

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

А. А. Петровский, И. С. Азаров

ЦИФРОВАЯ ОБРАБОТКА МЕДИАННЫХ

*Рекомендовано УМО вузов Республики Беларусь
по образованию в области информатики и радиоэлектроники
в качестве учебно-методического пособия для студентов учреждений,
обеспечивающих получение высшего образования по специальности
1-40 02 02 «Электронные вычислительные средства»*

Минск БГУИР 2012

УДК 621.391(076)

ББК 32.811я73

ПЗ0

Р е ц е н з е н т ы:

главный научный сотрудник

объединенного института проблем информатики

Национальной академии наук Беларуси,

профессор, доктор технических наук В. В. Старовойтов;

доцент кафедры радиотехники и цифровых медиатехнологий

Белорусского государственного университета,

кандидат технических наук В. И. Микулович

Петровский, А. А.

ПЗ0 Цифровая обработка медиаданных : учеб.-метод. пособие /
А. А. Петровский, И. С. Азаров. – Минск : БГУИР, 2012. – 56 с. : ил.
ISBN 978-985-488-790-6.

Содержит описание методов оценки частотно-временных параметров нестационарных сигналов, включая гармоническое преобразования на основе преобразования Фурье, метод оценки мгновенных гармонических параметров и метод согласованной подгонки. Приведен пример гибридной декомпозиции аудиосигнала с использованием изложенных подходов.

Издание предназначено для студентов специальности 1-40 02 02 «Электронные вычислительные средства».

УДК 621.391(076)

ББК 32.811я73

ISBN 978-985-488-790-6

© Петровский А. А., Азаров И. С., 2012

© УО «Белорусский государственный университет информатики и радиоэлектроники», 2012

1 Синусоидальное представление сигнала

Синусоидальное (гармоническое) представление сигнала широко применяется в современных системах обработки звука, с его помощью можно эффективно описывать периодические компоненты, которые составляют значительную часть различных звуков и речи. Синусоидальная и гармоническая модели были с успехом использованы в системах кодирования речи, таких, как мультиполосный вокодер [6, 10] и трансформкодер [17], а также в кодировании широкополосных звуковых сигналов [14].

Хотя предположение о том, что произвольная функция может быть представлена в виде суммы синусоидальных составляющих, принадлежит Жану Баттисту Фурье и было высказано еще в 1807 году, модель для представления дискретных звуковых сигналов в виде суммы синусоидальных составляющих с изменяющимися во времени амплитудами, частотами и фазами была представлена сравнительно недавно [20]. В работе [19] представлена модель, которая рассматривает вокализованную речь как сумму гармонически связанных синусоид с амплитудами и фазами, вычисляемыми непосредственно из спектра дискретного преобразования Фурье, а невокализованная речь рассматривается как сумма случайно распределенных синусоид со случайными начальными фазами. Позже данная модель была значительно усовершенствована [11].

Синусоидальное представление предполагает, что моделируемый сигнал является комбинацией синусоид с медленно изменяющимися амплитудами и частотами и, следовательно, не всегда эффективно для описания шумовых и взрывных звуков. Таким образом, одна из особенностей использования синусоидальных представлений применительно к звуковым сигналам заключается в том, что часто нет необходимости точно описывать весь исходный сигнал в виде суммы синусоид. Намного эффективнее с практической точки зрения разделить сигнал на отдельные составляющие, для каждой из которых применить затем свою модель описания. Такая форма представления называется гибридной и использует разные способы параметрического представления сигнала. В со-

