

**УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»**

УДК 004.934 + 004.4'277

АЗАРОВ
Илья Сергеевич

**ГАРМОНИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ
С ЧАСТОТНО-ВРЕМЕННЫМ МАСШТАБИРОВАНИЕМ
ДЛЯ СИСТЕМ ОБРАБОТКИ И СИНТЕЗА
ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ**

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

по специальности 05.13.17 – теоретические основы информатики

Минск 2009

Работа выполнена в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель – **Петровский Александр Александрович**, доктор технических наук, профессор, заведующий кафедрой электронных вычислительных средств учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Официальные оппоненты: **Листопад Николай Измайлович**, доктор технических наук, профессор, директор учреждения «Главный информационно-аналитический центр Министерства образования Республики Беларусь»

Хейдоров Игорь Эдуардович, кандидат физико-математических наук, доцент, доцент кафедры радиофизики Белорусского государственного университета

Оппонирующая организация – Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси»

Защита состоится 26 ноября 2009 г. в 16.00 на заседании совета по защите диссертаций Д 02.15.04 при учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г. Минск, ул. П. Бровки, 6, ауд. 232-1, тел. 293-89-89, e-mail: dissovvet@bsuir.by.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

Автореферат разослан 22 октября 2009 г.

КРАТКОЕ ВВЕДЕНИЕ

Гармонический анализ является средством выполнения временно-частотных преобразований и служит одним из основных инструментов при создании современных систем мультимедиа, обрабатывающих и передающих звуковые сигналы. Под гармоническим анализом (гармоническим преобразованием) понимается способ оценки параметров синусоидальной модели сигнала, который описывает заданный сигнал как сумму периодических функций (синусоид) с меняющейся во времени амплитудой и частотой. Ключевым вопросом при использовании синусоидальной модели является определение гармонических параметров, так как от точности их оценки зависит результат применения всего подхода в целом. Несмотря на то, что существует большое количество способов для выполнения гармонического анализа, это научное направление постоянно совершенствуется, так как требуемая во многих приложениях точность анализа пока не достигнута.

Наряду с синусоидальной и гармонической моделями, одним из наиболее популярных средств кодирования и анализа речи является линейное предсказание, используемое в системах компрессии, синтеза речи по тексту, конверсии голоса и многих других приложениях. Линейное предсказание обладает рядом полезных свойств, к которым относятся оценка формантных траекторий и оценка спектральной огибающей. Данные свойства обеспечивают его применимость в декодерах, системах идентификации диктора и синтеза речи по тексту. Линейное предсказание описывает некоторое общее поведение сигнала на всем окне анализа, не всегда обеспечивая достаточную временную локализацию. Подобная особенность сильно затрудняет анализ сигналов с частотной модуляцией средствами линейного предсказания. Так как частотно-модулированные (ЧМ) сигналы хорошо представимы гармонической моделью, целесообразно ее объединение с линейным предсказанием. Объединение двух моделей является перспективным направлением научного исследования, так как позволит расширить область применения линейного предсказания в задачах цифровой обработки сигналов. Более того, оно может дать дополнительную физическую интерпретацию синусоидальному описанию сигнала.

Полученные научные результаты имеют практическое применение в приложениях мультимедийных и промышленных систем.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с крупными научными программами и темами

Диссертационная работа выполнена в соответствии с научно-техническими заданиями и планами работ кафедры "Электронные вычислительные средства" и НИЛ 3.1 по проекту фонда фундаментальных исследований "Системы параметри-

ческого кодирования аудио/речевых сигналов: методы синтеза и анализа, реализация на реконфигурируемой вычислительной платформе" (№Т08МС040), НИР "Конверсия голоса на основе гармонического преобразования для применения в системах мультимедиа" на основании гранта Министерства Образования Республики Беларусь (№ГР20090518).

Цель и задачи исследования

Цель работы – разработка специальных методов и алгоритмов для вычисления мгновенных гармонических параметров речевых и звуковых сигналов и использования данного параметрического представления в системах обработки звуковой информации.

Для достижения поставленной цели предусмотрено решение следующих задач:

1. Проанализировать существующие способы вычисления гармонических параметров с целью выбора наиболее современного и эффективного подхода в качестве отправной точки исследования.
2. Разработать специальные методы точной оценки гармонических параметров, локализованных во временной и частотной областях, для достижения точной параметризации сигнала и адекватного его разделения на шумовую и периодическую составляющие.
3. Разработать специальные методы для вычисления параметров голосового тракта на основе гармонической модели с целью повышения разрешения формантного анализа в речевых приложениях.
4. Создать программную реализацию предлагаемых методов анализа и оценить их применимость для сигналов с различным содержанием периодических и шумовых компонент.
5. Программно реализовать способы обработки параметрического представления сигнала, позволяющие выполнять модификации временного масштаба, тона и тембра в контексте следующих систем мультимедиа: кодирование аудио и речи, конверсия голоса, обработка вокала, синтез речи по тексту и редактирование шума.

Объектом исследования являются системы мультимедиа, использующие синусоидальную или гармоническую модель представления сигнала. Предмет исследования – методы гармонического преобразования, позволяющие получить параметрическое представление сигнала в виде набора мгновенных параметров.

Решение рассматриваемых в диссертации задач требует применения методов цифровой обработки сигналов, математического моделирования и объектно-ориентированного программирования.

