

Панов Александр Ефимович

**Алгоритмы и конвейерное вычислительное устройство
помехоустойчивой сегментации вокализованных участков
речевого сигнала**

Специальность 05.13.05 – Элементы и устройства вы-
числительной техники и систем управления

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Курск – 2008

Работа выполнена в ГОУ ВПО «Курский государственный технический университет» на кафедре вычислительной техники в совместной научно-исследовательской лаборатории Центра информационных технологий в проектировании РАН и Курского государственного технического университета: «Информационные распознающие телекоммуникационные системы»

Научный руководитель : доктор технических наук, профессор
Типикин А.П.

Официальные оппоненты: доктор технических наук, профессор
Филист С.А.

кандидат технических наук
Старков Е.Ф.

Ведущая организация: ФГУП «Курский НИИ» МО РФ

Защита состоится 25 декабря 2008г. в 14 часов на заседании диссертационного совета Д.212.105.02 в Курском государственном техническом университете по адресу: 305040, г. Курск, ул. 50 лет Октября, 94, конференц-зал.

С диссертацией можно ознакомиться в библиотеке КурскГТУ.

Автореферат разослан 25 ноября 2008 года.

Ученый секретарь
диссертационного совета

Титенко Е.А.

Актуальность работы. Первичная обработка речевого сигнала (РС) существенно влияет на качество функционирования информационно-речевых систем. При выделении содержащейся в РС информации необходимо располагать как можно большим количеством сведений о его структуре. Вероятность правильного распознавания речи и идентификации диктора зависит от точности определения траектории основного тона и сегментирования РС, то есть разделения его на вокализованные и невокализованные участки, разделения вокализованных участков на интервалы смежных вокализованных фонем, выделения самых коротких фонемных интервалов длительностью около четырех квазипериодов тональных колебаний. Эти же данные необходимы для повышения степени сжатия речевой информации. Причем эти процедуры первичной сегментации должны выполняться в реальном времени в темпе поступления в инструментальную ЭВМ оцифрованного РС.

Неразборчивость и шумовой фон реальной речи, большая нестабильность формирования звуковых колебаний речевым аппаратом существенно усложняют решение названных выше задач сегментации РС. В связи с этим известные методы и алгоритмы первичной обработки и сегментации РС имеют большую математическую и вычислительную сложность. Они ориентированы в основном на программную реализацию и пока имеют недостаточные помехоустойчивость и скорость работы. Необходимость дальнейшего повышения помехоустойчивости требует еще большего увеличения сложности названных алгоритмов, что при программной реализации приведет к дополнительному снижению скорости первичной обработки и сегментации РС.

В тоже время успехи микроэлектроники в повышении степени интеграции элементной базы ЭВМ позволяют аппаратно реализовать в одной специализированной СБИС очень сложные вычислительные алгоритмы и достичь требуемой высокой производительности за счет параллельной организации вычислений. Для этого требуется разработать метод повышения помехоустойчивости первичной обработки РС путем усовершенствования вычислительных алгоритмов сегментации и способ увеличения ее производительности за счет параллельной аппаратной реализации основных ее процедур оценивания мгновенных значений квазипериодов основного тона и сегментации вокализованных фонем.

В связи с вышеизложенным актуальной является **научно-техническая задача:** повышение помехоустойчивости аппаратно-ориентированных алгоритмов выделения в реальном времени траектории основного тона и вокализованных фонем.

Объект исследования: специализированное вычислительное устройство предварительной обработки речевого сигнала.

Предмет исследования: помехоустойчивые алгоритмы и устройство сегментации вокализованных участков РС в реальном времени.

Работа выполнена в 2005-2008 г.г. по плану совместных НИР научно-исследовательской лаборатории Центра информационных технологий в проектировании РАН и Курского государственного технического университета: «Ин-

формационные распознающие телекоммуникационные интеллектуальные системы».

Цель работы: разработка методов, алгоритмов и конвейерного устройства помехоустойчивого выделения основного тона и вокализованных фонем речевого сигнала.

Для достижения поставленной цели решены следующие задачи.

1. Анализ методов, алгоритмов и устройств помехоустойчивого оценивания основного тона и сегментации вокализованных участков речевого сигнала.
2. Модификация корреляционного метода оценивания основного тона путем введения многооконной скользящей обработки сигнала с целью повышения помехоустойчивости определения траектории основного тона.
3. Разработка метода сегментации вокализованных участков речи на фонемы, позволяющего повысить вероятность нахождения границ между ними.
4. Разработка вычислительного алгоритма потокового нахождения множества корреляционных функций в нескольких, разных по длительности, окнах скользящей обработки, как предпосылки для конвейеризации процедур определения траектории основного тона.
5. Разработка алгоритмов, структурной и функциональных схем специализированного конвейерного вычислительного устройства сегментации вокализованных участков речевого сигнала, работающего в реальном масштабе времени.
6. Программное моделирование алгоритмов функционирования разработанного устройства и оценка выигрыша по скорости и помехоустойчивости разработанных методов, алгоритмов и устройства. Оценка аппаратной сложности устройства при реализации на ПЛИС.

Научная новизна результатов работы состоит в следующем:

1. Разработан модифицированный корреляционный метод оценивания мгновенных значений квазипериодов основного тона речевого сигнала, отличающийся применением при скользящей обработке нескольких временных окон различных длительностей и отслеживанием переменного квазипериода наиболее подходящим по длительности окном, позволяющий повысить помехоустойчивость воспроизведения траектории основного тона.
2. Разработан метод сегментации вокализованных участков речи на фонемы, основанный на вычислении расстояния между нормализованными по частоте спектральными плотностями речевого сигнала в разных его квазипериодах, отличающийся учетом непостоянства тональной частоты, ограничением диапазона усреднения спектров по числу квазипериодов, специальными правилами локализации переходного участка между фонемами и позволяющий повысить вероятность нахождения границ между ними и помехоустойчивость процедур сегментации.
3. Разработан рекурсивный вычислительный алгоритм потокового нахождения множества корреляционных функций сигнала в нескольких разных окнах скользящей обработки, позволивший аппаратно совместить их одно-

временное определение и организовать конвейерное выполнение всех процедур определения траектории основного тона.

4. Разработаны алгоритмы, структурная и функциональные схемы специализированного вычислительного устройства сегментации вокализованных участков речевого сигнала, отличающегося двухкаскадной аппаратно-программной конвейерной организацией с программной реализацией в цифровом процессоре сигналов во втором каскаде заключительной процедуры сегментации на фонемы и аппаратной реализацией первого каскада в виде шестиступенчатого конвейера процедур оценивания тона, позволяющего достичь быстродействия, достаточного для выполнения предварительной обработки сигнала в реальном времени.

Практическая ценность.

