

Информационная модель слухового восприятия

В. Е. Гай,* Н. В. Гай,† О. А. Яковлев,‡ В. Г. Сажин§

*Нижегородский государственный технический университет имени Р. Е. Алексеева,
Институт радиоэлектроники и информационных технологий,
кафедра Вычислительных систем и технологий,
Россия, 603950, Нижний Новгород, ул. К. Минина, д. 24
(Статья поступила 09.11.2014; Подписана в печать 28.11.2014)*

В работе описывается предлагаемая модель информационных преобразований, проходящих в слуховой системе. Проводятся аналогии между процессами преобразования информации в слуховой системе и предлагаемой моделью. Указанная модель строится на принципах теории активного восприятия (ТАВ), предложенной для моделирования информационных процессов в системе зрительного восприятия. На основе предложенной модели создано несколько методов обработки звуковых сигналов: метод вычисления частоты основного тона голоса человека, метод выделения полезного сигнала. Проведённые экспериментальные исследования подтвердили эффективность предложенных методов.

PACS: 43.60.-c

УДК: 681.391, 577.354

Ключевые слова: моделирование слуховой системы, обработка сигналов.

1. ОСНОВНЫЕ ПОНЯТИЯ И ПОСТАНОВКА ЗАДАЧИ

Человек, как и любое другое живое существо, для выживания в окружающем мире наделён сенсорной системой. Одна из компонент сенсорной системы — слуховая система, предназначенная для восприятия звуковых колебаний.

Аналогично понятию «зрение», введённого Д. Марром [1], попробуем описать понятие «слух». Слух — процесс порождения по звуковым колебаниям окружающей среды (воздушной или водной) описания, полезного для наблюдателя и не перегруженного несущественной информацией.

Цель работы — создание модели информационных преобразований, проходящих в слуховой системе для построения эффективных алгоритмов обработки звуковых сигналов. Указанная модель будет построена на принципах теории активного восприятия (ТАВ) [2], предложенной для моделирования информационных процессов в системе зрительного восприятия.

2. КРАТКИЕ СВЕДЕНИЯ О ФИЗИОЛОГИИ СЛУХОВОЙ СИСТЕМЫ

Слуховой анализатор — совокупность соматических, рецепторных и нервных структур, деятельность которых обеспечивает восприятие звуковых колебаний [3]. Слуховой анализатор позволяет слышать человеку звук в пределах от 16 Гц до 22 кГц при передаче колебаний по воздуху, и до 220 кГц при передаче зву-

ка по костям черепа. Слуховой анализатор состоит из уха (периферическая часть), слухового нерва и сложной системы нервных связей и центров мозга (высшие отделы).

Ухо состоит из наружного, среднего и внутреннего уха. Наружное ухо включает в себя: слуховой канал и ушную раковину, среднее ухо: барабанную перепонку, овальное окно, молоточек, наковальню и стремечко, внутреннее ухо: улитку и нервы.

3. МОДЕЛИ СЛУХОВОГО АНАЛИЗАТОРА

А. Модель Н.В. Позина

Приведённая в [4] модель описывает в виде аналогового устройства, преобразования сигналов, выполняемые в среднем и внутреннем ухе (см. рис. 1).

По данным физиологии среднее ухо имеет свойства фильтра низких частот, поэтому в первом приближении среднее ухо можно рассматривать в качестве интегратора. Из рис. 1 следует, что среднее ухо моделируется соединением апериодического и колебательного звеньев.

Основная мембрана в модели представлена в виде набора частоточувствительных элементов, упорядоченно расположенных вдоль оси частот.

В работе [5] для моделирования функционирования базилярной мембраны, входящей в состав улитки, используется набор частотных фильтров, совмещённых с устройствами, вычисляющими вторую разность.

Существует различные модели периферического отдела слуховой системы, в которых указывается, что среднее ухо можно смоделировать в виде фильтра нижних частот [4, 6], а базилярную мембрану — в виде набора полосовых фильтров [4, 7].

