

На правах рукописи



Тиунов Сергей Дмитриевич

**МОДЕЛЬ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКС ПРОГРАММ  
ДЛЯ АКУСТИЧЕСКОГО АНАЛИЗА ГОЛОСА В ЗАДАЧАХ  
ДИАГНОСТИКИ ГОЛОСОВЫХ РАССТРОЙСТВ**

Специальность 05.13.18

Математическое моделирование, численные методы и комплексы программ

**АВТОРЕФЕРАТ**

диссертации на соискание учёной степени

кандидата технических наук

Томск — 2014

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР)

Научный руководитель: доктор технических наук, профессор  
**Мещеряков Роман Валерьевич**

Официальные оппоненты:  
**Кочегуров Владимир Александрович,**  
доктор технических наук, профессор  
Национальный исследовательский  
Томский политехнический университет,  
профессор кафедры прикладной математики  
**Шамин Алексей Алексеевич,**  
кандидат технических наук,  
ООО НТП «Киберцентр»,  
начальник отдела разработки микропроцессорных  
терминалов

Ведущая организация: Федеральное государственное бюджетное образова-  
тельное учреждение высшего профессионального  
образования «Алтайский государственный универ-  
ситет»

Защита состоится «29» декабря 2014 г. в 15 ч. 15 мин. на заседании диссер-  
тационного совета Д 212.268.02 Томском государственном университете систем  
управления и радиоэлектроники по адресу: 634050, г. Томск, пр. Ленина, 40.

С диссертацией можно ознакомиться в научной библиотеке Томского госу-  
дарственного университета систем управления и радиоэлектроники по адресу:  
г. Томск, ул. Вершинина, 74 и на сайте Томского государственного университета  
систем управления и радиоэлектроники: <http://tusur.ru/ru/science/news/diss.html>

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2014 г.

Ученый секретарь  
диссертационного совета



Р. В. Мещеряков

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** В задачах обработки речевого сигнала, включая вычисление параметров сигнала, классификацию фонем, сегментацию и идентификацию диктора, наиболее остро стоит проблема точности вычислений. Если какой-либо алгоритм обработки речевого сигнала достигает определенного уровня точности, то для него находятся и приложения. При этом часто внимание не уделяется эффективности вычислений, что приводит к алгоритмам, сложным с вычислительной точки зрения. С другой стороны ряд приложений речевых технологий, таких как обучение иностранному языку, тренировка пения, запись текста под диктовку и многие другие, требуют не только высокой точности работы алгоритмов, но и предоставления результатов обработки речевого сигнала в реальном времени. Способы и методы понижения вычислительной сложности алгоритмов, которые не влияют на их точность, являются актуальными, поскольку позволяют использовать данные алгоритмы в новых областях применения.

Методам и алгоритмам эффективной цифровой обработки сигналов посвящены работы J. Cooley, L. Rabiner, S. Winograd, R. Agarwal, R. Blahut, M. Bodrato и другие. Алгоритмы нахождения частоты основного тона рассматривали и предлагали в своих работах L. Rabiner, M. Schroeder, M. Sondhi, A. Acero, G. Fant, A. Noll, Л. А. Чистович, М. В. Хитров, В. П. Бондаренко, А. Л. Ронжин и многие другие.

В области диагностики фониатрических заболеваний с помощью речевых технологий (т.н. акустический анализ голоса) требуется как точность вычислений, которая позволяет использовать акустические параметры голоса для постановки диагноза, так и скорость вычислений. Акустический анализ является лишь вспомогательным инструментом, частью комплексной методики лечения таких заболеваний как ларингит, поэтому не будет использоваться, если существенно увеличит время приема пациента. Параметры, вычисляемые в ходе акустического анализа, основываются на частоте основного тона сигнала, и чем надежнее алгоритм определения частоты основного тона, тем надежнее метод акустического анализа в целом.

**Целью** диссертационной работы является разработка моделей, алгоритмов и программных средств, позволяющих проводить акустический анализ голоса в реальном времени, при этом точность разрабатываемых алгоритмов должна быть обеспечена за счет их математической тождественности уже разработанным и проверенным алгоритмам. Это позволит использовать данные алгоритмы в новых областях применения, требующих их работы в реальном времени, таких как диагностика фониатрических заболеваний.

Для достижения поставленной цели были решены следующие **задачи**:

1) обзор и анализ существующих моделей работы голосовых складок при наличии голосового расстройства, анализ существующих методов акустического анализа голоса, рассмотрение структуры оптимизируемого алгоритма вычисления ЧОТ;

2) выбор или разработка моделей работы голосовых складок при наличии голосового расстройства, моделирование и анализ результатов;

3) разработка эффективных численных методов цифровой фильтрации сигнала, одновременной маскировки, поиска канала ЧОТ по шаблонам;

4) анализ разработанных численных методов с точки зрения точности и времени работы по сравнению с существующими аналогами и вычислительной мощностью современных ЭВМ;

5) разработка комплекса программ, использующего эффективные алгоритмы вычисления ЧОТ, и его апробация в задачах акустического анализа голоса.

**Объектом исследований** является моделирование процесса голосообразования при наличии голосового расстройства.

**Предметом исследований** являются модели и алгоритмы, применяемые для разработки и вычисления акустических параметров голоса, полезных при диагностике голосовых расстройств.

**Методы исследований.** Для решения поставленных в работе задач использовались методы математического моделирования и вычислительного эксперимента, методы функционального анализа, методы структурного и объектно-ориентированного программирования, методы планирования эксперимента.

**Научная новизна.**

1) Предложена новая *модель работы воспаленных голосовых складок*, объединяющая трехмассовую модель голосовых складок и модель ноцицепции для учета процесса ноцицепции при голосообразовании и влияния воспаления на ноцицепцию.

2) Предложен новый *численный метод одновременной маскировки* на основе цифровой фильтрации и сформулированы *ограничения*, при которых этот метод можно использовать.

3) Разработан новый *алгоритм поиска ближайшего двоичного шаблона*, который использует следующие методы повышения эффективности вычислений: использование таблиц поиска и векторизации, использование динамического программирования и исключение тривиальных операций.