Положения, выносимые на защиту

1. Метод синтеза фильтров анализа, отличающийся тем, что импульсная характеристика полученного фильтра может быть модулирована в соответствии с заданным частотным контуром, обеспечивая точную оценку периодических компонентов с резким изменением частоты. На выходе фильтра формируется аналитический сигнал, из которого могут быть непосредственно вычислены мгновенные гармонические параметры (амплитуда, частота, фаза).

2. Метод оценки контура частоты основного тона, основанный на использовании фильтров анализа, отличающийся тем, что позволяет провести предварительную оценку контура в узкой низкочастотной полосе сигнала, а затем уточнить полученный результат в процессе анализа гармоник основного тона.

3. Методы гармонического анализа звуковых и речевых сигналов, позволяющие оценивать мгновенные гармонические параметры с повышенной точностью, учитывая частотные модуляции периодических составляющих сигнала.

4. Алгоритм гармонического анализа звуковых и речевых сигналов, отличающийся тем, что он позволяет выделить продолжительные и устойчивые гармонические составляющие сигнала без его предварительной обработки.

5. Метод преобразования мгновенных гармонических параметров в коэффициенты линейного предсказания, отличающийся тем, что он обеспечивает локализованное (мгновенно) описание сигнала при помощи коэффициентов линейного предсказания.

6. Методы параметрического описания сигнала для создания систем обработки звука, вокала и речи, позволяющие изменять просодические параметры с низким уровнем слышимых артефактов, а так же метод конкатенации речевых сегментов для системы синтеза речи по тексту, отличающийся тем, что он позволяет синтезировать речевой сигнал с ослабленным эффектом «компьютерного акцента».

Личный вклад соискателя

Все результаты и положения, выносимые на защиту, получены автором лично. Основной вклад научного руководителя Петровского А.А. заключался в постановке цели и задач исследования, в выборе соответствующих методов исследования, в разработке общей концепции гармонического анализа на основе частотно-модулированных фильтров. В публикациях с соавторами вклад соискателя определяется рамками излагаемых в диссертации результатов.

Апробация результатов диссертации

Результаты выполненных работ докладывались и обсуждались на: 12-й и 13-й Международных конференциях “Современные средства связи” (Нарочь, Бе-

ларусь, 2007, Минск, Беларусь, 2008); the 5-th international conference “Neural networks and artificial intelligence” (Минск, Беларусь, 2008); 11-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2008); the 16-th and 17-th European signal processing conferences “EUSIPCO 2008” and “EUSIPCO 2009” (Lausanne, Switzerland, 2008, Glasgow, Scotland 2009); the 13-th international conference “Speech and Computer” (Saint-Petersburg, Russia, 2009); the 12-th international conference on digital audio effects “DAFx-09” (Como, Italy, 2009); the 10-th international conference “Pattern Recognition and Information Processing” (Minsk, Belarus, 2009); the 126-th audio engineering society convention (Munich, Germany, 2009); научно-технических конференциях БГУИР (Минск, Беларусь, 2006–2008).

Результаты диссертационной работы внедрены и используются:

- Государственным научным учреждением “ОИПИ НАН Беларуси” (лаборатория “Распознавания и синтеза речи”);
- компанией «Сакрамент ИТ», являющейся резидентом Парка высоких технологий;
- ЗАО «Ай-Ти Мобайл» г. Москва;
- в учебном процессе специальности “Электронные вычислительные средства” (БГУИР).

Опубликованность результатов диссертации

Основные результаты диссертации опубликованы в 15 научных работах, в том числе: 3 статьи в научно-технических журналах, рекомендованных ВАК для опубликования результатов диссертационных исследований, 7 статей в сборниках материалов научно-технических конференций и симпозиумов, 2 тезиса докладов на научных конференциях, 1 глава монографии, 1 препринт и 1 статья в сборнике трудов. Суммарный объем публикаций составляет около 7 авторских листов. Результаты работы включены в 3 отчёта о НИР.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, пяти глав, заключения, библиографического списка из 123 использованных литературных источников, списка 15 публикаций соискателя и 3 приложений. Полный объем диссертации составляет 160 страниц. Из них: 54 иллюстрации на 50 страницах, 11 таблиц на 11 страницах, библиографический список - 12 страниц, 3 приложения на 22 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во *введении* обоснована актуальность темы диссертации, показана необходимость разработки и исследования методов и алгоритмов гармонического анализа для устройств мультимедиа, определены основные направления исследований.

Первая глава посвящена анализу существующих способов вычисления гармонических параметров с целью выбора наиболее современного и эффективного подхода в качестве отправной точки исследования. Показано, что синусоидальное (гармоническое) представление сигнала широко применяется в современных системах обработки, так как с его помощью можно эффективно описывать периодические компоненты, составляющие значительную часть различных звуков и речи. Рассмотрены методы оценки гармонических параметров и коэффициентов линейного предсказания.

Так как синусоидальное моделирование в большинстве случаев можно применить непосредственно к входу системы, входной сигнал удобно рассматривать как комбинацию периодической и остаточной компонент. Сигнал $s(n)$ можно записать в виде соотношения:

$$s(n) = \sum_{k=1}^K \text{MAG}_k(n) \cos \varphi_k(n) + r(n), \quad (1)$$

где $\text{MAG}_k(n)$ – мгновенная амплитуда k -й синусоиды;

K – число синусоид;

$\varphi_k(n)$ – мгновенная фаза k -й синусоиды;

$r(n)$ – сигнал-остаток.