*E-mail: iamuser@inbox.ru

†E-mail: natalja.v.gai@gmail.com

‡E-mail: yakovlev.oleg2010@gmail.com

§E-mail: darthy.first@gmail.com

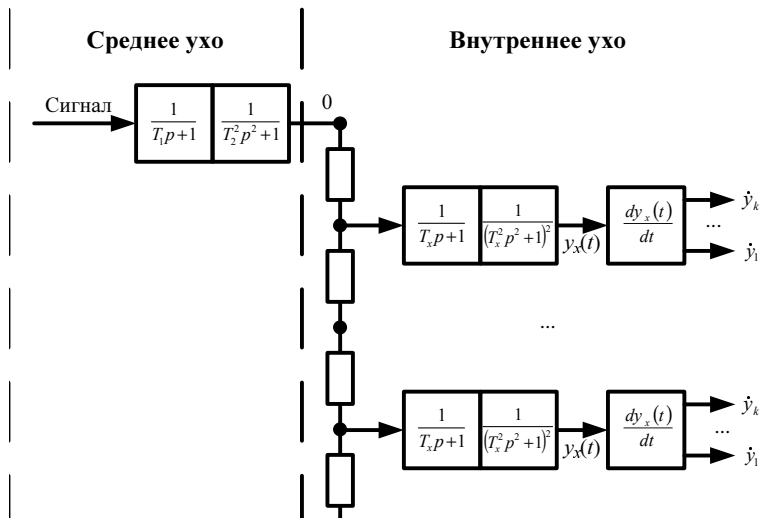


Рис. 1: Модель среднего уха с элементами основной мембраны

В. Модель Л. А. Чистович

В работах [8–10] описывается модель обработки амплитудной огибающей в частотных каналах слухового спектрального анализатора (модель обнаружения амплитудных неравномерностей). Основные положения модели заключаются в следующем:

- происходит слуховая обработка не суммарной огибающей исходного сигнала, а огибающих сигнала с выходов отдельных каналов слухового спектрального анализатора, связанных с достаточно коротким участком базилярной мембраны улитки;
- на вход модели поступают выходы 128 частотных каналов анализатора; основу модели составляет набор трёх типов полосовых фильтров с центральными частотами 7, 25 и 75 Гц;
- функционально канал обработки огибающей в частотном канале соответствует преобразованию сигналов в нейронах волокон слухового нерва;
- в каждом канале обработки огибающей происходит нелинейное преобразование амплитуды огибающей типа логарифмирования, причём на пороговых уровнях характеристика линейна, затем осуществляется низкочастотная фильтрация, обеспечивающая выделение изменений амплитуды огибающей во времени; при этом в каждом канале функционируют несколько параллельных полосовых фильтров, соответствующих различным по качеству восприятия изменениям огибающей сигнала: «слоговой», «г»-образных звуков, «хриплости»;
- выходные сигналы фильтров разделяются по знаку и информация о «положительных» и «отри-

цательных» изменениях амплитуды огибающей некоторым образом объединяется;

- в каждом частотном канале существуют детекторы, обнаруживающие локальные во времени изменения амплитуды сигнала.

С. Модель слухового анализатора на основе ТАВ

Принимая во внимание описанные модели, сформулируем основные положения моделирования процессов обработки звукового сигнала, в рамках ТАВ:

- среднее ухо представляет собой интегрирующий коммутатор и в предлагаемой модели используется для выделения сигнала длительности T для дальнейшего анализа на мембране;
- мембрана улитки внутреннего уха моделируется с помощью U -преобразования (см. рис. 2) и описывается с помощью двух преобразований: интегрирования (Q -преобразование) и дифференцирования; данные преобразования реализуются с помощью двух математических операций: сложения и вычитания; в рамках модели, результат работы анализатора — грубо-точное дерево разложения исходного сигнала. Таким образом, в предлагаемой модели на этапе анализа формируется спектральное представление сигнала, т. е. разложение его по системе фильтров.