**Практическая значимость** диссертационной работы. Разработанные алгоритмы и программные средства позволяют расширить область применения алгоритма вычисления частоты основного тона, основанного на модели слуха, и использовать его в приложениях, требующих работу *в реальном времени*, в

частности для акустического анализа голоса в задачах диагностики голосовых расстройств. Разработанные алгоритмы и программы могут быть использованы для *снижения требований к аппаратному обеспечению* других речевых приложений, при этом не влияя существенным образом на качество обработки.

**Основные положения, выносимые на защиту.**

1) Предложенная модель работы воспаленных голосовых складок позволяет *моделировать симптомы голосовых расстройств*: потеря голоса и короткое время непрерывной фонации.

2) Предложенный численный метод одновременной маскировки позволяет вычислять одновременную маскировку *за меньшее количество операций*, чем существующий алгоритм, при выполнении определенных ограничений, исследованных в диссертационной работе.

3) Разработанный алгоритм поиска ближайшего шаблона позволяет находить ближайший шаблон из фиксированного набора шаблонов *за меньшее количество операций*, чем существующий алгоритм.

4) Разработанный комплекс программ позволяет вычислять частоту основного тона речевого сигнала *в реальном времени с точностью исходного алгоритма*, основанного на модели слуха.

**Достоверность** изложенных в работе результатов обеспечивается строгостью применения математического аппарата, сравнением полученных результатов с известными знаниями и существующими аналогами, проведенным тестированием разработанного программного обеспечения в разрезе точности вычислений и времени выполнения, проведенным внедрением результатов работы.

**Внедрение результатов диссертационной работы.** Разработанный комплекс программ для вычисления параметров речевого сигнала используется в кабинете фониатрии городской поликлиники №10 г. Томска для задач акустического анализа голоса пациентов, а также в учебном процессе кафедры комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) ТУСУР.

**Апробация работы.** Основные результаты работы докладывались на следующих конференциях: Ежегодная международная открытая научная конференция «Современные проблемы информатизации», 2010; Региональная научно-практическая конференция «Многоядерные процессоры и параллельное программирование», 2011; Международная конференция «Речь и компьютер» SPECOM-2011; 5 междисциплинарный семинар «Анализ разговорной русской речи» — AP3-2011; II региональная научно-практическая конференция «Многоядерные процессоры и параллельное программирование», 2012; X Международная научно-практическая конференция «Электронные средства и системы управления», 2014.

**Личный вклад.** Основные научные результаты получены лично автором. Оригинальные методы обработки речи, которые были взяты за основу в данной работе, были разработаны Бондаренко В. П., Мещеряковым Р. В., Коневым А. А. Автором был осуществлен анализ возможностей оптимизации данных алгоритмов, разработка новых методов и алгоритмов, позволяющих получать те же самые результаты с меньшими вычислительными затратами. Разработанные методы и алгоритмы были реализованы в виде комплекса программ также лично автором.

**Публикации.** Основные результаты по теме диссертации изложены в 8 печатных изданиях, 2 из которых изданы в журналах, рекомендованных ВАК, 6 — в материалах и тезисах докладов.

**Объем и структура работы.** Диссертация состоит из введения, шести глав, заключения и приложения. Полный объем диссертации **133** страниц текста с **45** рисунками и **16** таблицами. Список использованных источников содержит **106** позиций.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

В **первой главе** приводится обзор литературных источников по теме работы.

Для целей моделирования голосообразования при патологии рассматриваются существующие математические модели голосовых складок, воспаления, ноцицепции (боли).

Для организации взаимосвязи модели голосовых складок и модели ноцицепции рассматривались существующие исследования по измерению чувствительности ноцицепторов (нейронов, специально предназначенных для ноцицепции) в воспаленной ткани и в норме. Найденные исследования отличаются многообразием наблюдаемых животных, органов и методов измерения, и как следствие, в них сильно различаются значения величин, но можно наблюдать общую форму зависимости.

Рассматриваются методы акустического анализа голоса. Эти методы являются частью комплексной диагностики фониатрических заболеваний, и представляют собой набор параметров голоса как звуковой волны. Для акустического анализа голоса используются такие параметры как HNR (отношение гармоника-шум), Jitter (частотная нестабильность), Shimmer (амплитудная нестабильность).

Все эти параметры так или иначе связаны с частотой основного тона (ЧОТ), и на основе более надежных алгоритмов вычисления ЧОТ можно построить и более надежные алгоритмы для вычисления акустических параметров голоса. Непосредственно ЧОТ также используется для акустического анализа голоса. Приводится краткая классификация и описание существующих методов вычисления частоты основного тона.

Приводится краткое описание структуры оптимизируемого алгоритма вычисления частоты основного тона (рисунок 1). Рассматриваемый алгоритм состоит из трех основных этапов: вычисление спектрограммы  $S$  (матрица размера  $K \times N$ ), вычисление одновременной маскировки  $M$  (матрица размера  $K \times N$ ) и поиск номеров наиболее близких шаблонов ЧОТ  $k$  (вектор длины  $N$ ), а также нескольких промежуточных этапов.