временных системах компрессии, например, используют раздельное кодирование синусоидальной, шумовой и транзиентной (переходной) составляющих [14, 28]. Первая из них представляет собой сумму тригонометрических функций (синус и косинус); вторая – нерегулярный сигнал без четкой локализации спектра, который на слух воспринимается как шум; третья – непериодические (часто непродолжительные) фрагменты с резким изменением энергии и формы. Гибридные модели позволяют повысить эффективность кодирования сигналов за счет использования психоакустических критериев оценки и применения специальных способов шумоподавления [3, 34, 35, 36]. Аналогичное гибридное параметрическое представление используется в системах обработки звука для создания специальных эффектов (изменение масштаба времени, изменение основного тона, тембра и др.) [11, 14], а также в системах синтеза речи по тексту [7, 13]. Представление речевого сигнала при помощи гибридной модели представляет особый интерес, фрагменты речи различной природы звукообразования (вокализованные и невокализованные) относятся к отдельным составляющим модели (периодической и остаточной соответственно). При гармоническом анализе речи сигнал разделяют на вокализованные и невокализованные фрагменты, применяя синусоидальное описание только к первым из них. Вокализованная речь часто моделируется при помощи синусоид, гармонически расположенных в частотной области (гармоническая модель), что значительно сокращает число параметров, необходимых для хранения, обработки или передачи.

Так как синусоидальное моделирование в большинстве случаев можно применить непосредственно к входу системы, входной сигнал удобно рассматривать как комбинацию периодической и остаточной компонент [21]. Таким образом, сигнал $s(n)$ можно записать в виде соотношения

$$s(n) = \sum_{k=1}^K \text{MAG}_k(n) \cos \varphi_k(n) + r(n), \quad (1.1)$$

где $MAG_k(n)$ – 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:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0), \quad (1.2)$$

где F_s – частота дискретизации;

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

Наряду с синусоидальной моделью широко применяется (особенно в вокодерных системах) гармоническая модель, которая предполагает, что значения мгновенных частот являются кратными частоте основного тона $f_0(n)$ и могут быть вычислены по следующей формуле:

$$f_0(n) = k f_0(n). \quad (1.3)$$

Гармоническая модель используется для кодирования речевых сигналов с высоким коэффициентом сжатия [1, 8, 18], т. к. обеспечивает чрезвычайно эффективное описание вокализованных фрагментов речи.

Задачей гармонического анализа является определение синусоидальных параметров $MAG_k(n)$, $f_k(n)$ и $\varphi_k(n)$ для заданного момента или интервала времени. Считается, что амплитуда и фаза синусоидальных компонент изменяется медленно, поэтому можно сделать следующие предположения:

– каждая синусоида может быть ограничена в частотной области узкой частотной полосой;

– синусоидальные компоненты разделены в частотной области (их можно выделить на всем протяжении анализируемого фрейма фильтрами с не-

перекрывающимися полосами пропускания), в противном случае они создают переходную (транзиентную) компоненту;

– синусоидальные компоненты достаточно продолжительны, в противном случае они формируют либо переходную компоненту, либо шум.

Таким образом, искомые параметры синусоидальной модели являются гладкими, непрерывными функциями с ограниченным частотным диапазоном. Разделение сигнала на периодическую и остаточную части, как и оценка гармонических параметров, является фундаментальной задачей синусоидального моделирования. Точность оценок, как правило, оказывает существенное влияние на качество работы систем обработки звуковой информации. Неточное разделение вносит в обрабатываемый сигнал слышимые артефакты, которые затем на последующих стадиях обработки не могут быть исключены. В качестве основного инструмента для выполнения гармонического анализа используется дискретное преобразование Фурье (ДПФ), что накладывает определенные ограничения. В этом случае предполагается, что анализируемый сигнал является квазистационарным, т. е. на протяжении некоторого периода времени его параметры остаются неизменными. Прежде всего, используя ДПФ, сложно получить адекватное параметрическое описание неустойчивых тональных звуков. Другая проблема заключается в сложности анализа сигналов с быстро изменяющимся тоном. Например, ДПФ очень ограничено применимо к оценке параметров гармоник высокого порядка вокализованной речи из-за свойственного им быстрого изменения частоты. Тем не менее, используя ДПФ, были получены достаточно неплохие результаты в системах синтеза речи по тексту [7] и кодирования речи [26].

2 Частотно-временные преобразования

Одной из самых важных частей любого гармонического анализатора является частотно-временное преобразование, выбор которого определяется особенностями решаемой задачи.