В [11] отмечается, что Q -преобразованию можно поставить в соответствие тонические нейроны, а преобразованию дифференцирования — фазические. В предлагаемой модели можно провести такую же аналогию. Из нейрофизиологических экспериментов известно [10, 12], что существует два типа реакций центральных слуховых нейронов: фазическая и тониче-

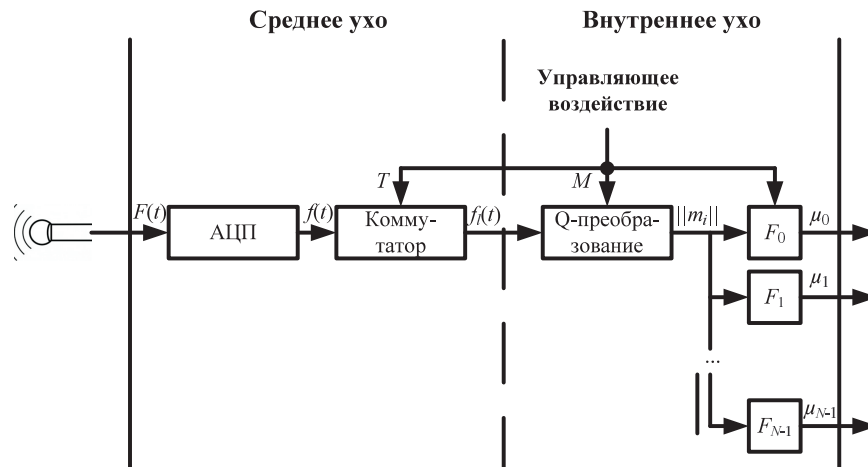


Рис. 2: Модель среднего и внутреннего уха

ская. Фазические нейроны чувствительны к изменению акустических свойств сигнала, т. е. они выделяют акустические события (именно для этого и предназначено преобразование дифференцирования, для выделения структуры сигнала). Обнаружены фазические нейроны, которые избирательно откликаются на изменение частоты тонального сигнала: одни из них возбуждаются при наличии изменения, другие чувствительны только к понижению или повышению частоты. Тонические нейроны характеризуются непрерывным откликом на звуковой сигнал. Обнаружены тонические нейроны, которые точно реагируют на периодичность гласного: их импульсные реакции имеют фазовую привязку к каждому периоду основного тона [13]. Обнаружено, что у лягушки тонические нейроны имеют тенденцию к воспроизведению огибающей сложных звуковых сигналов [14].

Параметры модели: T — длина анализируемого звукового сигнала, M — число уровней разложения, N — число используемых фильтров, $fl(t)$ — сигнал, получаемый на выходе коммутатора, $\{F_i\}$ — множество фильтров, моделирующих мембрану улитки, $\{m_i\}$ — вектор масс (в [14] названы визуальными массами), результат выполнения Q -преобразования, $\{\mu_i\}$ — множество чисел, полученных в результате U -преобразования, $F(t)$ — аналоговый звуковой сигнал.

Пусть $f(t)$ — анализируемый звуковой сигнал (цифровой), наблюдаемый на конечном отрезке времени. Результат применения U -преобразования к сигналу f — многоуровневое спектральное представление $D = \{d_{ij}\}$ (грубо-точное дерево разложения сигнала), $i = \overline{1, M}$, $j = \overline{1, K_i}$, где M — число уровней разложения, K_i — количество сегментов сигнала на i -ом уровне разложения, d_{ij} — спектр, включающий N спектральных коэффициентов (число используемых фильтров), $d_{ij}\{k\}$ — k -ый спектральный коэффициент ($k = \overline{1, N}$), f_{ij} — сегмент сигнала f , по которому вычислен спектр d_{ij} .

В рамках ТАВ 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.

В работе [15] отмечается, что в слуховой системе достоверно реализуются следующие математические операции: сложение, вычитание, интегрирование по времени, а в работе [16] высказывается гипотеза о том, что слуховой аппарат реагирует не только на мгновенные значения сигнала, но и на его функциональные преобразования, например, на первую и вторую производные [17]. Эти данные, а также данные, приведённые в [4] согласуются с предложенной моделью.

Теория активного восприятия не ограничивается только формированием спектрального представления сигнала. 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. Методы создания групп описаны в [18–21].

В [7] указывается, что в слуховой системе в переднем улитковом (кохлеарном) ядре и заднем улитковом (кохлеарном) ядре находятся мультиполярные нейроны, которые соединяются с несколькими волокнами слухового нерва. Данная структура подобна архитектуре построения полных и замкнутых групп.