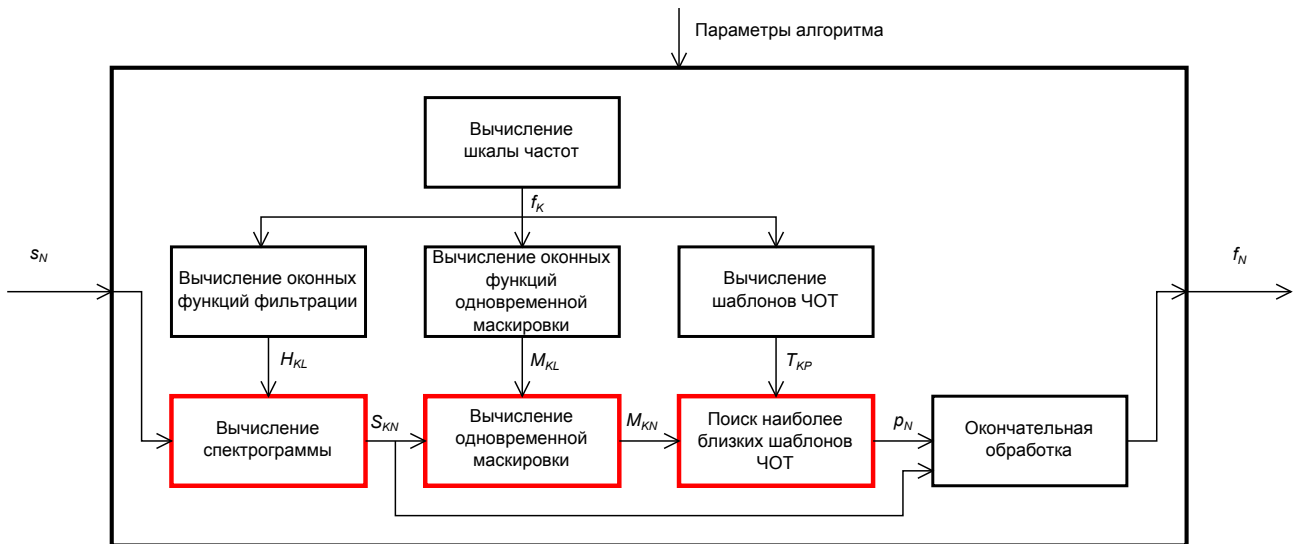


Рисунок 1 – Структура алгоритма вычисления ЧОТ

Рассматриваемый алгоритм вычисления ЧОТ основан на том, что матрица  $M$  зависит от частоты входного сигнала настолько сильно, что можно распознать ЧОТ входного сигнала по одновременной маскировке. Для этого используются двоичные шаблоны маскировки, каждый шаблон соответствует некоторой частоте. Наиболее близкий шаблон и указывает на частоту основного тона.

**Вторая глава** описывает математическую модель работы воспаленных голосовых складок и результаты моделирования голосообразования с использованием данной модели.

Математическая модель работы воспаленных голосовых складок включает в себя модель голосовых складок, модель ноцицепции, а также прямую и обратную связи между ними (рисунок 2).

Модель голосовых складок и модель ноцицепции были заимствованы у авторов В. Story и N. Britton и использованы без изменений.

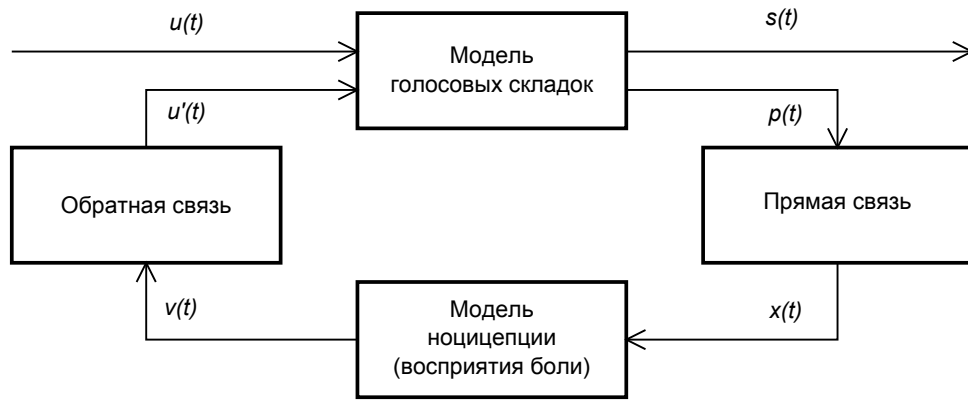
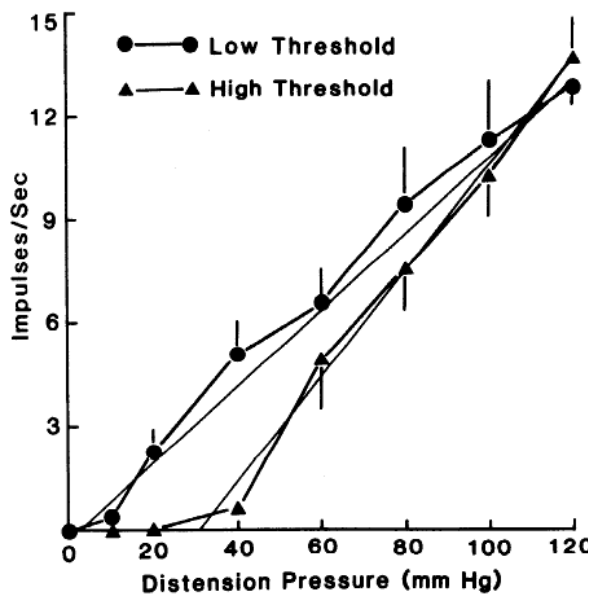
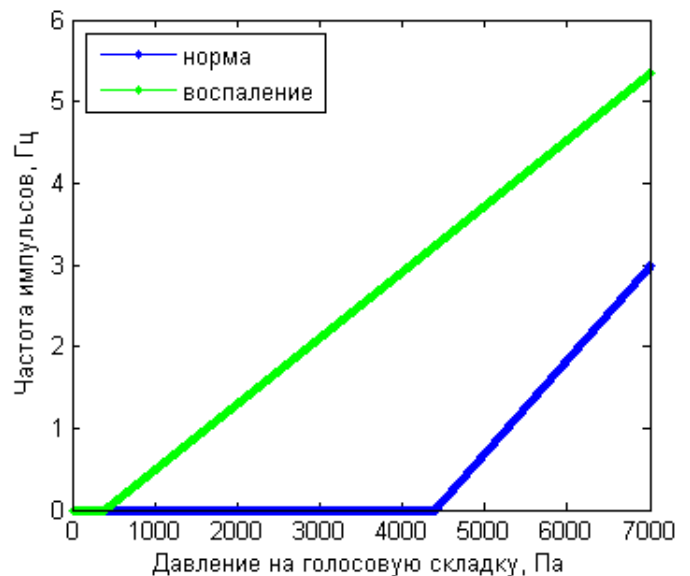


Рисунок 2 – Структура предлагаемой модели

Для организации прямой связи была использована аппроксимация экспериментальной зависимости частоты импульсов нервного волокна от давления на стенки пищевода (Sengupta, 1990), как наиболее близкой из найденных (рисунок 3а).