1. Разработанное устройство может быть встроено в звуковую плату ПЭВМ для расширения ее функциональных возможностей по предварительной обработке речевых сигналов, разгружая центральный процессор. Выходные данные устройства являются необходимыми исходными данными для определения признаков при распознавании речи и идентификации диктора на программном уровне ПЭВМ, а также при сжатии речевой информации.
2. Разработанная структурно-функциональная организация устройства позволяет не менее чем в 30 раз повысить производительность по сравнению с программной реализацией в многоядерных микропроцессорах и цифровых процессорах сигналов и обрабатывать речевые сигналы в реальном масштабе времени на тактовой частоте 50-400 МГц.
3. В результате экспериментальных исследований функционирования программной модели устройства на фонотеке, включающей до 100 слов и фраз, показано, что разработанные методы и алгоритмы обеспечивают следующие показатели помехоустойчивости. Траектория основного тона воспроизводится с вероятностью 0.4...0.96 и погрешностью 0.02...0.015, границы между вокализованными фонемами находятся с вероятностью от 0.4 до 0.95 с погрешностью не более двух квазипериодов основного тона в диапазоне отношения сигнал/шум от -3дБ до 8дБ, что в среднем в 2 раза лучше показателей помехоустойчивости известных методов и алгоритмов.

Реализация и внедрение. Результаты работы использованы в ОКБ «Авиаавтоматика» (г. Курск), в ООО «ЮнионСофт технолоджиз», а также в учебном процессе Курского государственного технического университета.

Апробация работы. Основные положения и результаты диссертационной работы докладывались и обсуждались на международных конференциях: МНТК «Распознавание-2003» (г. Курск, 2003), МНТК «Распознавание-2005» (г. Курск, 2005), МНТК «Распознавание-2008» (г. Курск, 2008), 22 межвузовской научно-технической конференции «Молодежь и XXI век» (г. Курск, 2003), МНТК «Information and telecommunication technologies in intelligent systems» (Mallorca, Spain, 2007), а также на научных семинарах кафедры ВТ КурскГТУ.

Публикации. По материалам диссертации опубликовано 10 работ, в том числе 3 статьи, 2 из которых – в журналах, входящих в перечень ВАК, 5 тезисов докладов и 2 свидетельства об официальной регистрации программ для ЭВМ.

Личный вклад автора. В работах, опубликованных в соавторстве, лично автором в [1,3,4] разработан модифицированный корреляционный метод оценивания мгновенных значений квазипериодов основного тона; в [2,5] разработан метод сегментации вокализованных участков речи на фонемы; в [6,8] проведена оценка помехоустойчивости названных методов сегментации вокализованных участков речевого сигнала; в [7] разработан рекурсивный алгоритм потокового вычисления множества корреляционных функций, структурные и функциональные схемы специализированного вычислительного устройства сегментации вокализованных участков речевого сигнала.

На защиту выносятся следующие результаты:

1. Модифицированный корреляционный метод определения траектории основного тона речевого сигнала, отличающийся применением нескольких, убывающих по длительности временных окон скользящей обработки, позволяющий повысить помехоустойчивость и создать предпосылки для параллельного выполнения процедур оценивания основного тона.
2. Метод сегментации вокализованных участков речи на фонемы, основанный на вычислении расстояния между нормализованными по частоте спектральными плотностями речевого сигнала в разных его квазипериодах и позволяющий повысить вероятность нахождения границ между ними и помехоустойчивость процедур сегментации.
3. Рекурсивный вычислительный алгоритм потокового нахождения множества корреляционных функций сигнала в нескольких разных окнах скользящей обработки, позволивший аппаратно совместить их одновременное определение и организовать конвейерное выполнение всех процедур определения траектории основного тона.
4. Алгоритмы, структурная и функциональные схемы специализированного вычислительного устройства сегментации вокализованных участков речевого сигнала, отличающегося двухкаскадной аппаратно-программной конвейерной организацией и аппаратной реализацией первого каскада в виде шестиступенчатого конвейера процедур оценивания тона, позволяющего достичь быстродействия, достаточного для выполнения предварительной обработки сигнала в реальном времени.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения и приложений. Список литературы включает 83 наименования.

Содержание работы

Во введении обоснована актуальность, научная новизна исследований, определена цель работы и задачи исследований, дана общая характеристика работы, сформулированы основные положения, выносимые на защиту.

В первой главе проведен анализ методов, алгоритмов и быстродействующих устройств помехоустойчивого оценивания основного тона и сегментации вокализованных участков речевого сигнала.

Среди методов оценивания частоты основного тона следует отметить эмпирические, корреляционные, спектральные, гомоморфные (кепстральные) методы, а также метод наименьших квадратов. В некоторых работах предлагается одновременное использование нескольких методов, например, анализ поведения определителя автокорреляционной матрицы, пиковый метод, метод аппроксимации временной функции основного тона. При этом достоверное значение принимается мажоритарно.

Анализ корреляционных методов оценивания частоты основного тона, как наиболее устойчивых к шумам, показал, что все они работают в ограниченном диапазоне частот и не могут восстановить непрерывную траекторию основного тона в широком диапазоне частот человеческого голоса. Для расширения рабочего диапазона частот обрабатываемого сигнала и дальнейшего повышения помехоустойчивости целесообразно перейти от скользящей обработки сигнала в окне фиксированной длительности к многооконной обработке с отслеживанием переменного квазипериода наиболее подходящим по длительности окном, позволяющей повысить помехоустойчивость воспроизведения траектории основного тона.

Обзор методов сегментации показал, что известные методы сегментации РС являются обособленными и работают независимо от алгоритмов определения траектории частоты основного тона, требуют ресурсоемких процедур вычисления, имеют низкую вероятность сегментации и неустойчивы к внешним помехам.

Поэтому для повышения помехоустойчивости предварительной обработки речи, в том числе определения траектории частоты основного тона и выделения вокализованных фонем из речевого сигнала, необходимо создание комплексного алгоритма сегментации РС, с аппаратно-программной поддержкой процедур расчета значений корреляционной функции и принятия решений, а также выделения лучших усредненных спектральных характеристик фонем, необходимых для дальнейшей обработки РС в информационно-речевых системах.

Во второй главе разрабатываются новые методы определения траектории частоты основного тона (ЧОТ) и сегментации РС на вокализованные фонемы.

Сегментация выполняется в два этапа:

1. Определение частоты основного тона в вокализованных сегментах РС и построение траектории основного тона по найденным значениям.

2. Сегментация вокализованной речи на вокализованные фонемы по траектории основного тона.

На первом этапе происходит дополнительная селекция по признаку вокализованный/невокализованный участок РС (рис. 1).

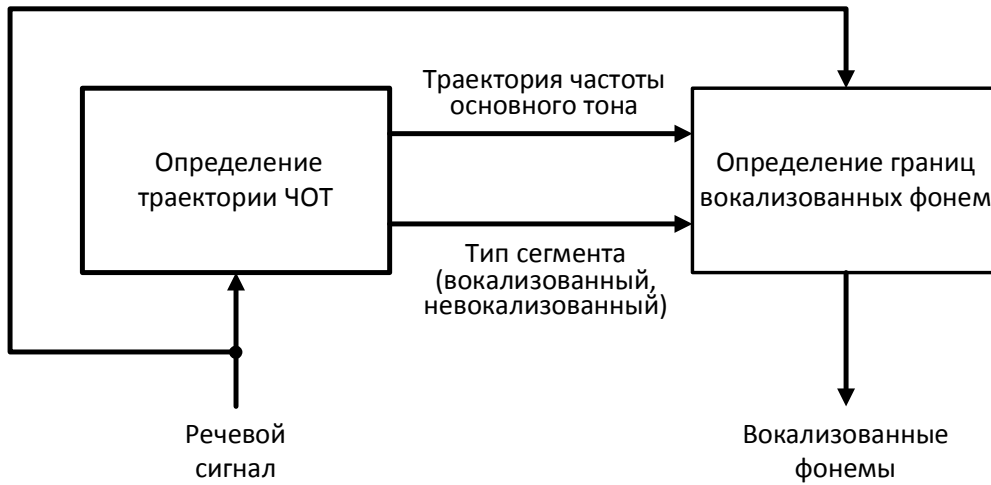


Рис. 1. Сегментация речевого сигнала на вокализованные фонемы.