Отметим, что число нейронов, в различных структурах слуховой системе, при движении к слуховой коре, увеличивается [4]. Такую же закономерность можно наблюдать и в ТАВ при движении от операторов к замкнутым группам.

4. ПОСТРОЕНИЕ МЕТОДОВ ОБРАБОТКИ ЗВУКОВОГО СИГНАЛА НА ОСНОВЕ ПРЕДЛОЖЕННОЙ МОДЕЛИ

Рассмотрим методы обработки звуковых сигналов, основанные на предложенной модели. Отметим, что все алгоритмы обработки выполняют анализ сигнала в рамках его сформированного грубо-точного представления.

А. Вычисление огибающей сигнала

В предлагаемой модели решение задачи выделения огибающей сигнала выполняется на начальных уровнях дерева разложения сигнала. Звуковой сигнал в ТАВ рассматривается как системное образование. Для обнаружения системных элементов в ТАВ используется интегральное преобразование, а для выявления связей между элементами — пространственное дифференцирование. U -преобразование, представляющее собой композицию преобразований интегрирования и дифференцирования, реализуется с помощью системы фильтров. В результате проведённых исследований установлены свойства спектров, относящихся к полезному сигналу, а также паузе и шуму:

- спектр полезного сигнала содержит элементы, значительно отличающиеся друг от друга по величине;
- спектр помехи содержит элементы близкие друг к другу по значению (элементы спектра сигнала, представляющего собой паузу, равны нулю).

В связи с этим, метод выделения огибающей заключается в вычислении среднеквадратического отклонения первых трёх спектральных коэффициентов, находящихся на i -ом уровне разложения:

$$E_j = STD[d_{ij}\{1\}, d_{ij}\{2\}, d_{ij}\{3\}], \quad j = \overline{1, 2^{i-1}}, \quad (1)$$

где $STD[\bullet]$ — оператор вычисления среднеквадратического отклонения.

Существуют исследования, в которых указывается, что выделение огибающей сигнала — одна из функций слуховой системы:

- в [22] на основе нейрофизиологических данных предполагается, что выделение огибающей звукового сигнала выполняется на уровне кохлеарных ядер;
- в [23] указывается, что в слуховой системе существуют нейроны, которые кодируют не форму звуковой волны, а некоторое её функциональное преобразование, например, огибающую.

В. Выделение полезного сигнала

Алгоритм выделения полезного сигнала в качестве исходных данных использует результат выделения огибающей E сигнала $f(t)$:

- вычисление многоуровневого разложения d_E огибающей E (при формировании разложения используется только фильтр F_0);
- формирование множества сигналов S из коэффициентов разложения, находящихся на i -ом уровне ($i \geq 3$):

$$S = \{s_i\}, \quad s_{ij} = d_{ij}\{0\}, \quad i = \overline{3, M}, \quad j = \overline{1, 2^{i-1}}; \quad (2)$$

- кластеризация каждого сигнала, входящего в множество S и выбор карты сегментации на каждом уровне (число кластеров зависит от уровня разрешения, при сегментации участки с большей амплитудой принимаются за полезный сигнал);
- объединение результатов сегментации на нескольких уровнях.

Известно, что на нижних уровнях слуховой системы, информация кодируется с большим разрешением, чем на верхних [24]. Верхние участки слухового пути отвечают за обработку более сегментированной информации, чем низшие. Предложенный метод выделения полезного сигнала не противоречит описанному принципу обработки информации в слуховой системе, так как выделение полезного сигнала происходит на первых этапах обработки принятого сигнала и на низких уровнях разрешения.

С. Оценка частоты основного тона

Рассмотрим модели, описывающие процесс оценки частоты основного тона в слуховой системе и сравним их с предлагаемым методом.

В [4] на основе экспериментальных данных предлагается модель, описывающая процесс определения слуховой системой высоты звуков посредством временного анализа:

- фильтрация с помощью слухового частотного анализатора входного сигнала;
- кодировка временных особенностей формы волны сигнала в каждом частотном канале;
- сигнал, представленный в импульсной форме, подаётся на блок предварительной оценки высоты; по каждому частотному каналу происходит оценка временного периода, соответствующего высоте звука;

- по совокупности предварительных оценок периода происходит окончательное определение высоты звука.