(а) Экспериментальная зависимость



(б) Модельная зависимость

Рисунок 3 – Чувствительность нервных волокон для нормальной и воспаленной ткани

Математическая запись модельной зависимости представлена ниже:

$$x(t) = \begin{cases} 0 & \text{если } p(t) - p_t \leq 0 \\ k \cdot (p(t) - p_t) & \text{если } p(t) - p_t > 0 \end{cases} \quad (1)$$

где  $k$  – коэффициент пропорциональности [Гц/Па],

$p_t$  – порог активации нервных волокон ноцицепторов [Па].

Используемые значения параметров  $k$  и  $p_t$ :

$$k = 1.1459 \times 10^{-3} \text{ Гц/Па} \quad p_t = 4400 \text{ Па} \quad \text{для нормальной ткани}$$

$$k = 0.8089 \times 10^{-3} \text{ Гц/Па} \quad p_t = 400 \text{ Па} \quad \text{для воспаленной ткани}$$



Графики полученных зависимостей для нормальной и воспаленной тканей показаны на рисунке 36.

Для организации обратной связи каких-либо экспериментальных данных найдено не было, поэтому была предложена простая модель зависимости уровня активации мышцы от потенциала ноцицептора в виде дифференциального уравнения.

$$\frac{da}{dt} = \begin{cases} -\frac{a - a_{min}}{T} \cdot \left( \frac{V - V_{thr}}{V_{thr} - V_{min}} \right)^2, & \text{если } V \geq V_{thr} \\ -\frac{a - a_0}{T} \cdot \left( \frac{V - V_{thr}}{V_{thr} - V_{min}} \right)^2, & \text{если } V < V_{thr} \end{cases} \quad (2)$$

где  $a$  — уровень активации мышцы,

$a_{min}$  — его минимальное значение (соответствует разьединенным голосовым складкам),

$a_0$  — его целевое (нормальное) значение,

$V$  — обозначает потенциал нейрона,

$V_{min}$  — его минимальное значение (состояние покоя),

$V_{thr}$  — пороговое значение (при котором регистрируется боль).

Поведение модели обратной связи можно описать так: если регистрируется боль, то уровень активации мышцы стремится к минимальному значению (то есть прекращению фонации), а в противном случае — к нормальному значению (то есть нормальной фонации).

Далее в главе приведены результаты моделирования голосообразования в норме и при патологии (то есть при наличии воспаления). Моделирование этого процесса в норме дает результаты, напоминающие результаты моделирования голосообразования без учета ноцицепции, поскольку порог боли  $p_t$  высокий и боль не регистрируется. Моделирование при наличии воспаления показывает иной результат: звуковой сигнал на время прекращается, затем начинается с меньшей силой и снова прекращается, и так далее (рисунок 4). Таким образом, удалось смоделировать такие симптомы ларингита как потеря голоса и короткое время непрерывной фонации.

**Третья глава** посвящена исследованию возможностей повышения эффективности вычисления спектрограммы. Вычисление спектрограммы сводится к цифровой фильтрации, рассматриваются методы потоковой цифровой фильтрации: метод пересечения с накоплением и метод пересечения с добавлением.

Также сравниваются методы свертки фиксированного размера: итерационный метод и метод, основанный на применении теоремы о свертке и быстрого преобразования Фурье (БПФ). Производится сравнение вычислительной сложности алгоритма, составленного по итерационному методу, и алгоритма на ос-

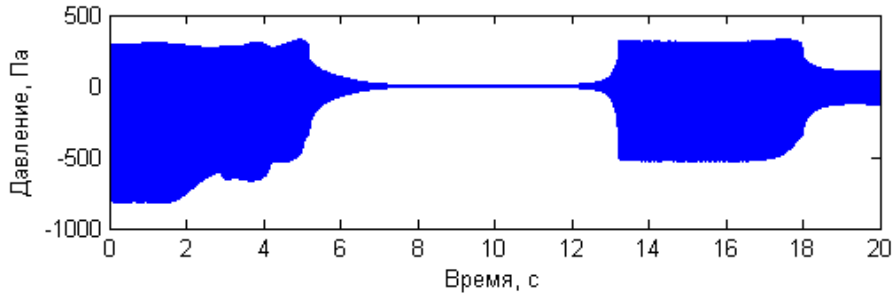


Рисунок 4 – Моделирование давления на выходе голосового тракта при патологии

нове БПФ для различных длин циклической свертки. Оказывается, что алгоритм на основе БПФ имеет меньшую вычислительную сложность.

Производится расчет требуемой вычислительной мощности в GFLOPS для работы системы фильтров в реальном времени для различных значений  $K$  (количество фильтров),  $f_q$  (частота дискретизации) и  $L$  (длина окна фильтрации). Получаются значения от 0,118 до 1,772 GFLOPS — это значит, что современные процессоры, обладающие вычислительной мощностью от нескольких GFLOPS до более чем 100 GFLOPS могут обеспечить работу системы фильтров в реальном времени. Также при рассмотрении результатов расчета обнаружено, что для каждой длины окна фильтрации  $L$  существует *оптимальная* длина циклической свертки, при которой минимизируется требуемая вычислительная мощность.

Приводится сравнение по точности вычислений двух программ, реализующих алгоритм на основе БПФ и итерационный алгоритм соответственно. Результаты получены в зависимости от параметра  $f_0$  — резонансной частоты фильтрации. Для итеративного алгоритма максимальная относительная ошибка составляет  $10^{-9}$ , а для алгоритма на основе БПФ —  $10^{-12}$ . Таким образом, оба алгоритма не снижают точность существенным образом, и поэтому могут быть использованы для эффективного вычисления спектрограммы.