Для определения частоты основного тона используется модифицированный корреляционный метод, основанный на скользящей обработке сигнала рядом убывающих по длительности окон, размеченных рабочими и штрафными зонами.

Корреляционный метод определения квазипериода колебаний основного тона (ОТ) оцифрованного речевого сигнала основан на выделении максимумов корреляционной функции, соответствующих границам квазипериодов.

Автокорреляционная функция речевого сигнала вычисляется в скользящем прямоугольном окне обработки по следующей формуле.

$$r(k) = \frac{1}{N \cdot \sigma^2} \sum_{i=1}^{N-k} (x_i - m)(x_{i+k} - m), \quad (1)$$

где $m = \frac{1}{N} \sum_{i=1}^N x_i$ – математическое ожидание сигнала в окне обработки;

$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - m)^2}$ – среднеквадратическое отклонение;

$k = \overline{0, N}$ – дискретные значения аргумента корреляционной функции, N – число дискретных отсчетов времени в окне обработки.

Однако, из-за большой нестабильности речевого сигнала величины названных пограничных максимумов автокорреляционной функции изменяются в широком диапазоне, что приводит к снижению помехоустойчивости.

Влияние нестабильности формы и размаха квазипериодических тональных колебаний, а также биений речевого сигнала на величины пограничных максимумов автокорреляционной функции может быть уменьшено путем автономного центрирования и нормализации каждого k -го коррелирующего отрезка сигнала с длиной $(N-k)$ дискретных отсчетов времени. Соответствующая модификация выражения автокорреляционной функции (1) заключается в следующем:

$$\rho(k) = \frac{1}{(N-k) \cdot \sigma_{1k} \cdot \sigma_{2k}} \sum_{i=1}^{N-k} (x_i - m_{1k})(x_{i+k} - m_{2k}), \quad (2)$$

где $m_{1k} = \frac{1}{N-k} \sum_{i=1}^{N-k} x_i$, $m_{2k} = \frac{1}{N-k} \sum_{i=1}^{N-k} x_{i+k}$ – матожидания первого и второго коррелирующих отрезков сигнала в окне обработки;

$$\sigma_{1k} = \sqrt{\frac{1}{N-k} \sum_{i=1}^{N-k} (x_i - m_{1k})^2}, \sigma_{2k} = \sqrt{\frac{1}{N-k} \sum_{i=1}^{N-k} (x_{i+k} - m_{2k})^2} -$$

среднеквадратичные отклонения;

$k = \overline{0, N}$ – дискретные значения аргумента корреляционной функции, N – число дискретных отсчетов времени в окне обработки.

Из-за убывания величин пограничных максимумов исходной автокорреляционной функции к концу окна обработки по формуле (1), достоверному оцениванию частоты основного тона F_0 по первому максимуму A_1 (рис. 2, б) препятствуют максимумы A_θ вложенных колебаний, которые могут превышать первый тональный максимум.

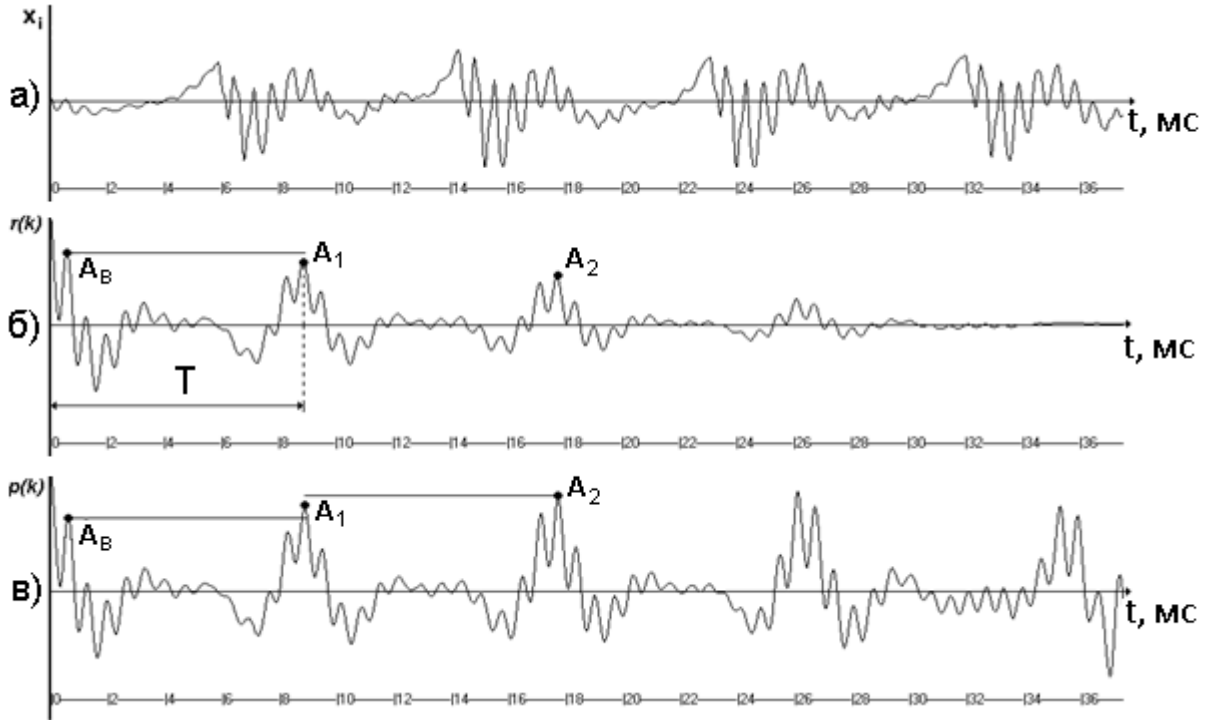


Рис. 2. Пример корреляционной обработки речевого сигнала: а – речевой сигнал; б – автокорреляционная функция; в – модифицированная корреляционная функция.

Применение модифицированной формулы (2) позволяет повысить стабильность и величины пограничных максимумов корреляционной функции (рис. 2, в). При этом существенно снижается вероятность влияния вложенных максимумов A_θ на достоверность оценки периода T , так как при обработке по формуле (2) обычно $A_1 > A_\theta$. Однако нестабильность речевого сигнала столь высока, что даже при его обработке по формуле (2) пограничные максимумы корреляционной функции изменяются в широком диапазоне $[0,5; 1]$ (рис. 2, в). Достоверность оценки можно существенно повысить путем следующей многооконной скользящей обработки и разбиения окон на несколько рабочих и штрафных зон.