Мгновенная фаза $\varphi_k(n)$ и мгновенная частота $f_k(n)$ соотносятся следующим образом:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0) \quad (2)$$

где F_s – частота дискретизации;

$\varphi_k(0)$ – начальная фаза k -й синусоиды.

Наряду с синусоидальной моделью широко применяется гармоническая модель, которая предполагает, что значения мгновенных частот $f_k(n)$ являются кратными частоте основного тона $f_0(n)$. Данная модель используется для кодирования речевых сигналов с высоким коэффициентом сжатия, так как обеспечивает чрезвычайно эффективное описание вокализованных фрагментов речи.

Задачей гармонического анализа является определение синусоидальных параметров $\text{MAG}_k(n)$, $f_k(n)$ и $\varphi_k(n)$ для заданного момента или интервала времени.

Считается, что амплитуда и фаза синусоидальных компонент изменяется медленно, а искомые параметры синусоидальной модели $MAG_k(n)$ и $f_k(n)$ являются гладкими, непрерывными функциями с ограниченным частотным диапазоном.

В качестве основного инструмента для выполнения гармонического анализа используется преобразование Фурье, однако при его помощи сложно получить адекватное параметрическое описание неустойчивых тональных звуков. Другая проблема заключается в сложности анализа сигналов с быстро изменяющимся тоном. Для того чтобы локализовать частотно-модулированные гармонические компоненты применяется гармоническое преобразование, которое является преобразованием Фурье с временным масштабированием. Методы мгновенного гармонического анализа позволяют достичь максимально высокого разрешения одновременно во временной и частотной областях, однако необходимость предварительного применения узкополосной фильтрации накладывает серьезные ограничения.

Линейное предсказание, наряду с синусоидальной моделью, является одним из самых используемых подходов в задачах цифровой обработки речи. Для оценки коэффициентов предсказания используются два основных метода: автокорреляционный и ковариационный.

Используя гармоническую модель, локализованный сигнал можно описать при помощи линейного предсказания с мгновенными коэффициентами. Это позволяет значительно улучшить временное разрешение модели и выполнять поиск мгновенных формантных траекторий, что положительно скажется на приложениях, использующих формантные оценки.

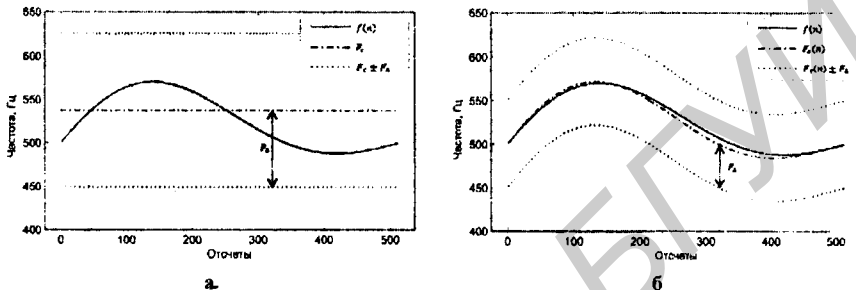
Анализ используемых гармонических преобразований в системах обработки звука и речи показывает, что для повышения точности оценок требуется применение комбинированных методов, позволяющих получать мгновенные гармонические параметры нестационарных сигналов. Показаны ограничения каждого из подходов, которые усложняют либо делают неприемлемым их использование в том или ином приложении.

Для связи синусоидальной модели с линейным предсказанием с целью получения локализованной оценки параметров вокального тракта, необходим поиск решения, позволяющего минимизировать ошибку предсказания периодических сигналов. В связи с вышеизложенным, использование модификации существующих способов определения коэффициентов предсказания с учетом свойства периодичности является наиболее перспективным направлением.

Вторая глава посвящена описанию разработанного метода оценки мгновенных гармонических параметров сигнала. Предложен способ синтеза специального фильтра анализа, обеспечивающего правильную оценку мгновенных параметров компонентов с частотной модуляцией. Данный подход объединяет в себе элемен-

ты преобразований со встроенным временным масштабированием и методов оценки на основе аналитических сигналов.

Для получения модулированной импульсной характеристики ширина полосы пропускания F_{Δ} принимается постоянной, а центральная частота полосы рассматривается как функция от времени, $F_c(n)$ (рисунок 1).



а – стационарный фильтр; б – модулированный фильтр

Рисунок 1 – Мгновенный гармонический анализ с использованием различных фильтров

Мгновенные гармонические параметры фильтрованного сигнала определяются по формулам:

$$\text{MAG}(n) = \sqrt{A^2(n) + B^2(n)}; \quad (3)$$

$$\varphi(n) = \arctan\left(\frac{-B(n)}{A(n)}\right); \quad (4)$$

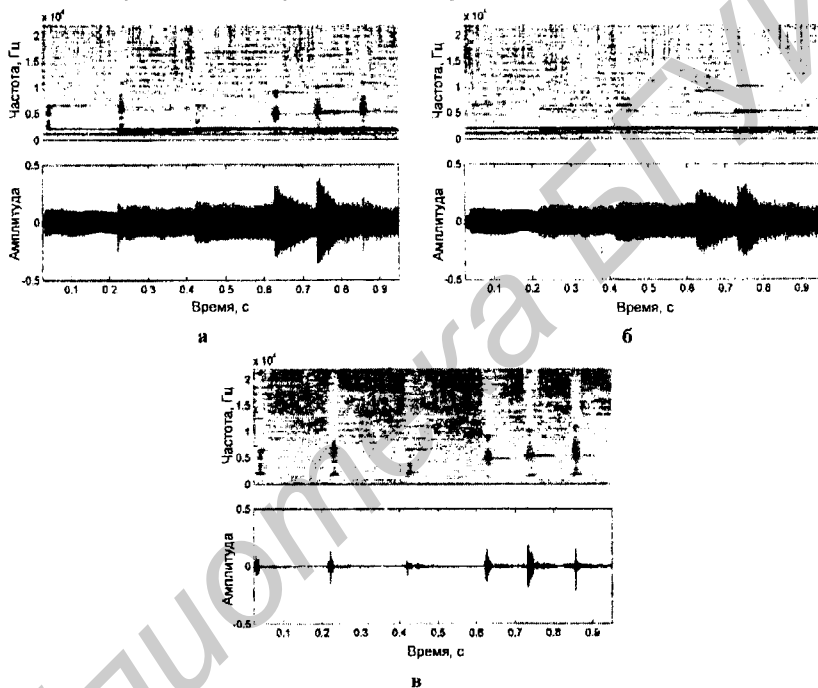
$$f(n) = \frac{\varphi(n+1) - \varphi(n)}{2\pi} F_s; \quad (5)$$

где
$$A(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_{\Delta}} \sin\left(\frac{2\pi(n-i)}{F_s} F_{\Delta}\right) \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right);$$

$$B(n) = \sum_{i=0}^{N-1} \frac{-s(i)F_s}{2\pi(n-i)F_{\Delta}} \sin\left(\frac{2\pi(n-i)}{F_s} F_{\Delta}\right) \sin\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right).$$

Предложен способ оценки мгновенных параметров звуковых и речевых сигналов. При обработке звуковых сигналов анализатор выполняет оценку парамет-

ров с постобработкой для выделения продолжительных периодических компонент. Для локализации периодических компонент в частотной области в синусоидальном анализаторе используется итеративный перерасчет оцениваемых частот. После того, как все энергетические максимумы локализованы, оцениваются требуемые синусоидальные параметры (амплитудная огибающая, мгновенная частота и фаза). Результат разделения звукового сигнала на периодическую и остаточную компоненты путем оценки гармонических параметров приведен на рисунке 2.



а – исходный сигнал; б – периодическая составляющая;
в – остаточная составляющая

Рисунок 2 – Разделение звукового сигнала на периодическую и остаточную составляющие гармоническим анализатором

При обработке речевого сигнала процедура оценки состоит из двух основных этапов:

- оценка частоты основного тона и классификация фрейма как вокализованного или невокализованного;

– оценка гармонических параметров с уточнением частоты основного тона.

Частота основного тона определяется предварительно, чтобы найти приближительные частотные траектории гармоник, необходимые для синтеза частотно-модулированных фильтров. Задача оценки частоты основного тона сводится к нахождению тонального компонента с минимальной частотой и достаточно высокой энергией. Диапазон изменения частоты основного тона в речевых приложениях обычно находится в интервале 60-420 Гц. Частоты периодических компонент речевого сигнала в этом диапазоне изменяются довольно медленно, что делает возможным их оценку с использованием стационарных фильтров. Для того чтобы облегчить процесс вычисления алгоритма оценки, можно воспользоваться ограниченностью частотного интервала поиска и применить предварительную низкочастотную фильтрацию с децимацией. В том случае, когда не удается выделить продолжительный синусоидальный компонент, соответствующие фреймы сигнала классифицируются как невокализованные. После оценки каждой гармоники контур частоты основного тона уточняется с учетом полученных параметров.

В *третьей главе* решается задача конверсии параметров гармонической модели в коэффициенты предсказания, которая формулируется следующим образом. Для наборов гармонических амплитуд $МАG_1(n), \dots, МАG_K(n)$ и частот $f_1(n), \dots, f_K(n)$ требуется найти набор коэффициентов предсказания $a_1(n), \dots, a_p(n)$, который обеспечивает спектральную огибающую $\Lambda(e^{i\omega})$, максимально аппроксимирующую исходный набор гармонических параметров.

Используя свойство периодичности сигнала ошибки предсказания для данного вектора амплитуд $МАG_1(n), \dots, МАG_K(n)$ и вектора частот $f_1(n), \dots, f_K(n)$, его энергию можно минимизировать на бесконечном интервале (p – порядок предсказания):

$$E_a^2 = \sum_{k=1}^K МАG_k(n)^2 \left(\left[1 - \sum_{i=1}^p a_i \cos(f_k(n)i) \right]^2 + \left[\sum_{i=1}^p a_i \sin(f_k(n)i) \right]^2 \right), \quad (6)$$

что приводит к следующей системе вычисления коэффициентов предсказания:

$$\sum_{i=1}^p a_i q(|i-j|) = -q(j), \quad (7)$$

где $j = 1, 2, \dots, p$;

$$q(l) = \sum_{k=1}^K МАG_k(n) \cos(f_k(n)l), \quad (l \geq 0).$$

Для получения устойчивого фильтра-предсказателя выводится следующая система:

$$\sum_{i=1}^p a_i q(|i-j|) = -q(j), \quad (8)$$

где $j = 1, 2, \dots, p$;

$$q(l) = \begin{cases} \sum_{k=1}^K \text{MAG}_k(n) \cos(f_k(n)l) + \Delta d, & l = 0 \\ \sum_{k=1}^K \text{MAG}_k(n) \cos(f_k(n)l), & l > 0 \end{cases}.$$

Значение Δd определяет, насколько близко будет фильтр к нестабильности в случае, если система (7) дает в результате фильтр-предсказатель $A(z)$ с корнями, лежащими на единичной окружности; данный параметр так же влияет на точность предсказания. Для оценки эффективности предлагаемого метода конверсии приводятся результаты его практического применения к продолжительному отрезку речи.

