спектрального представления сигнала [1]. В ТАВ входит раздел «Алгебра групп», посвящённый анализу зависимостей между спектральными коэффициентами разложения. Обнаруженные зависимости допускают своё использование на этапах принятия решения и понимания анализируемого звукового сигнала.

Примем для множества фильтров $\{F_i\}$, что $(+1 \rightarrow 1)$ и $(-1 \rightarrow 0)$. Получим множество бинарных операторов $\{V_i\}$. Для этих операторов допустимы теоретико-множественные операции объединения (сложения) и пересечения (умножения). Результатом имеем алгебру $A_V = \langle \{V_i\} : +, \times \rangle$ [1]. В алгебре A_V существуют алгебраические группы:

1) P_{ni} – группы на трёх элементах (названы полными), образованы на тройках операторов (V_i, V_j, V_k) , для которых справедливы соотношения: $V_i + V_j + V_k \equiv e_1$ – единица; $V_i V_j V_k$ – образ, формируемый с помощью операции умножения и описание группы P_{ni} ;

2) P_{si} – группы на четырёх элементах (названы замкнутыми), образованы на четвёрке операторов (V_i, V_j, V_n, V_m) , где $(V_i, V_j, V_k) \in P_{ni}$, $(V_n, V_m, V_k) \in P_{nj}$, с описанием $V_i V_j + V_n \overline{V_m}$ и единицей – $V_i + V_j + V_n + \overline{V_m} \equiv e_1$.

Алгебра групп используется для формирования описания звукового сигнала. С помощью замкнутых и полных групп выполняется спектрально-корреляционный анализ. Полные группы позволяют выявить корреляционные связи между операторами. Замкнутые – корреляционные связи между полными группами.

5. Идентификация звукового сигнала

Количество музыкальных композиций, хранящихся в настоящее время в сети Интернет, велико (например, сервис Яндекс.Музыка хранит около пяти миллионов записей, сервис Shazam – пять миллиардов). Очевидно, что в такой ситуации актуальна задача быстрого и точного поиска по имеющимся музыкальным записям. Рассмотрим алгоритмы, предлагаемые для формирования цифрового отпечатка звукового сигнала и для поиска сигнала в базе данных.

5.1. Формирование отпечатка сигнала

Пусть $f(t)$ – звуковой сигнал, наблюдаемый на конечном отрезке времени. Результат применения U -преобразования к сигналу f – многоуровневое спектральное представление $D = \{d_{ij}\}$, $i = \overline{1, K}$, $j \in \overline{1, M_i}$, где K – число уровней разложения, M_i – количество сегментов сигнала на i -м уровне разложения, d_{ij} – спектр, включающий L спектральных коэффициентов (число используемых фильтров), $d_{ij}\{k\}$ – k -й спектральный коэффициент ($k = \overline{1, L}$), f_{ij} – сегмент сигнала f , по которому вычислен спектр d_{ij} .

Рассмотрим алгоритм формирования отпечатка звукового сигнала:

1) построение на основе U -преобразования спектрального представления D сигнала f (при построении разложения сегменты f_{ij} не перекрываются); в предлагаемом алгоритме вычисляется только один уровень U -преобразования, в связи с этим значение индекса i для краткости записи не приводится;

2) формирование огрубленного представления $C = \{c_j\}$ спектрального представления $D = \{d_j\}$ (степень огрубления – параметр k) [5];

3) выбор ключевых сегментов сигнала f :

$$\begin{aligned} &\forall j \in \overline{1, M}, \\ &\forall t \in \overline{1, L}, \\ &\text{если } |c_j(t)| \leq k \text{ или } |c_j(t)| \geq l \text{ (} k < l \text{), то} \\ &\quad K_f.data[j] = f_j; \\ &\quad K_f.shift[j] = (j - 1) \cdot |f_j| + 1, \end{aligned}$$

где K_f – набор ключевых сегментов сигнала f , $|f_j|$ – количество отсчётов в сегменте f_j , $K_f.shift[j]$ – смещение j -го сегмента относительно начала сигнала.