Частота основного тона изменяется в широком диапазоне (бас 75-330 Гц, тенор 120-500 Гц, меццо-сопрано 170-700 Гц, сопрано 230-1100 Гц), поэтому целесообразно выполнять многооконную скользящую обработку сигнала, а длительности окон выбрать таким образом, чтобы самое короткое окно содержало не более 2-х периодов самого высокого сопрано, а самое длинное – не более 2-х периодов самого низкого баса. Всего разных окон обработки, длительности которых определяются рекуррентным соотношением $N_{k+1} = N_k \cdot \frac{N_k}{12}$, необходимо тридцать пять. Нумерация окон осуществляется от большего к меньшему. Длительность первого окна $N_0 = 40$ мс (рис. 3). Размеченные окна (рис. 3, б) используются при обработке модифицированной корреляционной функции (рис. 2, в).

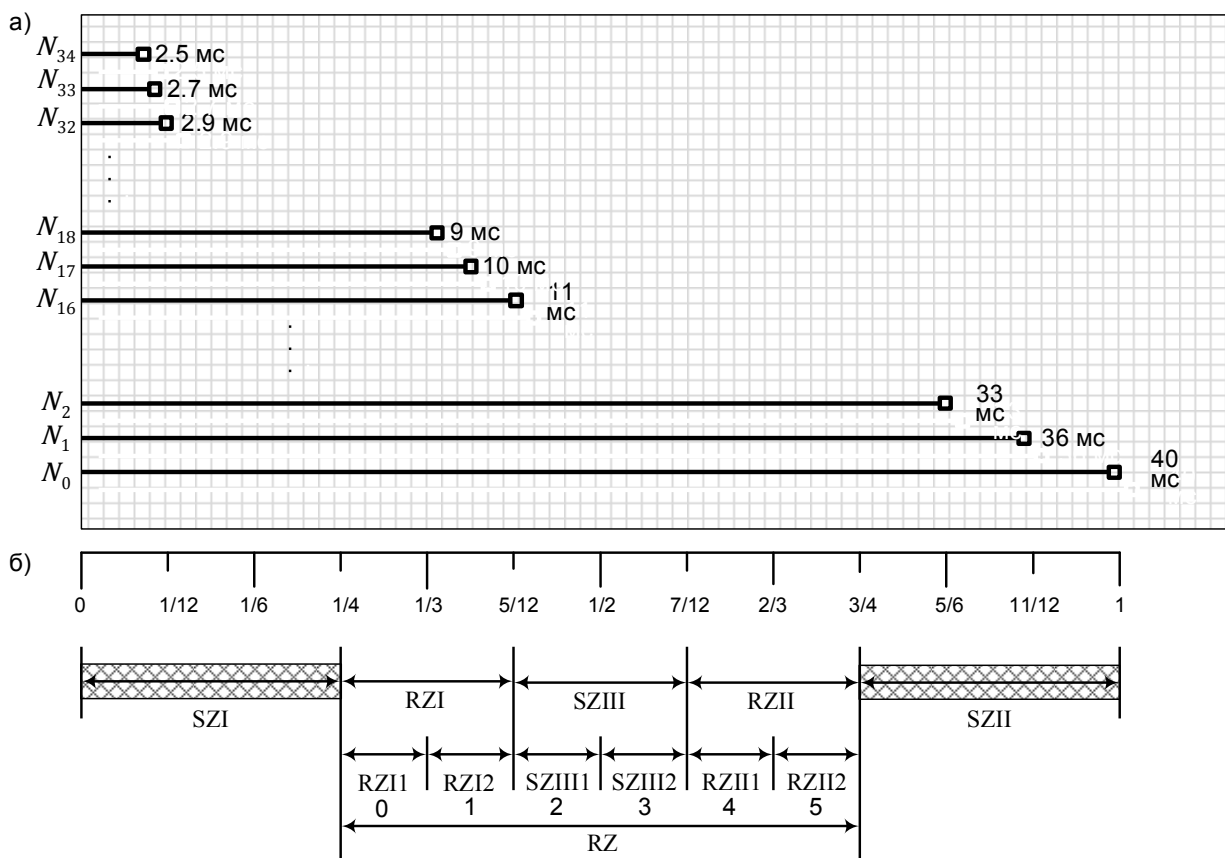


Рис. 3. Временные окна скользящей обработки сигнала (а) и разметка окна на рабочие RZ и штрафные SZ зоны (б).

Рабочие зоны RZI и RZII делятся на подзоны RZI1, RZI2 и RZII1, RZII2 соответственно и образует рабочую зону RZ. Каждая подзона составляет одну двенадцатую часть окна. Штрафные зоны – SZI, SZII, SZIII. Поиск глобального, в пределах окна, максимального экстремума корреляционной функции в штрафных зонах SZI или SZII не выполняется. Штрафная зона SZIII состоит из двух подзон – SZIII1, SZIII2.

Новизна разработанного метода оценивания длительности квазипериода T состоит в применении многооконной скользящей обработки и анализе попадания тональных максимумов $\{A_i\}$ в рабочие и штрафные зоны окон обработки.

Разработанный метод сегментации состоит в следующем. На вокализованных участках речи после оценки квазипериода ОТ начальное положение всех окон сдвигается во времени на величину найденного квазипериода $pitch$. При обработке невокализованных участков сдвиг окон осуществляется на постоянную минимальную величину, равную 1 мс.

Обработка речевого сигнала алгоритмом сегментации на вокализованные фонемы выполняется синхронно меткам основного тона. При этом вокализованные участки, длительность которых меньше минимальной длительности фонемы, отбрасываются и в дальнейшей обработке не участвуют. Минимальная длительность вокализованной фонемы D_{min}^{V+} принята равной 40 мс.

Начало и конец работы алгоритма сегментации задаются алгоритмом определения траектории тона по начальной отметке выхода на квазипериодический участок и конечной отметке прекращения квазипериодических колебаний. Границей предыдущей вокализованной фонемы считается начало текущего вокализованного участка.

В третьей главе разрабатываются алгоритмы, структурная и функциональные схемы устройства определения траектории основного тона и сегментации РС на вокализованные фонемы.

Конвейерное вычислительное устройство помехоустойчивой сегментации вокализованных участков речевого сигнала состоит из двух каскадов: устройства определения траектории основного тона (УОТ) и устройства сегментации вокализованных участков на вокализованные фонемы (УСВУ) (рис. 4).

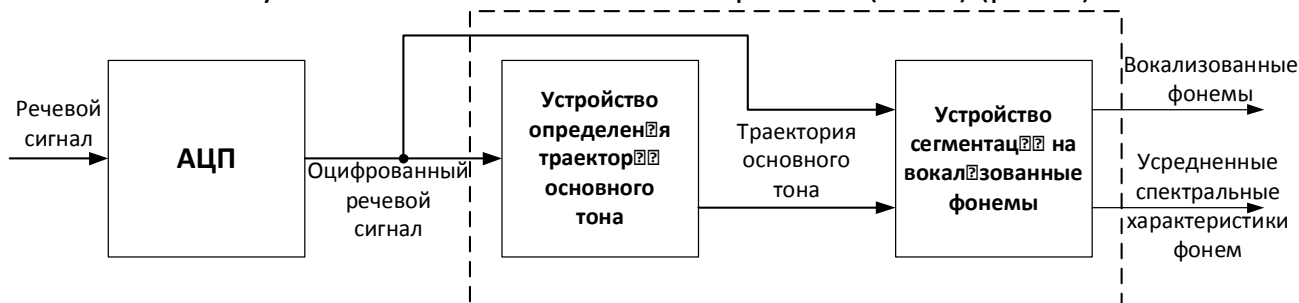


Рис. 4. Структурная схема конвейерного вычислительного устройства помехоустойчивой сегментации вокализованных участков РС

На основе описанных выше методов разработаны следующие алгоритмы функционирования устройств УОТ и УСВУ.