Предложенный метод конверсии так же может применяться для оценки коэффициентов предсказания непосредственно из преобразования Фурье.

В *четвертой главе* описана структура гармонического анализатора для оценки параметров звуковых и речевых сигналов. Предложены соответствующие алгоритмы анализа. Приведены результаты тестирования гармонического анализатора. Для тестирования качества сепарации сигналов на периодическую и шумовую компоненты применялась специальная процедура. В качестве тестовых сигналов использовались синтезированные звуки, представляющие собой комбинацию гармонической и шумовой составляющих, соотношение которых задается коэффициентом отношения гармоника/шум $HNR = 10 \lg \frac{\sigma_H^2}{\sigma_s^2}$, где σ_H^2 – энергия гармонической составляющей тестового сигнала; σ_s^2 – энергия шумовой компоненты тестового сигнала.

В процессе тестирования выполнялась сепарация сигналов с использованием различных частотно-временных преобразований (дискретное преобразование Фурье с адаптированным частотным разрешением, дискретное преобразование Фурье с временным масштабированием, узкополосная фильтрация с последующим преобразованием Гильберта, предлагаемое гармоническое преобразование на основе стационарных и модулированных фильтров) и вычислялись следующие оценки:

– отношение энергии периодической составляющей, синтезированной при помощи оценочных значений гармонических параметров, $\tilde{\sigma}_H^2$ к энергии шумовой компоненты σ_e^2 : $\bar{HNR} = 10 \lg \frac{\tilde{\sigma}_H^2}{\sigma_e^2}$;

– отношение энергии периодической компоненты тестового сигнала σ_H^2 к энергии синтезируемой периодической компоненты $\tilde{\sigma}_H^2$: $O/E = 10 \lg \frac{\sigma_H^2}{\tilde{\sigma}_H^2}$;

– отношение сигнал/шум синтезируемой периодической компоненты $SNR_H = 10 \lg \frac{\tilde{\sigma}_H^2}{\sigma_{eH}^2}$, где σ_{eH}^2 - энергия сигнала, вычисляемого как разность между

исходной периодической частью тестового сигнала и синтезированной при помощи оценочных значений.

Результаты эксперимента показали, что в речевых приложениях предлагаемый способ вычисления гармонического преобразования является наиболее предпочтительным, так как он позволяет выполнять наиболее точное разделение периодической и шумовой составляющих нестационарного сигнала (предлагаемый метод обеспечил более высокий уровень SNR_H на 2,5-15dB), исключение составило сильно зашумленный сигнал ($HNR < 10dB$), однако на практике сигнал с таким уровнем шума обычно требует применения специальных методов шумоподавления перед выполнением гармонического анализа вне зависимости от применяемого способа частотно-временного преобразования.

Показано применение анализатора в качестве входного блока гибридного кодера. Описаны специальные способы обработки гармонических параметров для получения звуковых эффектов. Приведена структура системы модификации голоса с гибридным представлением сигнала, в которой экспериментально используется разработанный гармонический анализатор.

Представлено применение анализатора для задачи конверсии голоса. Для этого случая реализована специальная система конверсии. Приводятся полученные результаты.

Система обработки голоса может быть разделена на три основных функциональных блока: блок гармонического анализа, блок обработки и блок синтеза. Блок гармонического анализа выполняет оценку гармонических параметров с разделением входного сигнала на периодическую и остаточную части - рисунок 3. Все модификации голоса выполняются в блоке обработки, причем периодическая часть и остаточная обрабатываются отдельно. Периодическая часть сигнала обра-

бающую при помощи интерполяции. Исходный вокализованный фрагмент анализируется гармоническим анализатором, затем огибающие спектра вычисляются с использованием мгновенных значений гармонических амплитуд и частоты основного тона.

Экспериментальное применение изложенных методов для конверсии голоса, позволяет полностью заменить индивидуальность исходного диктора путем изменения спектральных огибающих и контура частоты основного тона. Спектральные огибающие модифицировались при помощи кодовой книги, которая формировалась способом эффективного обучения. Выходной сигнал сохранял разборчивость, что так же подтверждает целесообразность и действенность предлагаемого подхода. Результаты практического применения свидетельствуют о том, что на основе гибридной модели представления сигнала возможно создание ряда высококачественных систем мультимедиа, таких как голосовые процессоры эффектов, голосовые конверторы, распознаватели речи, кодеры и декодеры.