**Четвертая глава** посвящена исследованию возможностей повышения эффективности вычисления одновременной маскировки. Вычисление в оригинальном алгоритме одновременной маскировки производится по формуле:

$$\begin{aligned} \forall n &= \overline{0, N-1} : \\ \forall k &= \overline{0, K-1} : \\ M(k, n) &= \left( \sum_{l=0}^{L-1} |S(k', n)|^2 \cdot H(k, l) \right) \cdot b < |S(k, n)|^2 \end{aligned} \quad (3)$$

$$\text{где } k' = \begin{cases} 0, & \text{если } 0 > k + l - L/2 \\ k + l - L/2, & \text{если } 0 \leq k + l - L/2 \leq K - 1, \\ K - 1, & \text{если } k + l - L/2 > K - 1 \end{cases}$$

$b = const.$

В общем случае, вычисление по этой формуле включает в себя  $K \cdot N \cdot L$  уникальных умножений, и поэтому не может быть оптимизировано. В главе приводится расчет требуемой мощности вычислительной машины для расчета одновременной маскировки по наивному алгоритму: для различных параметров  $K$  (количество каналов фильтрации) и  $f_q$  (частота дискретизации) требуемая вычислительная мощность составляет от 0.03 до 0.69 GFLOPS.

Рассматривается следующее ограничение на матрицу  $H$ :

$$\forall i, j = \overline{0, K-1} : H_i = H_j \quad (4)$$

При выполнении ограничения (4) матрицу  $H$  можно вычислить через линейную свертку:

$$\begin{aligned} \forall n &= \overline{0, N-1} : \\ \forall k &= \overline{0, K-1} : \\ M(k, n) &= b \cdot \sum_{l=0}^{L-1} |S'(k+l, n)|^2 \cdot H(i, l) < |S(k, n)|^2 = \\ &= b \cdot M'(k+L-1) > |S(k, n)|^2 \end{aligned} \quad (5)$$

где  $M' = |S'^n|^2 \otimes \bar{H}'$ ,

$\otimes$  — оператор линейной свертки,

$\forall k = \overline{0, K-1}, \forall l = \overline{0, L-1}, \forall n = \overline{0, N-1} : S'(k+l, n) = S(k', n)$ ,

$\forall l = \overline{0, L-1} : \bar{H}'(l) = H(i, L-1-l)$ .

Формулу (5) можно вычислить через эффективные методы цифровой фильтрации, рассмотренные в предыдущей главе.

Приводятся детали вычисления матрицы  $H$ : помимо различных параметров на значения матрицы  $H$  влияет шкала частот  $f$ . Таким образом, условие эффективного вычисления одновременной маскировки налагает ограничения на форму шкалы частот. Выводится общий вид шкалы частот, который удовлетворяет данным ограничениям:

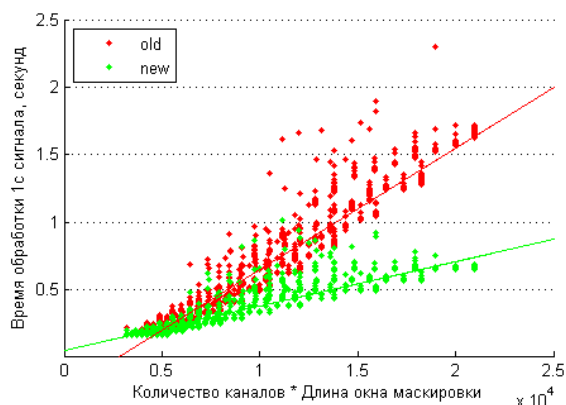
$$f(x) = \exp(a \cdot x + b) - \gamma \quad (6)$$

где  $a, b$  — параметры шкалы частот,

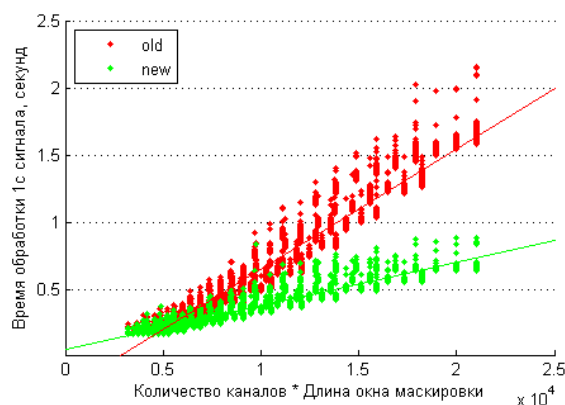
$\gamma$  — константа.

Далее в главе описывается и рассматривается алгоритм вычисления одновременной маскировки через цифровую фильтрацию. Проводятся эксперименты по измерению времени и точности работы по сравнению с оригинальным алгоритмом вычисления одновременной маскировки. Было проведено планирование эксперимента: выбраны звуковые сигналы различной природы (шумовые и речевые), различные практически применимые параметры алгоритма (всего 121

сочетание параметров), в случае вычисления времени работы одни и те же вычисления производились несколько раз. Результаты экспериментов по времени работы представлены на рисунке 5. Из представленных графиков рассеяния видно, что предлагаемый алгоритм производит вычисления за меньшее количество времени.



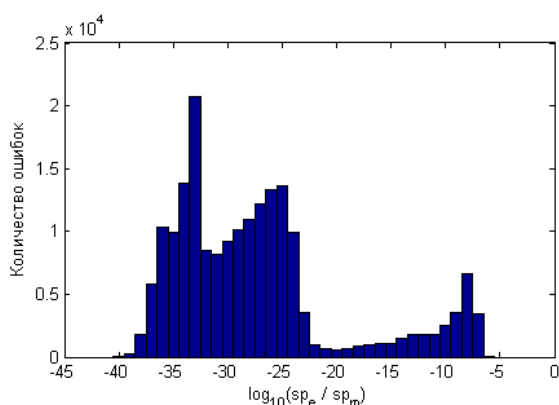
(а) Набор сигналов шума



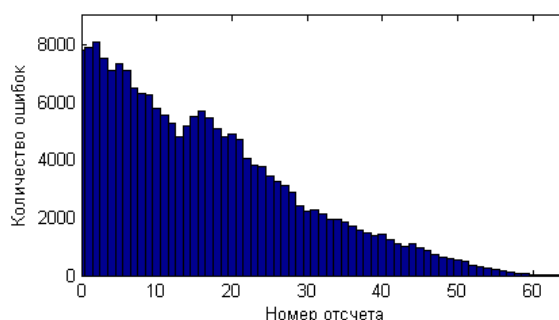
(б) Набор речевых сигналов

Рисунок 5 – Статистика времени работы исходного (old) и предлагаемого (new) алгоритмов

На рисунке 6а представлено распределение ошибок вычисления в зависимости от относительной величины входного спектра. Видно, что ошибки вычисления происходят лишь при малых относительных значениях спектра (менее  $10^{-5}$ ), а основная часть ошибок — при очень малых значениях (менее  $10^{-20}$ ). При рассмотрении аналогичных распределений в рамках отдельных сигналов форма зависимости сохраняется.