Алгоритм определения траектории частоты основного тона в каждый момент времени может находиться в одном из двух режимов: режим подбора оптимального окна (С) или режим адаптации (А).

Режим подбора оптимального окна (окно обработки установлено в начало обрабатываемого речевого сегмента):

C0. Текущий режим алгоритма $cur_mode = running$. Номер текущего окна обработки $cur_win_num = 0$.

C1. Вычислить корреляционную функцию в прямоугольном окне длительности $win_len_tab[cur_win_num]$, где win_len_num – таблица длительностей окон обработки.

C2. Если номер окна обработки $cur_win_num \geq 34$, то перейти к п. C11.

C3. Найти значения и аргументы экстремумов корреляционной функции в рабочих подзонах RZI1, RZI2, RZII1, RZII2 и штрафной зоне SZIII: $max11 = \max_{x \in RZI1}(x)$, $max12 = \max_{x \in RZI2}(x)$, $max21 = \max_{x \in RZII1}(x)$, $max22 = \max_{x \in RZII2}(x)$, $max3 = \max_{x \in SZIII}(x)$.

C4. $max1 = \max(max11, max12)$; $max2 = \max(max21, max22)$. Если $max1 < max2$, то п. C6.

C5. Если $max11 > max12$, то $max1 = max11$, $max2 = max21$, иначе $max1 = max12$, $max2 = max22$, п. C7.

C6. Если $max21 > max22$, то $max1 = max11$, $max2 = max21$, иначе $max1 = max12$, $max2 = max22$.

C7. Если в окне более одного периода, т.е. если $max3 > 0.75 \cdot max1$, то п. C10.

C8. Если $\frac{abs(max1-max2)+2-max1-max2}{max2^3} > 9.6$, то п. C10.

C9. Период основного тона $pitch = \arg(max1)$, оптимальное окно, содержащее только два периода основного тона $opt_win_num = cur_win_num$, участок речи, соответствующий метке периода, признается вокализованным $vocal = true$, а алгоритм переходит в режим адаптации $cur_mode = adapt$. Окно обработки сдвигается на $pitch$. Перейти к п. A1.

C10. Перейти к следующему окну $cur_win_num = cur_win_num + 1$, п. C2.

C11. Период основного тона $pitch = 1\text{мс}$, $vocal = false$, $cur_mode = running$. Окно обработки сдвигается на $pitch$. Перейти в п. C0.

Режим адаптации:

A0. Количество проходов алгоритма $cnt_run = 0$.

A1. Если $cnt_run > 2$ или если $opt_win_num < 0$ или $opt_win_num > 34$, то перейти к п. C0, иначе вычислить корреляционную функцию в прямоугольном окне длительности $win_len_tab[opt_win_num]$.

A2. Найти значения и аргументы экстремумов в рабочих зонах RZI, RZII и подзонах штрафной зоны SZIII: $max1 = \max_{x \in RZI}(x)$, $max2 = \max_{x \in RZII}(x)$, $max31 = \max_{x \in SZIII1}(x)$, $max32 = \max_{x \in SZIII2}(x)$.

A3. $max3 = \max(max31, max32)$. Если $max3 > max1$ и $max3 > max2$, то п. A4, иначе п. A5.

A4. Если $max31 > max32$, то $opt_win_num = opt_win_num - 1$, иначе $opt_win_num = opt_win_num + 1$, $cnt_run = cnt_run + 1$. Перейти к п. A1.

A5. Если $max3 > 0.9 \cdot max1$, то п. C0.

A6. Если $1 - \frac{\arg(max2) - \arg(max1)}{\arg(max2)} > 0.1$, то п. C0.

A7. $pitch = \arg(max1)$, $vocal = true$, $cur_mode = adapt$, окно обработки сдвигается на $pitch$, перейти к п. A0.

Алгоритм сегментации вокализованных участков на вокализованные фонемы:

S1. Для каждого k -ого квазипериода ОТ вычислить спектральную плотность $S_k(w)$ и среднюю спектральную плотность $\overline{S_{k-1}}(w)$. При расчете $\overline{S_{k-1}}(w)$ выполнить интерполяцию сплайнами каждого квазипериода ОТ так, чтобы их границы соответствовали k -ому квазипериоду ОТ. При вычислении средней

спектральной плотности первые $Q_{min}^{V+}/2$ квазипериодов пропускаются:

$$\overline{S_{k-1}}(w) = \frac{1}{k-1-\frac{Q_{min}^{V+}}{2}} \sum_{j=Q_{min}^{V+}/2}^{k-1} S_j(w + l\Delta w), l = 0, N-1.$$

S2. Определить значения сегментирующей функции как значение корреляции между текущей и средней спектральными плотностями:

$$\varphi(t_i) = r(S_i(w), \overline{S_{i-1}}(w)), \text{ где } r - \text{коэффициент взаимной корреляции.}$$

S3. На каждом шаге, после вычисления i – ой ординаты сегментирующей функции $\varphi(t_i)$ выполнить проверку условия:

если $|m_\varphi - \min(\varphi(t_{i+2}), \min(\varphi(t_i), \varphi(t_{i+1})))| > 4 \cdot \sigma_\varphi + c$, то границей b считается $(arg(\varphi(t_i)) + 2)$ -ой квазипериод ОТ. При этом

$$m_\varphi = \sum_{i=Q_{min}^{V+}/2}^{k-1} \frac{\varphi(t_i)}{k-1-Q_{min}^{V+}/2}; \sigma_\varphi = \sqrt{\frac{1}{k-1-Q_{min}^{V+}/2} \cdot \sum_{i=Q_{min}^{V+}/2}^{k-1} (\varphi(t_i) - m_\varphi)^2},$$

k – текущий обрабатываемый период, причем $k \geq Q_{min}^{V+}$; c – некоторая малая величина, исключающая появление ложной границы в случае малого σ_φ .

Дальнейший поиск границ продолжается с текущей границы, найденной в п. S3, при этом значения сегментирующей функции обнуляются. Алгоритм выполняется до тех пор, пока есть вокализованные участки, формируемые алгоритмом определения траектории основного тона.

В алгоритме используется относительная минимальная длительность фоны $Q_{min}^{V+} = D_{min}^{V+} \cdot \overline{F_0}$, где $\overline{F_0}$ – среднее значение частоты основного тона диктора, вычисленное по всем текущим значениям периодов траектории основного тона обрабатываемого РС.

Новизна разработанного метода и алгоритма сегментации заключается в нормализации спектральных плотностей $S(w)$ по частоте (п. S1), автоматическом подборе интервала усреднения спектральных плотностей квазипериодов (п.п. S1, S2) и специальных правилах локализации границы между соседними фонемами (п. S3).

На рис. 5 показана структурно-функциональная схема устройства определения траектории основного тона, представляющего собой шестиступенчатый конвейер. Условные обозначения выходных и выходных линий связи в блоках схемы соответствуют наименованиям переменных, принятых в приведенных формулах и алгоритмах, и показывают последовательность их конвейерного вычисления и способ аппаратной ориентации алгоритмов.