В *главе пять* разработанный способ гармонического анализа применяется для модификации системы речи по тексту. Предлагается использование специального акустического процессора для конкатенации речевых сегментов в параметрической форме. Акустическая база описывается при помощи гармонической модели. Описание базы данных гармонической моделью является перспективной альтернативой методу конкатенации во временной области, так как позволяет:

- осуществлять непосредственный контроль над просодическими характеристиками речевого сигнала;
- выполнять высококачественное соединение речевых сегментов при помощи простого линейного сглаживания;
- существенно сжимать акустическую базу данных;
- выполнять синтез выходного сигнала с низкими вычислительными затратами, используя целочисленную арифметику.

Синтезатор состоит из трех основных блоков: блока изменения просодических характеристик, блока конкатенации и блока синтеза. Блок изменения просодических характеристик выполняет изменение контура частоты основного тона сегментов, изменение временного масштаба времени и энергии согласно с данными, формируемыми просодическим процессором. Блок конкатенации производит соединение сегментов со сглаживанием параметров на границах. Блок синтеза выполняет интерполяцию и непосредственный синтез волновой формы выходного сигнала.

Для оценки общего качества предлагаемой системы синтеза речи по тексту были подготовлены две акустические базы данных: одна для диктора с мужским голосом и вторая для диктора с женским голосом. При помощи каждой из полу-

ченных баз были синтезированы контрольные фразы. Для сравнения эти же фразы были синтезированы при помощи конкатенации во временной области. Для синтеза в обоих случаях использовались одни и те же правила конкатенации и изменения просодических параметров. Контрольные фразы были синтезированы с частотой дискретизации 22 кГц.

Полученные результаты позволили установить, что предлагаемый метод синтеза производит речь, которая ближе к оригиналу по сравнению с методом конкатенации во временной области. Более того, интонации звучат более натурально, благодаря усовершенствованному способу изменения просодики. Разница между предлагаемым и используемым методами наиболее заметна для женского голоса, так как конкатенация во временной области плохо подходит для фрагментов с высокой частотой основного тона и требует специальной коррекции для получения хороших результатов.

Проведенные эксперименты показали высокий потенциал предлагаемого подхода, так как он позволяет получать более высокое качество синтезированного сигнала по сравнению с подходами, основанными на конкатенации во временной области.

В *приложениях* к диссертации приведены технические характеристики и программная реализация основных функций гармонического анализатора, описание метода моделирования переходных фрагментов сигнала методом согласованной подгонки, использованного в системе гибридного кодера. Приведены акты внедрения результатов диссертационной работы.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. На основе исследования существующих способов гармонического анализа применительно к системам обработки речи и звука показано, что для достижения высокой точности анализа требуется применение специальных комбинированных методов, позволяющих оценивать мгновенные параметры сигналов с частотной модуляцией [3-А, 13-А, 14-А].

2. На основе анализа существующих подходов к формантному и спектральному оцениванию при помощи методов линейного предсказания показано, что возможно существенное повышение точности оценок частотно-модулированных сигналов при использовании гармонической модели описания сигнала с последующей конверсией параметров в коэффициенты линейного предсказания [1-А, 13-А].

3. Предложен способ точной оценки гармонических параметров звуковых и речевых сигналов, позволяющий локализовать описание сигнала, как во временной, так и в частотной областях. Метод использует специальный фильтр анализа, позволяющий оценивать мгновенные параметры синусоидальных компонент сигнала [2-А, 4-А, 8-А, 11-А, 15-А]. Импульсная характеристика фильтра синтезируется при помощи аналитических выражений и представляет собой непрерывную функцию, которая зависит от полосы пропускания, определяемой заданными границами частотного диапазона. Выражение импульсной характеристики получено в форме, позволяющей синтезировать частотно-модулированный фильтр, учитывающий изменения частоты анализируемых компонент. Таким образом становится возможным применение узкополосной фильтрации высокого порядка к синусоидальным компонентам с резкими изменениями частотного контура. Оценка мгновенных гармонических параметров позволяет выполнять точное разделение сигнала на периодическую и остаточную части.

4. Предложен метод конверсии мгновенных гармонических параметров в коэффициенты линейного предсказания, который может использоваться для вычисления параметров голосового тракта на основе гармонической модели [1-А, 5-А]. Метод позволяет выполнять мгновенный формантный анализ в речевых приложениях. Исследована устойчивость фильтра-предсказателя. Получено решение в виде системы линейных уравнений, которая позволяет получить устойчивый фильтр. Используя разработанный метод конверсии, коэффициенты линейного предсказания могут быть получены непосредственно из преобразования Фурье, что может быть полезным в приложениях перцигуального кодирования, анализа и синтеза речи.

5. Предложенный метод гармонического анализа использован в экспериментальных приложениях кодирования и обработки звука и речи [6-А, 9-А, 10-А, 12-А, 14-А]. Полученные результаты свидетельствуют о том, что предлагаемый метод может применяться для практической обработки речевых и звуковых сигналов, позволяя с высокой точностью определять мгновенные параметры периодических компонент сигнала. Предложены и экспериментально реализованы методы обработки гармонических параметров с целью создания специальных звуковых эффектов, таких как изменение тембра, тона и временного масштаба синтезируемого сигнала. Экспериментальное применение разработанных методов для коррекции певческого голоса позволяет достигать необходимых целевых параметров, сохраняя натуральность звучания. Применение разработанных методов для конверсии голоса, позволяет получать разборчивый речевой сигнал с параметрами целевого диктора.