(а) в зависимости от величины входного спектра



(б) в зависимости от номера отсчета

Рисунок 6 – Статистика ошибок вычисления предлагаемого алгоритма

На рисунке 6б показано распределение ошибок по номерам отсчетов. Видно, что ошибки вычисления происходят в основном при обработке первых отсчетов спектрограммы, которые как правило имеют малые значения из-за того,

что при вычислении спектрограммы сигнал принимается равным нулю при значениях времени  $t < 0$ .

**Пятая глава** посвящена повышению эффективности алгоритма поиска ближайшего двоичного шаблона. Рассматривается оригинальный алгоритм поиска по шаблону: используемые формулы и оценка количества требуемых операций:

$$N_{\Delta} = N \cdot P \cdot K \quad (7)$$

$$N_A = N \cdot P \cdot (K - 1) \quad (8)$$

где  $N$  — количество временных отсчетов,

$K$  — количество каналов фильтрации,

$P$  — количество шаблонов, соответствующих различным значениям ЧОТ,

$N_A$  — количество сложений,

$N_{\Delta}$  — количество операций  $\delta$  (вычисление расстояния Хэмминга).

Рассматриваются следующие возможности повышения эффективности алгоритма: исключение тривиальных операций, использование динамического программирования и использование таблиц поиска и векторизации.

Использование таблиц поиска позволяет заменить операции  $\delta$  и некоторые сложения на поиск по таблице. Для этого входной вектор размера  $K$  разбивается на  $C$  частей размера  $L$  каждая. У каждой из данных частей возможно  $R = 2^L$  различных вариантов (поскольку входной вектор двоичный). Тогда можно составить таблицу (матрицу) поиска  $D$  размера  $C \times R \times P$ , в ячейках которой будет содержаться расстояние между  $c$ -той частью  $p$ -того шаблона и  $r$ -тым вариантом  $c$ -той части некоторого входного вектора. Иллюстрация такой таблицы приведена на рисунке 7 для фиксированного  $p$ . Тогда суммирование векторов по каждой  $c$ -той части приводит нас к вектору  $D_P$ , в которой каждый  $p$ -тый элемент соответствует расстоянию входного вектора до  $p$ -того шаблона. Теперь для нахождения шаблона, наиболее близкого к входному вектору, достаточно найти наименьшее значение в векторе  $D_P$ .

Для сложения векторов чисел предлагается использовать метод векторизации: специальные векторные инструкции процессоров (SIMD-инструкции), которые позволяют сложить сразу несколько пар чисел.

Использование динамического программирования заключается в том, что используются результаты вычисления предыдущего отсчета. В сочетании с использованием таблиц поиска и векторизации, это сводится к корректировке вектора расстояний  $D_P$ : для этого нужно посмотреть, какие части входного вектора изменились, и для измененных частей вычесть соответствующие вектора расстояний (из таблицы поиска  $D$ ) для предыдущих значений входного вектора и прибавить — для текущих значений.

К элементов, С частей по L элементов



Рисунок 7 – Таблица поиска (матрица расстояний)  $D_{CRP}$

Исключение тривиальных операций сводится к исключению сложения с нулем.

В главе разрабатывается, описывается и оценивается эффективный алгоритм поиска ближайшего шаблона. Блок-схема алгоритма приведена на рисунке 8.

Количество операций сложения, необходимых для работы алгоритма:

$$N_A = \frac{N \cdot P \cdot (C - 1)}{F_{vec} \cdot F_{dyn} \cdot F_{tr}} \quad (9)$$

где  $F_{vec}$  — фактор векторизации (во сколько раз разрядность у машинного регистра больше, чем элемента расстояния),

$F_{dyn}$  — фактор динамичности (обратная средняя доля изменяемых частей входного вектора),

$F_{tr}$  — фактор тривиальности (обратная доля нулей в таблице поиска).

Факторы  $F_{vec}$ ,  $F_{dyn}$ ,  $F_{tr}$  обозначают, во сколько раз уменьшается количество необходимых операций, при применении векторизации, динамического программирования и исключения тривиальных операций соответственно. Значения  $F_{vec}$  и  $F_{tr}$  всегда больше единицы при практически используемых значениях параметров алгоритма. Значение  $F_{dyn}$  предположительно больше единицы, но зависит от качества входного сигнала: для гармонических сигналов  $F_{dyn}$  высоко, а для шумовых — низко.

В таблице 1 приведена оценка требуемой вычислительной мощности компьютера для работы оригинального и предлагаемого алгоритмов в реальном времени. Из таблицы видно, что предлагаемый алгоритм является гораздо менее требовательным, чем оригинальный алгоритм. Однако при этом требуется некоторый объем памяти: от нескольких мегабайт до нескольких гигабайт, что является вполне доступным для современных компьютеров.