Новизна устройства состоит в том, что вторая и третья ступени конвейера синтезированы в виде параллельных пятиканальных вычислителей на основе следующего преобразования выражения модифицированной корреляционной функции (2), позволившего составить следующий рекурсивный вычислительный алгоритм потокового нахождения множества корреляционных функций сигнала в нескольких разных окнах скользящей обработки:

$$\rho(N, k) = \frac{S_5(N, k) - \frac{1}{N-k} \cdot S_1(N, k) \cdot S_2(N, k)}{\sqrt{\left(S_3(N, k) - \frac{1}{N-k} \cdot S_1(N, k)^2\right) \cdot \left(S_4(N, k) - \frac{1}{N-k} \cdot S_2(N, k)^2\right)}},$$

где $S_1(N, k) \dots S_5(N, k)$ – частичные суммы, определяемые следующими формулами: $S_1(N, k) = \sum_{i=1}^{N-k} x_i$; $S_2(N, k) = \sum_{i=1}^{N-k} x_{i+k}$; $S_3(N, k) = \sum_{i=1}^{N-k} x_i^2$; $S_4(N, k) = \sum_{i=1}^{N-k} x_{i+k}^2$; $S_5(N, k) = \sum_{i=1}^{N-k} x_i x_{i+k}$.

Таким образом, $\rho(N_i, k) = f(S_1(N_i, k), S_2(N_i, k), S_3(N_i, k), S_4(N_i, k), S_5(N_i, k))$, а $S_1(N, k) = S_1(N-1, k) + x_i$; $S_2(N, k) = S_2(N-1, k) + x_{i+k}$; $S_3(N, k) = S_3(N-1, k) + x_i^2$; $S_4(N, k) = S_4(N-1, k) + x_{i+k}^2$; $S_5(N, k) = S_5(N-1, k) + x_i x_{i+k}$. Соответственно, по каждому текущему значению k модифицированная корреляционная функция $\rho(N, k)$ может быть вычислена сразу для всех N , а затраты на вычисление $\rho(N, k)$ для следующего k сводятся к минимуму.

Основные временные затраты при вычислении значения каждого очередного квазипериода приходятся на блок 2 вычисления частичных сумм и блок 3 вычисления автокорреляционной функции (рис. 5). Для вычисления корреляционной функции в N_{max} окнах, с максимальной длительностью окна, равной N_{max} , требуется $Q_{corr} = \frac{1}{2} N_{max} (N_{max} - 3)$ тактов. Суммарное количество тактов, необходимых для нахождения одного периода основного тона РС устройством: $Q_{total} = Q_{corr} + Q_{БВМ} + Q_{БПР}$, где $Q_{БВМ}$ – количество тактов, необходимое для нахождения зональных максимумов в текущем окне (блок 5); $Q_{БПР}$ – количество тактов, необходимое для принятия решения о наличии периодичности в текущем окне обработки (блок 6). В худшем случае при работе на шумовом сигнале для обработки одной ординаты сигнала требуется Q_{total}/N_{1ms} тактов, где N_{1ms} – количество точек РС, соответствующего минимальной величине сдвига начала окон, равной одной миллисекунде. Для различных значений частоты дискретизации РС величины отношения Q_{total}/N_{1ms} приведены в таблице 1.

Таблица 1. Оценка быстродействия УОТ, требуемого для обработки в реальном времени при разных частотах дискретизации входного РС

Частота дискретизации, Гц	N_{max}	Q_{total}/N_{1ms}	Требуемая тактовая частота, МГц
22050	900	17582	388
11025	450	8443	93
8000	326	6397	51

Для сравнения показателей производительности был использован современный персональный компьютер с процессором Intel Pentium Dual-Core 1.73 ГГц. Для оценки быстродействия в качестве экспериментального сигнала использовался розовый шум с частотой дискретизации 22050 Гц и длиной выборки 220000 точек. Среднее время выполнения алгоритма составило 85.141 сек. Исходя из этого, расчетное количество тактов с учетом тактовой частоты процессора составило 668409 тактов на одну ординату сигнала. Моделирование алгоритма определения траектории частоты основного тона на сигнальном процессоре TMS320VC5402 с тактовой частотой 100 МГц проведено для РС с частотой дискретизации 8 КГц и длиной выборки 22000 точек. Время работы составило 199.829 секунд, соответственно, количество тактов, необходимых для обработки одной точки РС, составило 918317 тактов на одну ординату сигнала.