6. Разработанные методы оценки гармонических параметров и их обработки практически применены для модификации системы синтеза речи по тексту [7-А]. Описание акустической базы данных параметрами гармонической модели позволяет выполнять конкатенацию речевых сегментов с изменением просодических параметров, а так же согласовывать параметры соседних фреймов для соединения сегментов без слышимых артефактов. Достигнуто более высокое общее качество синтезируемого сигнала по сравнению с аналогичными системами синтеза речи по тексту, использующими конкатенацию во временной области. Разработанный метод синтеза речи по тексту может быть реализован на мобильной вычислительной платформе.

Рекомендации по практическому использованию результатов

Концепция использования мгновенных гармонических параметров для частотно-модулированных сигналов с целью получения более точного параметрического описания звука и речи позволяет усовершенствовать системы мультимедиа для обработки звука (системы синтеза речи по тексту, кодеры и вокодеры, конверторы голоса и голосовые процессоры). Полученные результаты свидетельствуют о перспективности дальнейших исследований в области гармонического анализа, а так же методов обработки гармонических параметров.

Перспектива развития гармонической модели и методов гармонического преобразования заключается в поиске специфических способов анализа и обработки, зависящих от целевого приложения системы. Одним из достоинств предлагаемого подхода является гибкость и адаптивность к требованиям решаемых задач. Преимуществом так же является гибкость представления акустической информации, которая адаптируется необходимым образом, отвечая требованиям той или иной задачи.

Предполагается, что разработанный акустический процессор может быть использован в составе системы синтеза речи по тексту, работающей как на персональном компьютере, так и на мобильном вычислительном устройстве. Дальнейшие исследования будут направлены на применение разработанного метода гармонического анализа в смежных областях, например, акустической диагностики вращающихся механизмов.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в научных журналах

- 1-А. Азаров, И.С. Анализ речи на основе мгновенного линейного предсказания / И.С. Азаров, А.А.Петровский // Информатика. – 2008. - № 2. – С. 71–82.
- 2-А. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала / И.С. Азаров, А.А. Петровский // Речевые технологии. – 2008. – № 1 (1). – С. 67–77.
- 3-А. Азаров, И.С. Непрерывное и дискретное гармонические преобразования для декомпозиции речевого сигнала на периодическую и шумовую компоненты / И.С. Азаров, А.А. Петровский // Доклады БГУИР. – 2008. – № 4 (34). – С. 92–105.

Материалы конференций

- 4-А. Azarov, E. Continuous representation of the discrete signal in the frequency domain / E. Azarov, A. Petrovsky // Neural networks and artificial intelligence: proceedings of the 5-th intern. conf., Belarus, Minsk, May 27–30, 2008. – Minsk, 2008. – P. 102–106.
- 5-А. Azarov, E. Linear prediction of speech based on instantaneous harmonic parameters / E. Azarov, A. Petrovsky // Speech and Computer: proceedings of the 13-th intern. conf., Russia, Saint-Petersburg, June 21–25, 2009. – Saint-Petersburg, 2009. – P. 215–220.
- 6-А. Azarov, E. Text and speaker independent voice conversion / E. Azarov, A. Petrovsky // Pattern recognition and information processing: proceedings of the 10-th intern. conf., Belarus, Minsk, May 19–21, 2009. – Minsk, 2009. – P. 195–198.
- 7-А. Text-to-speech system with acoustic processor based on the instantaneous harmonic analysis / E. Azarov, A. Petrovsky, B. Lobanov, L. Tsurulnik // Speech and Computer: proceedings of the 13-th intern. conf., Russia, Saint-Petersburg, June 21 25, 2009. – Saint-Petersburg, 2009. P. 414–418.
- 8-А. Estimation of the instantaneous harmonic parameters of speech [electronic resource] / E. Azarov, A. Petrovsky, M. Parfieniuk. - EUSIPCO 2008: proc. of the 16th European Signal Processing Conference, Switzerland, Lausanne, August 25-29, 2008. - Lausanne., 2008. – Mode of access: <http://eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569101820.pdf>. - Date of access: 09.09.2008.
- 9-А. Instantaneous harmonic analysis for vocal processing [electronic resource] / E. Azarov, A. Petrovsky. - DAFX-09: proc. of the 12th International Conference on Digital Audio Effects, Italy, Como, September 1-4, 2009. – Como., 2009. –

Mode of access: http://dafx09.como.polimi.it/proceedings/papers/paper_25.pdf. - Date of access: 04.09.2009.

10-A. Petrovsky, A. Combining advanced sinusoidal and waveform matching models for parametric audio/speech coding / A. Petrovsky, E. Azarov, A. Petrovsky // EUSIPCO 2009: proc. of the 17th European Signal Processing Conference. – Glasgow., 2009. – P. 436-440.

Тезисы докладов

11-A. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала / И.С. Азаров, А.А. Петровский // Современные средства связи: тезисы докладов 12-ой международной научно-технической конференции, Беларусь, Нарочь, 24–28 сентября 2007 г. – С. 87-88.

12-A. Азаров, И.С. Конверсия голоса с использованием дискретного гармонического преобразования / И.С. Азаров, А.А. Петровский // Современные средства связи: тезисы докладов 13-ой международной научно-технической конференции, Беларусь, Минск, 7–9 октября 2008 г. – С. 61.