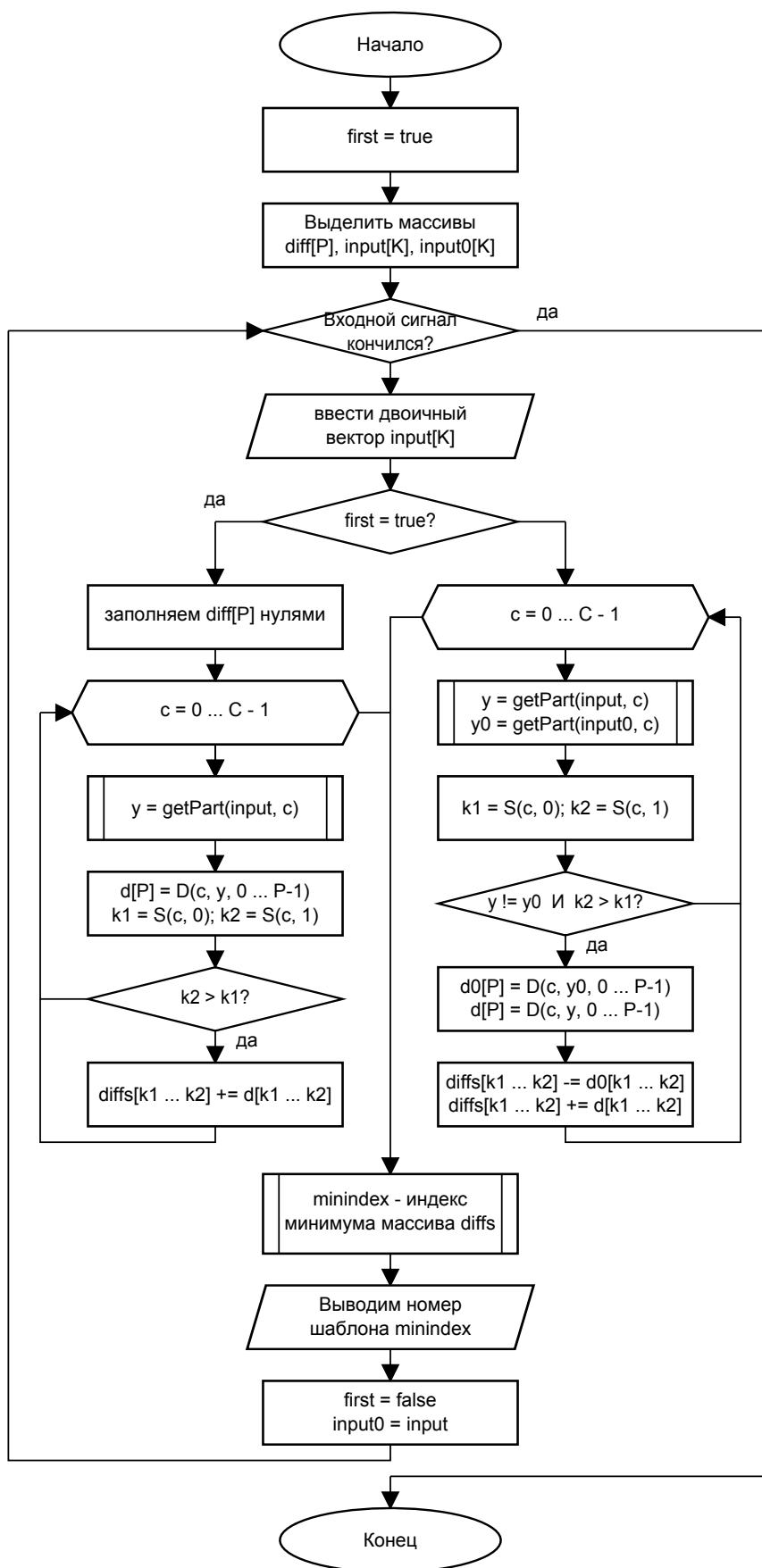


Рисунок 8 – Алгоритм быстрого поиска оптимального шаблона

Таблица 1 – Расчеты требуемой мощности вычислительной машины для поиска канала ЧОТ в реальном времени

Необходимая вычислительная мощность, млн. операций в сек.									
Операции	Оригинальный алгоритм								
Операции $\delta$	102,4	409,6	1638,4	102,4	409,6	1638,4	102,4	409,6	1638,4
Сложения	101,6	408	1635,2	101,6	408	1635,2	101,6	408	1635,2
$L$	Предлагаемый алгоритм								
4	1,55	12,60	50,80	0,78	6,30	25,40	0,39	3,15	12,70
8	0,75	6,20	25,20	0,38	3,10	12,60	0,19	1,55	6,30
16	0,35	3,00	12,40	0,18	1,50	6,20	0,09	0,75	3,10
$f_q$ , кГц	16	16	16	16	16	16	16	16	16
$K$	128	256	512	128	256	512	128	256	512
$P$	50	100	200	50	100	200	50	100	200
$F_{vec}$	8	4	4	8	4	4	8	4	4
$F_{dyn}$	1	1	1	2	2	2	4	4	4
$F_{tr}$	2	2	2	2	2	2	2	2	2

**Шестая глава** посвящена разработке эффективных программных средств для вычисления частоты основного тона в реальном времени.

Архитектура разработанного комплекса программ основана на использовании *конвейера* (подобно конвейерам UNIX). Конвейер состоит из последовательного соединения фильтров и буферов: *фильтры* вводят данные, преобразуют их определенным образом и выводят, а *буферы* — это устройство памяти, которое сохраняет результаты одного фильтра для передачи их другому фильтру. Реализованные фильтры позволяют вычислять спектрограмму, одновременную маскировку и частоту основного тона (рисунок 9).

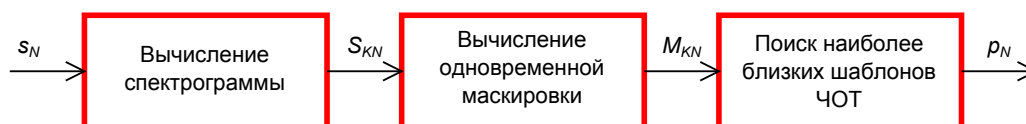


Рисунок 9 – Архитектура комплекса программ: конвейер для вычисления ЧОТ

Каждый фильтр работает в выделенном потоке программы, согласование работы фильтров происходит через буферы, которые поддерживают многопоточность благодаря применению событий и критических секций Windows.

В состав комплекса входят следующие модули:

- Библиотека SPL (Speech Parameter Library) — содержит компоненты для вычисления параметров и работы с вводом/выводом, интерфейс C++ и C (для работы со средой MATLAB).



- Тестовая программа — содержит процедуры тестирования библиотеки SPL — по корректности реализации, точности вычислений, времени выполнения.
- Программа вычисления частоты основного тона.
- Зависимости:
  - библиотека FFTW для быстрого вычисления БПФ;
  - библиотека MPFR для быстрых векторных операций.