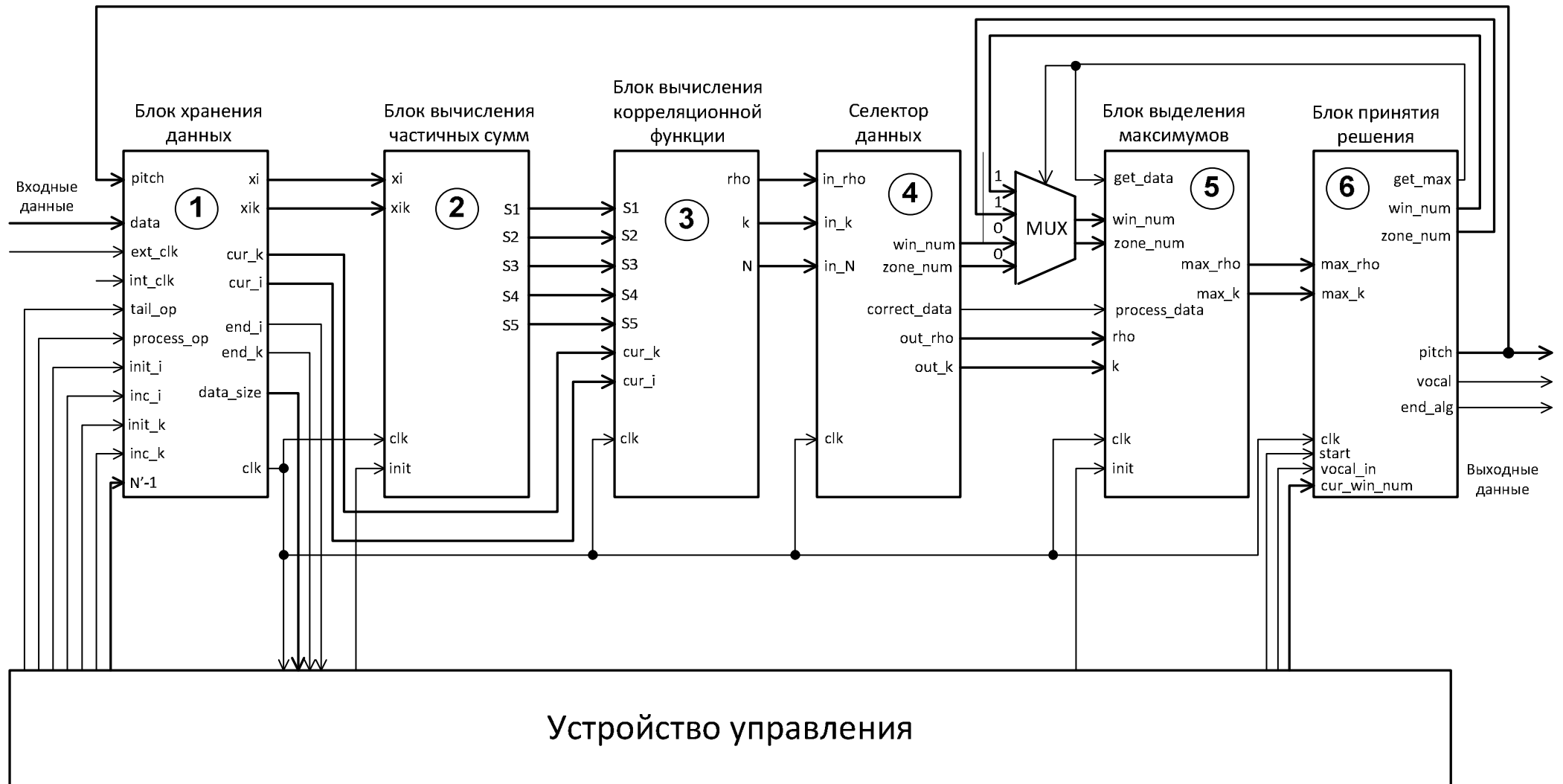


Рис. 5. Структурно-функциональная схема устройства определения траектории частоты основного тона

Таким образом, по соотношению чисел тактов работы, требуемых в худшем случае для обработки одной ординаты оцифрованного РС, разработанный параллельно-конвейерный вычислитель позволяет во-первых повысить производительность не менее чем в 30 раз, по сравнению с программной реализацией этих же алгоритмов в многоядерных микропроцессорах или цифровых процессорах сигналов и во-вторых выполнять первичную обработку РС в реальном времени при тактовых частотах 50...400МГц, существенно снизив за счет этого нагрев микросхем.

При реализации разработанного устройства УОТ для обработки речевых сигналов с частотой дискретизации 22КГц на современных ПЛИС семейств Xilinx Virtex-4, Xilinx Virtex-5. Аппаратная сложность УОТ составит 3805 слайсов и 34 матричных умножителя 18х18. Емкость требуемой памяти для УОТ – 84690 бит (рис. 5).

В устройстве УСВУ используются базовые алгоритмы цифровой обработки сигналов, такие как: вычисление спектральной плотности, интерполяция сплайнами и т.д. Поэтому УСВУ целесообразно реализовать на процессорах цифровой обработки сигналов.

В четвертой главе представлены результаты программного моделирования разработанных алгоритмов и проведена оценка помехоустойчивости сегментации РС на базе данных реальных речевых сигналов.

Выходные данные программной модели приведены на рис. 6. Квадратами отмечены границы вокализованных фонем, вертикальными линиями – границы квазипериодов. Траектория частоты основного тона представлена на графике ступенчатой линией.

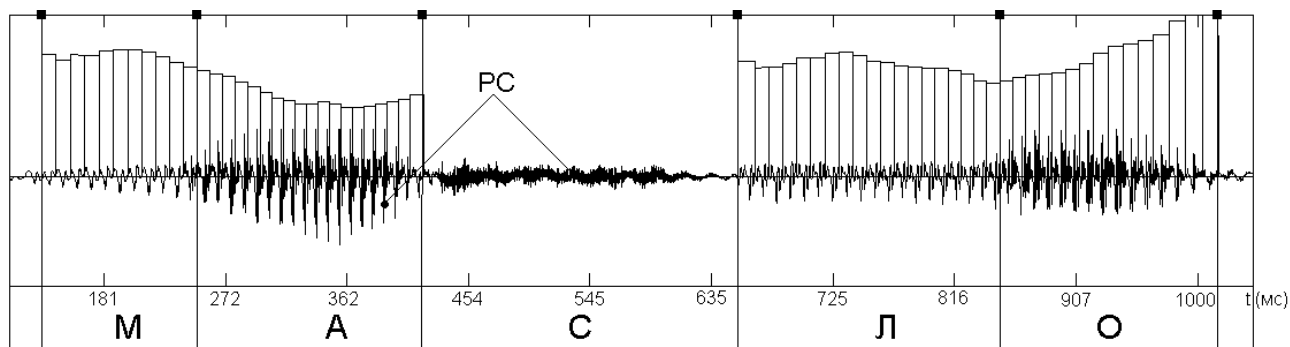


Рис. 6. Траектория частоты основного тона и разметка речевого сигнала на квазипериоды и вокализованные фонемы

Вычислительные эксперименты были поставлены на фонотеке реальных речевых сигналов, состоящей из 100 различных слов, произнесенных десятью дикторами мужского и женского полов. На каждое слово фонотеки накладывался «розовый» шум при отношении сигнал/шум 3, 0, -3 дБ. Начальный уровень шума в словах, использованных для восстановления эталонной траектории ОТ, соответствовал отношению сигнал/шум 8-10дБ. Оценка отношения сигнал/шум производилась по формуле:

$$SNR = 10 \cdot \log_{10} \left[\frac{\sum_{n=0}^{N-1} s^2(n)}{\sum_{n=0}^{N-1} e^2(n)} \right],$$

где s – ордината реального сигнала с начальным уровнем шума, e – ордината сигнала шума. Помехи типа «розового» шума (шума с тенденцией спада спектральной плотности 3 дБ на октаву в сторону высоких частот) использовались для зашумления исходных речевых сигналов в связи с тем, что они приводят к наибольшему снижению разборчивости речи.

Правильность функционирования системы определения мгновенных значений тональной частоты, множество значений которой образует траекторию ОТ, определялась следующими двумя показателями:

1. Погрешность восстановления траектории ОТ (δ).
2. Вероятность нахождения всех квазипериодов вокализованной фонемы (p).

Показатели оценивались по формулам:

$$\delta = \frac{\sqrt{\frac{1}{n_0} \sum_{i=1}^Q \delta_i^2}}{M_{F_0^*}}; p = n_0 / Q,$$

$$n_0 = \sum_{i=1}^Q \begin{cases} 0, \text{ если } \left(1 - \frac{F_{0i}}{F_{0i}^*}\right) > \varepsilon \\ 1, \text{ иначе} \end{cases}, \delta_i = \begin{cases} 0, \text{ если } \left(1 - \frac{F_{0i}}{F_{0i}^*}\right) > \varepsilon \\ (F_{0i} - F_{0i}^*), \text{ иначе} \end{cases}, M_{F_0^*} = \frac{1}{N} \sum_{i=1}^Q F_{0i}^*,$$

Q – эталонное количество квазипериодов, n_0 – количество выделенных квазипериодов ОТ с заданной погрешностью ε , F_{0i}^* – i -ое эталонное значение ординаты траектории ОТ, F_{0i} – i -ое измеренное значение ординаты траектории ОТ на заданном соотношении сигнал/шум, $M_{F_0^*}$ – математическое ожидание эталонной траектории ОТ, δ_i – модуль отклонения i -х значений ординат F_{0i}^* и F_{0i} . Результаты эксперимента приведены в таблице 2 (gma – условное обозначение разработанного метода, rpd – условное обозначение лучшего из известных методов, разработанного в центре цифровой обработки сигналов при Санкт-Петербургском государственном университете телекоммуникаций им. проф. М.А. Бонч-Бруевича, в 2005 году).