Другие публикации

13-A. Azarov, E. Estimation of the instantaneous harmonic parameters of speech / E. Azarov, A. Petrovsky // Recent advances in signal processing. – Vienna: In-Tech publisher, 2009. – 23 p.

14-A. Petrovsky, A. Harmonic representation and auditory model-based parametric matching and its application in speech/audio analysis / A. Petrovsky, E. Azarov, A.A. Petrovsky. – Munich, 2009. – 13 p. – (Preprint / Signal processing: proceedings of 126-th AES convention; № 7705).

15-A. Азаров, И.С. Следящий гармонический анализатор речевых сигналов / И.С. Азаров, А.А. Петровский // Труды Российского научно-технического общества радиотехники, электроники и связи имени А.С. Попова. – Москва, 2008. – № 10 (1). - С. 272-276.



РЕЗЮМЕ

Азаров Илья Сергеевич

ГАРМОНИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ С ЧАСТОТНО-ВРЕМЕННЫМ МАСШТАБИРОВАНИЕМ ДЛЯ СИСТЕМ ОБРАБОТКИ И СИНТЕЗА ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ

Ключевые слова: гармоническая (синусоидальная) модель, гармонический анализ звука и речи, гибридное кодирование сигнала, изменение частоты основного тона, тембра и временного масштаба, синтез речи по тексту.

Целью работы является разработка специальных методов и алгоритмов для вычисления мгновенных гармонических параметров речевых и звуковых сигналов и использования данного параметрического представления в системах обработки звуковой информации.

Объектом исследования являются системы мультимедиа, использующие синусоидальную или гармоническую модель представления сигнала. Предмет исследования – методы гармонического преобразования, позволяющие получить параметрическое представление сигнала в виде набора мгновенных параметров.

Для точной оценки гармонических параметров синтезирован специальный фильтр анализа с модулированной импульсной характеристикой, позволяющий производить узкополосную фильтрацию частотно-модулированных периодических компонент сигнала. Выход фильтра анализа представлен в аналитическом виде, что позволяет выполнять непосредственную оценку мгновенных гармонических параметров.

На основе синтезированных фильтров предлагается способ гармонического анализа звуковых и речевых сигналов, который может применяться в системах мультимедиа, таких как кодеры, вокодеры, конверторы голоса, голосовые процессоры и системы синтеза речи по тексту.

Разработан специальный способ конверсии гармонических параметров в коэффициенты предсказания. Показано, что применение мгновенного гармонического анализа в сочетании с методом конверсии позволяет выполнять мгновенную оценку формантных траекторий, что может быть использовано в задачах распознавания речи и конверсии голоса.

Для проверки эффективности предлагаемых методов гармонического анализа и обработки параметров гармонической модели рассматриваются вопросы их практического применения в некоторых речевых и звуковых приложениях.

Разработан акустический процессор для системы синтеза речи по тексту, позволяющий выполнять синтез при помощи параметрического представления речевых сегментов.

SUMMARY

Elias Azarov

HARMONIC TRANSFORM WITH FREQUENCY AND TIME WARPING FOR PROCESSING AND SYNTHESIS OF AUDIO AND SPEECH SIGNALS

Key words: harmonic (sinusoidal) model, harmonic analysis of speech and audio, hybrid coding, pitch modification, text to speech synthesis.

The purpose of the dissertation is developing of special methods and algorithms for instantaneous harmonic parameters extraction from speech and audio with subsequent application of the parametric representation for various sound processing systems.

The object of the investigation is a subclass of multimedia systems that use sinusoidal or harmonic representation of the signal. The subject of the investigation covers different methods of harmonic transform that produce parametric representation of the signal in the form of instantaneous parameters.

In order to get accurate estimates a special analysis filter with modulated impulse response has been synthesized. The filter provides narrow-band filtering of frequency-modulated signal's components. The output of the filter is presented in the form of a closed-form expression that allows direct parameters estimation.

A method of audio and speech analysis on the base of synthesized filters has been proposed. The method can be applied to different multimedia systems such as coders, vocoders, voice convertors, voice processors and text to speech synthesizers.

A special conversion technique has been presented that transforms harmonic parameters into prediction coefficients. It has been shown that combining of instantaneous harmonic analysis with the technique can provide instantaneous formant tracking, that can be very useful in speech recognition and voice conversion.

In order to show that the proposed analysis methods are highly effective some experimental applications has been carried out. The methods were applied to speech and audio processing systems.

A special acoustic processor for speech concatenation in harmonic domain has been researched and implemented. The processor is used in text-to-speech synthesis system.

Научное издание

АЗАРОВ ИЛЬЯ СЕРГЕЕВИЧ

**ГАРМОНИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ
С ЧАСТОТНО-ВРЕМЕННЫМ МАСШТАБИРОВАНИЕМ
ДЛЯ СИСТЕМ ОБРАБОТКИ И СИНТЕЗА
ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ**

Специальность 05.13.17 – теоретические основы информатики

**Автореферат диссертации на соискание ученой степени
кандидата технических наук**

Подписано в печать 19.10.2009.	Формат 60x84 ¹ / ₁₆ .	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.	Усл. печ. л. 1,63.
Уч.-изд. л. 1,3.	Тираж 60 экз.	Заказ 667.

Издатель и полиграфическое исполнение: Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
ЛСИ №02330/0494371 от 16.03.2009. ЛП №02330/0494175 от 03.04.2009.
220013, Минск, П. Бровки, 6.