Ключевой сегмент соответствует участку сигнала, на котором максимален отклик небольшого количества операторов (одного-двух). Это означает, что сегмент сигнала подпадает под образ операторов, давших максимальные отклики;

4) вычисление по каждому ключевому сегменту замкнутых групп и отбор из полученного набора первых N по массе групп:

$$\begin{aligned} &\forall j \in \overline{1, |K_f|}, \\ &W_f.groups[j] = GRP[K_f[j]]; \\ &W_f.diff[j] = MAX[W_f.groups[j]] - MIN[W_f.groups[j]]; \\ &W_f.shift[j] = K_f[j].shift, \end{aligned}$$

где $|K_f|$ – мощность множества K_f , $GRP[\bullet]$ – оператор, который вычисляет замкнутые группы по сегменту сигнала, $\max[\bullet]$ ($\min[\bullet]$) – оператор для поиска среди групп набора, $W_f.groups[j]$ – группы с максимальной (минимальной) массой, j – номер ключевого сегмента, W_f – предварительное описание сигнала f , $W_f.diff[j]$ – разница между максимальным и минимальным значением масс групп в $W_f.groups[j]$, $W_f.shift[j]$ – смещение j -го ключевого сегмента относительно начала сигнала;

5) отбор устойчивых сегментов сигнала (эксперименты показали, что устойчивость сегмента сигнала к искажениям тем лучше, чем больше разница между максимальной и минимальной массами групп, входящих в описание сигнала):

$$\begin{aligned} \forall j \in \overline{1, |M_f|}, \\ \text{если } M_f.diff[j] \geq T, \\ M_f.groups[k] = W_f.groups[j]; \\ M_f.shift[k] = W_f.shift[j]; \\ k = k + 1, \end{aligned}$$

где M_f – описание сигнала f , содержащее только устойчивые сегменты сигнала, T – значение порога.

5.2. Поиск отрывка звукового сигнала в базе данных

Допустим, с помощью алгоритма, приведённого в пункте 5.1, сформирована база данных описаний звуковых сигналов. Рассмотрим алгоритм поиска звукового сигнала в базе данных:

1) формирование отпечатка M_s искомого сигнала s (в отличие от алгоритма, описанного в пункте 5.1, шаг отбора устойчивых сегментов сигнала пропускается, сегменты, на которые разбивается сигнал, перекрываются);

2) поиск отпечатка M_s в базе данных заключается в сравнении пар элементов в описании i -го сигнала базы данных и сигнала s , которые находятся на одинаковых расстояниях: в случае, если такие пары найдены, вычисляется расстояние между их описаниями:

$$\begin{aligned} d_1 &= CMP[M_{f,i}.groups[k], M_s.groups[m]], \\ d_2 &= CMP[M_{f,i}.groups[l], M_s.groups[n]], \end{aligned}$$

где $CMP[\bullet]$ – оператор сравнения, который вычисляет количество одинаковых групп, в сравниваемых наборах групп ($k < l$, $m < n$);

3) если $d_1 > T$ и T – порог, то в массив R сохраняется абсолютное значение расстояния между элементами $M_{f,i}.groups[k]$ и $M_s.groups[m]$:

$$R[p] = |M_{f,i}.shift[k] - l|.$$

После окончания сравнения описаний i -го из базы данных и искомого сигнала s по массиву R формируется гистограмма. Гистограмма сигнала, который включает искомым сигнал s , будет иметь ярко выраженный максимум [6].

Заключение

В работе рассматривается информационный подход к описанию звукового сигнала, вводится понятие звукового сигнала как системного образования, описывается метод обработки сигнала в условиях априорной неопределённости. Данный метод заключается в последовательном применении к анализируемому сигналу операций интегрирования и дифференцирования, что позволяет выделить структурные элементы сигнала и определить связи между ними. Достоинствами рассматриваемого метода формирования описания сигнала

является низкая вычислительная сложность и простота реализации. В качестве примера описанного подхода к обработке звуковых сигналов рассматривается решение задачи идентификации звуковых записей.

Работа выполнена при поддержке Министерства образования и науки РФ в рамках договора № 02.G25.31.0061 от 12 февраля 2013 года (в соответствии с Постановлением Правительства Российской Федерации от 9 апреля 2010 г. № 218).

Работа, отраженная в данной статье, признана лучшей на Всероссийском конкурсе научных и инновационных проектов студентов, аспирантов и молодых ученых (МФТИ-2012).

Литература

1. *Утробин В.А.* Физические интерпретации элементов алгебры изображения // *Успехи физических наук.* — 2004. — Т. 174, № 10. — С. 1089–1104.
2. *Бертено М., Поджо Т.А., Торре В.* Некорректные задачи в предварительной обработке визуальной информации // *ТИИЭР.* — 1988. — Т. 76, № 8. — С. 17–40.
3. *Кабанихин С.И.* Обратные и некорректные задачи. — Новосибирск: Сибирское научное издательство, 2009.
4. *Бир С.* Кибернетика и управление производством. — М.: Наука, 1965.
5. *Gai V.E.* Signal comparison algorithm in terms of a priory uncertainty // *Proceedings of 8th Open German-Russian Workshop «Pattern recognition and Image understanding».* — 2011. — P. 75–78.
6. *Wang A.* The Shazam music recognition service // *Communications of the ACM.* — 2006. — V. 49, N. 8. — P. 44–48.

Поступила в редакцию 10.04.2013.