Таблица 2. Сравнительные результаты оценки помехоустойчивости методов определения траектории основного тона

SNR, дБ	gma		rpd	
	δ	p	δ	p
8	0.0150	0.96	0.0330	0.71
3	0.0177	0.69	0.0431	0.43
0	0.0187	0.51	0.0460	0.35
-3	0.0202	0.36	0.0580	0.26

По сравнению с известными методами и алгоритмами погрешность восстановления траектории ОТ снижена не менее чем в 2 раза, а вероятность ненахождения квазипериодов, равная $(1-p)$ – в 1.7...7 раз при отношении сигнал/шум SNR=3...8дБ и в 1.16...1.33 раза при SNR=-3...0дБ, или в среднем в 2 раза.

Оценка помехоустойчивости алгоритма сегментации вокализованных участков речевого сигнала приведена в таблице 3. Эталонная разметка речевых

сигналов в фонотеке была выполнена вручную. Фонема считалась найденной, если ее границы определены с точностью не менее чем в 2 квазипериода основного тона. Вероятность нахождения вокализованных фонем определялась как отношение количества найденных фонем к общему количеству фонем в фонотеке.

Таблица 3. Вероятность нахождения вокализованных фонем

SNR , дБ	всего вокализованных фонем	найденно вокализованных фонем	вероятность нахождения вокализованных фонем
8	498	473	0.95
3	498	354	0.71
0	498	319	0.64
-3	498	199	0.40

Вероятность определения ложных границ внутри участков вокализованных фонем – 0.11. Таким образом, разработанные методы и алгоритмы позволяют определять границы вокализованных фонем с вероятностью 0.4..0.95 и погрешностью не более двух квазипериодов основного тона в диапазоне изменения отношения сигнал/шум от –3дБ до 8дБ.

В **заключении** приведены основные результаты диссертационного исследования. В **приложении** приведены программы моделирования алгоритмов функционирования устройства, результаты тестирования помехоустойчивости алгоритма сегментации вокализованных участков на вокализованные фонемы, изображения речевых сигналов с восстановленной траекторией основного тона и изображения речевых сигналов, размеченных на вокализованные фонемы.

Основные результаты работы

В диссертации решена актуальная научно-техническая задача повышение помехоустойчивости аппаратно-ориентированных алгоритмов выделения в реальном времени траектории основного тона и вокализованных фонем. Получены следующие результаты:

1. Разработан модифицированный корреляционный метод определения траектории частоты основного тона речевого сигнала, отличающийся применением нескольких, убывающих по длительности, временных окон скользящей обработки, отслеживанием мгновенного значения квазипериода оптимальным окном и адаптацией его длительности к нестабильности частоты основного тона, позволяющий повысить помехоустойчивость и создать предпосылки для параллельного выполнения процедур оценивания основного тона.
2. Разработаны метод и вычислительный алгоритм сегментации вокализованных фонем, основанные на вычислении расстояния между нормализованными по частоте спектральными плотностями речевого сигнала в разных его квазипериодах и совместном использовании процедур определения траектории основного тона и сегментации фонем, позволившие повысить помехоустойчивость и вероятность нахождения границ между разными вокализованными фонемами.

3. Разработана структурно-функциональная организация специализированного вычислительного устройства помехоустойчивого определения траектории основного тона, отличающаяся параллельно-конвейерным вычислением нескольких корреляционных функций в 35 окнах скользящей обработки, позволяющая не менее чем в 30 раз повысить производительность по сравнению с программной реализацией в многоядерных микропроцессорах и цифровых процессорах сигналов.
4. Выполнено программное моделирование алгоритмов функционирования разработанного устройства. Проведены статистические исследования основных показателей их помехоустойчивости. Траектория основного тона воспроизводится с вероятностью 0.4...0.96 и погрешностью 0.02...0.015, границы между вокализованными фонемами находятся с вероятностью от 0.4 до 0.95 с погрешностью не более двух квазипериодов основного тона в диапазоне отношение сигнал/шум от -3дБ до 8дБ , что в среднем в 2 раза лучше показателей помехоустойчивости известных методов и алгоритмов.

Основные результаты диссертации опубликованы в следующих работах.

В изданиях из перечня ВАК:

1. Панов, А.Е. Расширение допустимых диапазонов искажения формы и изменения частоты тональных колебаний, выделяемых в речевом сигнале корреляционным методом [Текст] / А.Е. Панов, А.П. Типикин // Телекоммуникации. – 2003. - №11. – С. 13-16.
2. Типикин, А.П. Алгоритм определения границ между вокализованными фонемами [Текст] / А.П. Типикин, А.Е. Панов // Известия ТулГУ Серия выч. техн. Информационные технологии. Системы управления. Вып. 1. Вычислительная техника. – Тула: Изд-во: ТулГУ, 2006. – С. 16-21.

В других изданиях:

3. Панов, А.Е. Метод выделения мгновенных значений квазипериода основного тона речевого сигнала [Текст]/ А.Е. Панов, А.П. Типикин // Методы и средства систем обработки информации: сб. науч. ст. Вып. 4 Курск: Курск-ГТУ, 2007. – С. 70-76.
4. Панов, А.Е. Широкодиапазонное оценивание тона речевого сигнала корреляционным методом [Текст] / А.Е. Панов, А.П. Типикин // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Сб. мат-лов 6-й межд. конф. "Распознавание 2003" – Курск: 2003. – С.95-97.
5. Панов, А.Е. Метод определения границ вокализованных фонем по критерию грубого выброса [Текст] / А.Е. Панов, А.П. Типикин // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Сб. мат-лов 7-й межд. конф. "Распознавание 2005" – Курск: 2005. – С.42-44.
6. Panov, A.E. More robust pitch determination on speech signals [Текст] / A.E. Panov, A.P. Tipikin // Материалы международной конференции Information and telecommunication technologies in intelligent systems. ITT'IS 07 – Mallorca, 2007. – С.83-85.

7. Панов, А.Е. Параллельное вычислительное устройство определения траектории основного тона [Текст]/ А.Е. Панов, А.П. Типикин // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Сб. мат-лов 8-й межд. конф. "Распознавание 2008" – Курск: 2008. – С.41-43.
8. Панов, А.Е. Алгоритм помехоустойчивой сегментации речевого сигнала на вокализованные фонемы [Текст]/ А.Е. Панов, Е.Ю. Емельянова // Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Сб. мат-лов 8-й межд. конф. "Распознавание 2008" – Курск: 2008. – С.40-41.

Свидетельства о регистрации программ для ЭВМ:

9. Панов, А.Е. Программный комплекс с открытой архитектурой для широкодиапазонного выделения основного тона речевого сигнала // Свидетельство об официальной регистрации программ для ЭВМ №2006612757.
10. Панов, А.Е. Программный комплекс с открытой архитектурой для сегментации речевого сигнала на вокализованные фонемы // Свидетельство об официальной регистрации программ для ЭВМ №2006612549.

Соискатель

А.Е. Панов

Подписано в печать _____.____.08. Формат 60x84 1/16.

Гарнитура Calibri. Печать офсетная.

Усл. печ. л. _____. Тираж 100 экз. Заказ _____.

Курский государственный технический университет.

Издательско-полиграфический центр

Курского государственного технического университета.

305040, г. Курск, ул. 50 лет Октября, 94.