Предлагается подробная реализация описанной модели оценки высоты звука [4].

Из экспериментальных данных [4] известно, что информация о частоте заключена не только в пространственной картине возбуждения нейронов, но также и во временной организации последовательности импульсов у каждого нейрона. Таким образом, при оценке высоты звука выполняется анализ сигнала во времени.

В [7] рассматривается два класса моделей, используемых при оценке частоты основного тона: спектральная модель, основанная на сопоставлении с образцом (эталоном), и временная модель, основанная на автокорреляции.

Модель, основанная на сопоставлении с образцом [25] предполагает, что основной тон воспринимается даже тогда, когда фундаментальная частота отсутствует в спектре. Восприятие основного тона выполняется за счёт того, что в спектре присутствуют гармоники, связанные с фундаментальной частотой.

Недостатки модели, основанной на автокорреляции, сводятся к тому, что нейрофизиологам не удалось подтвердить наличие линии задержки в слуховой системе.

Оба типа моделей используют результаты частотного анализа, выполняемого на периферии слуховой системы. Отмечается, что использование спектральной модели, основанные на принципе не нашли обоснование с точки зрения нейрофизиологии. Предполагается, что основные операции по определению частоты основного тона выполняются в нижнем двухолмии.

Метод оценки частоты основного тона, основанный на предлагаемой модели, заключается в следующем [26]:

- анализируемый сигнал, по которому вычисляется высота звука, разбивается на сегменты;
- вычисляется спектр каждого сегмента;

- определяются характерные (ключевые) сегменты;
- по каждому характерному спектру вычисляются замкнутые группы, его описывающие;
- вычисляется расстояние между спектрами, описывающими похожие группы.

Результат работы алгоритма — наиболее часто встречающееся расстояние, и есть оценка высоты звука (частоты основного тона).

Сравнивая приведённые методы и модели оценки высоты звука необходимо отметить их подобие:

- в предлагаемом методе первый и второй шаг обработки сигнала реализуется за счёт использования U -преобразования, которое также, как и метод, описанный в [4], позволяет выполнить частотное разложение сигнала и закодировать его временные особенности (области);
- предлагаемый метод подобен автокорреляционному методу оценки частоты основного тона, однако за счёт того, что анализ зависимостей между данными выполняется не во временной области, а в частотной, к тому же по нескольким каналам (за счёт использования замкнутых групп), достигается высокая помехоустойчивость.

ЗАКЛЮЧЕНИЕ

Рассмотрены вопросы информационных преобразований, реализуемых системой слухового восприятия. Приводятся аналогии предложенной модели информационных преобразований с реальными процессами, протекающими в слуховой системе. Приводятся примеры использования модели в практике обработки звуковых сигналов.

-
- [1] *Март Д.* Зрение. Информационный подход к изучению представления и обработки зрительных образов. (М.: Радио и связь, 1987).
- [2] *Утробин В. А.* УФН. **174**, №. 10. С. 1089. (2004).
- [3] Словарь. 3-е изд. (Екатеринбург: Академический проект, Деловая книга. Серия: Gaudeamus. 2005).
- [4] *Позин Н. В.* Элементы теории биологических анализаторов. (М.: Наука, 1978).
- [5] *Лабутин В. К., Молчанов А. П.* Модели механизмов слуха. (М.: Энергия, 1973).
- [6] *Aage R.* Moller Hearing: its physiology and pathophysiology. (Academic Press, San Diego, 2000).
- [7] *Meddis R., Lopez-Poveda E. A., Fay R. R., Popper A. N.* Computational Models of the Auditory System. (Springer, 2010).
- [8] *Столярова Э. И.* Речевые технологии. №.2. С. 31. (2010).
- [9] *Люблинская В. В., Огородникова Е. А.* Речевые технологии. №. 2. С. 88. (2010).
- [10] *Чистович Л. А., Венцов А. В., Гранстерм М. П. и др.* Физиология речи. Восприятие человеком. (Л.: Наука, 1976).
- [11] *Утробин В. А.* Физико-математические интерпретации прогнозирующего восприятия и поведенческих свойств живых систем. Сложность. Разум. Постнеклассика. №. 13. С. 78. (2013).
- [12] *Кодзасов С. В., Кривнова О. Ф.* Общая фонетика. (М.: Изд-во РГГУ, 2001).
- [13] *Delgutte B.* Auditory Neural Processing of Speech. In Hardcastle W. J., Laver J. The Handbook of Phonetic