Разработанный комплекс программ позволяет вычислять частоту основного тона в реальном времени при этом занимая около 10% процессорного времени Intel Core i5 2.4 ГГц и менее 100 Мб оперативной памяти.

Комплекс был внедрен в фониатрическом кабинете городской поликлиники №10 г. Томска для целей акустического анализа сигнала и позволил повысить согласованность вычисляемых акустических параметров со слуховой оценкой врача-фониатра (до 10% случаев).

В **заключении** приведены выводы по работе, сформулированы основные результаты.

## **ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ**

В результате работы были разработаны модель, численные методы, алгоритмы и программное обеспечение на основе этих алгоритмов, которые позволяют проводить акустический анализ голоса в реальном времени с высокой точностью, обеспеченной ранее проведенными исследованиями и проверенными методами. Основные результаты работы представлены ниже.

1) Проведен обзор существующих моделей голосообразования, воспаления, ноцицепции, исследований чувствительности ноцицепторов в норме и при воспалении. Для моделирования голосообразования при голосовом расстройстве были выбраны трехмассовая модель голосовых складок, модель ноцицепции Бриттона, модель мышечного управления голосовыми складками и экспериментальная зависимость чувствительности нейрона от давления на ткани. Проанализирован оригинальный алгоритм вычисления основного тона, выделены наиболее ресурсоемкие части алгоритма: вычисление спектрограммы, одновременной маскировки и поиск канала ЧОТ по шаблону.

2) Разработана модель работы голосовых складок при воспалении. Проведено моделирование голосообразования при воспалении, удалось смоделировать некоторые симптомы ларингита (потеря голоса, короткое время непрерывной фонации).

3) Проанализированы возможности понижения вычислительной сложности этапа вычисления спектрограммы. Выбран и обоснован эффективный метод вычисления спектрограммы (цифровая фильтрация по методу перекрытия с на-

коплением) и конкретные реализации алгоритмов (алгоритм циклической свертки на основе БПФ, с комплексным окном фильтрации и вещественным входным сигналом). Теоретически оценена вычислительная сложность полученного алгоритма вычисления спектрограммы по сравнению с современным уровнем развития вычислительной техники, данная оценка показала, что алгоритм может быть использован для вычисления спектрограммы в реальном времени на современных процессорах широкого потребления в пределах практически используемых значений параметров вычисления. Реализована тестовая программа и проведен эксперимент по точности вычисления спектрограммы, который показал, что полученный алгоритм вычисления спектрограммы имеет допустимую ошибку ( $10^{-12}$ ) для практического применения.

4) Проанализированы возможности понижения вычислительной сложности этапа вычисления одновременной маскировки, в частности проанализированы условия вычисления одновременной маскировки с помощью методов цифровой фильтрации. Получен общий вид шкал частот, при использовании которых выполняется условие вычисления одновременной маскировки с помощью цифровой фильтрации. Проанализированы известные шкалы частот с точки зрения соответствия требуемому виду частот, найдена практически применяемая и обоснованная шкала частот, которая соответствует данному виду. Разработан и оценен алгоритм вычисления одновременной маскировки с помощью цифровой фильтрации.

5) Проанализированы возможности понижения вычислительной сложности этапа поиска ближайшего шаблона, предложено несколько путей понижения сложности вычисления по данному алгоритму. Проанализированы возможности сочетания рассмотренных способов понижения вычислительной сложности, оказалось, что предложенные способы ортогонально сочетаются, то есть понижают вычислительную сложность одинаково независимо друг от друга. Разработан, оценен и протестирован алгоритм поиска ближайшего шаблона, получена теоретическая оценка для количества операций данного алгоритма.

6) Разработан и протестирован программный комплекс для вычисления параметров речевого сигнала, который способен работать в реальном времени на современных процессорах широкого потребления в пределах практически используемых параметров вычисления. Проведено внедрение комплекса программ в кабинете фониапии городской поликлиники №10 г. Томска, программный комплекс позволил повысить надежность акустического анализа, и как следствие, повысить точность диагностики голосовых расстройств в целом (до 10 %).

## СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

*В журналах, рекомендованных ВАК Министерства образования и науки Российской Федерации:*

1. Тиунов С. Д., Мещеряков Р. В., Сизов А. Г. Оценка качества голоса лиц, страдающих дисфонией // Тр. СПИИРАН. 2012. Т. 20. С. 138–152.
2. Тиунов С. Д., Мещеряков Р. В., Черных Д. В. Оптимизация вычисления одновременной маскировки речевого сигнала // Труды СПИИРАН. 2014. Т. 32. С. 45–57.

*В других изданиях:*

3. Тиунов С. Д., Мещеряков Р. В. Оптимизация времени цифровой фильтрации в исследованиях, связанных с обработкой речи // Ежегодная международная открытая научная конференция «Современные проблемы информатизации». 2010.
4. Тиунов С. Д., Мещеряков Р. В. Анализ возможности проведения параллельных вычислений в модели слуховой системы // Региональная научно-практическая конференция «Многоядерные процессоры и параллельное программирование». 2011.
5. Tiunov S. D., Mescheriakov R. V., Mukhina V. I. Acoustical Analysis of Laryngitis Treatment Method Effectiveness // SPECOM 2011 Proceedings. 14th International Conference on Speech and Computer / Ed. by R. K. Potapova. 2011. P. 191–195.
6. Речевые технологии в задаче обучения студентов-носителей русского языка произношению на иностранном языке / С. Д. Тиунов, Р. В. Мещеряков, Ю. М. Лирмак [и др.] // 5 междисциплинарный семинар «Анализ разговорной русской речи» - AP<sup>3</sup>-2011. Санкт-Петербург: 2011.
7. Тиунов С. Д., Мещеряков Р. В. Оценка вычислительной сложности алгоритмов обработки речевого потока, основанных на модели слуха // II региональная научно-практическая конференция «Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов». 2012.
8. Тиунов С. Д. Моделирование голосообразования при патологии // Электронные средства и системы управления: Материалы докладов X Международной научно-практической конференции. Т. 2. Томск: В-Спектр, 2014. С. 19–23.