- Sciences. (Blackwell Publishing, Oxford, 1999).
- [14] Бибииков Н. Г. Акустический журнал. **48**. С. 447. (2002).
- [15] Сорокин В. Н., Чепелев Д. Н. Акустический журнал. **51**, №. 4. С. 536. (2005).
- [16] Гудонавичюс Р. В., Заездный А. М., Чиставичюс А. Б. Акустический журнал. **XIX**, №. 6. С. 824. (1973).
- [17] Dallos P. J. Neurosci. 5. P. 1591. (1985).
- [18] Утробин В. А. Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений. (Н. Новгород: Изд-во Нижегородского государственного технического университета им. П. Е. Алексеева, 2001).
- [19] Утробин В. А. Компьютерная обработка изображений. Принятие решений в пространстве эталонов. (Н. Новгород: Изд-во Нижегородского государственного технического университета им. П. Е. Алексеева, 2004).
- [20] Утробин В. А. Компьютерная обработка изображений. Анализ и синтез. (Н. Новгород: Изд-во Нижегородского государственного технического университета им. П. Е. Алексеева, 2003).
- [21] Утробин В. А. Компьютерная обработка изображений. Информационные модели этапа понимания. (Н. Новгород: Изд-во Нижегородского государственного технического университета им. П. Е. Алексеева, 2006).
- [22] Fontaine B., Steinberg L. J., Pena J. L. Sound envelope extraction in cochlear nucleus neurons: modulation filterbank and cellular mechanism. (BMC Neurosci, 2013).
- [23] Rieke F., Yamada W., Lewis E. R., Bialek W. Non-phase locked auditory cells and envelope detection In Neural Systems: Analysis and Modeling. (Springer US, 1993).
- [24] Fitch, R. H., Miller, S., Tallal, P. Annual Review of Neuroscience. **20**. P. 331. (1997).
- [25] A. de Cheveigne Pitch perception models. In: C. J. Plack, A. J. Oxenham, R. R. Fay, A. N. Popper, Pitch Neural Coding and Perception. (Springer, New York, 2005).
- [26] Гай В. Е. Цифровая обработка сигналов. №. 4. С. 65. (2013).

Information model of the auditory perception

V. E. Gai^a, N. V. Gai^b, O. A. Jakovlev^c, V. G. Sazhin^d

Chair of computing machines and technologies, Institute of radioelectronics and information technologies, Nizhny Novgorod State Technical University n. a. R. E. Alekseev, Nizhny Novgorod, 603950, Russian Federation
E-mail: ^aiamuser@inbox.ru, ^bnatalja.v.gai@gmail.com, ^cyakovlev.oleg2010@gmail.com, ^ddarthy.first@gmail.com

The research describes proposed model of information transformations taking place in auditory system. Drawing an analogy between processes of transformation information in auditory system and proposed model. This model is built on foundations of the theory of active perception (TAP) proposed for information processes modeling in visual perception. Based on proposed model were created several methods of processing of sound signals: computing method of human voice pitch frequency, method of useful signal extraction. Effectiveness of the proposed methods was confirmed by the conducted experimental researches.

PACS: 43.60.-c

Keywords: auditory system modeling, signal processing.

Received 09.11.2014.

Сведения об авторах

1. Гай Василий Евгеньевич — канд.техн.наук, доцент, доцент; тел.: (831) 436-82-28, e-mail: iamuser@inbox.ru.
2. Гай Наталья Васильевна — магистрант; тел.: (831) 436-82-28, e-mail: natalja.v.gai@gmail.com.
3. Яковлев Олег Алексеевич — студент; тел.: (831) 436-82-28, e-mail: yakovlev.oleg2010@gmail.com.
4. Сажин Владислав Геннадьевич — студент; тел.: (831) 436-82-28, e-mail: darthy.first@gmail.com.