2.1 Преобразование Фурье

Как было сказано выше, одним из наиболее используемых преобразований является дискретное преобразование Фурье, которое широко применяется в цифровом спектральном анализе для обработки речи и изображений, а также в задачах распознавания образов. Преобразование также используется при проектировании фильтров, т. к. устанавливает связь между импульсной и частотной характеристиками. Существует быстрое преобразование Фурье (БПФ) – простой алгоритм для эффективного вычисления ДПФ, который реализован на многих вычислительных платформах.

Преобразование Фурье (ПФ) непрерывного во времени сигнала $S(f)$ имеет вид

$$S(f) = \int_{-\infty}^{+\infty} s(t)e^{-jft} dt, \quad (2.1)$$

где f – частота;

j – мнимая единица.

Таким образом, сигнал $s(t)$ раскладывается на бесконечное число комплексных синусоид, параметры которых задаются функцией $S(f)$. Обратное преобразование Фурье (ОПФ) переводит сигнал из частотного представления во временную область и имеет следующий вид:

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(f)e^{jft} df. \quad (2.2)$$

В цифровой обработке сигналов используется дискретная версия преобразования, т. к. на практике анализируются последовательности отсчетов конечной длины. Соотношение для получения N -точечного дискретного преобразования Фурье имеет следующий вид:

$$S(f) = \frac{1}{N} \sum_{n=0}^{N-1} s(n)e^{-\frac{j2\pi n f}{N}} = \frac{1}{N} \sum_{n=0}^{N-1} s(n) \left[\cos\left(\frac{2\pi n f}{N}\right) - j \sin\left(\frac{2\pi n f}{N}\right) \right], \quad (2.3)$$

где $s(n)$ – дискретный сигнал;

f – нормализованная частота.

Синусоидальные компоненты преобразования могут быть в полярных или прямоугольных координатах, связь между которыми определяется формулой Эйлера:

$$e^{j\theta} = \cos \theta + j \sin \theta. \quad (2.4)$$

Выходной спектр ДПФ $S(f)$ является результатом вычисления свертки между выборкой, состоящей из входных отсчетов во временной области, и набором из N пар гармонических базисных функций (косинус и синус). Подобная процедура выполняется при вычислении обратного ДПФ для восстановления отсчетов сигнала $s(n)$ из спектра ДПФ $S(f)$. Соответствующее соотношение выглядит следующим образом:

$$s(n) = \sum_{f=0}^{N-1} S(f)e^{\frac{j2\pi n f}{N}} = \sum_{n=0}^{N-1} S(f) \left[\cos\left(\frac{2\pi n f}{N}\right) + j \sin\left(\frac{2\pi n f}{N}\right) \right]. \quad (2.5)$$

Выходной спектр ДПФ может быть представлен либо в полярной системе координат (амплитуда и фаза), либо в алгебраической форме (вещественная и

мнимая части). Обе указанные формы находятся во взаимно однозначном соответствии. Преобразование вещественных и мнимых частей в амплитуду и фазу выполняется по формулам:

$$S(f) = \text{Re}S(f) + j \text{Im}S(f), \quad (2.6)$$

$$\text{MAG}[S(f)] = \sqrt{\text{Re}S(f)^2 + \text{Im}S(f)^2}, \quad (2.7)$$

$$\varphi[S(f)] = -\arctan \frac{\text{Im}S(f)}{\text{Re}S(f)}. \quad (2.8)$$

Для гармонического анализа иногда применяется ДПФ с неравномерным частотным разрешением [4, 30, 37], позволяющее получить более точные оценки.

2.2 Гармоническое преобразование

Преобразование Фурье сложно применять для анализа гармонических частотно-модулированных (ЧМ) сигналов, т. к. выходной спектр ДПФ представляет собой набор стационарных компонент с фиксированными частотами. Для того чтобы локализовать частотно-модулированные гармонические компоненты, применяется преобразование Фурье с временным масштабированием. Исходный сигнал из временной оси t отображается на масштабную ось u при помощи некоторой масштабной функции $p(t)$. Сигнал при этом растягивается или сжимается во времени [2], для того чтобы перед вычислением ДПФ обеспечить стационарность частоты (рисунок 2.1). Для речи данная операция применяется с целью обеспечения стационарности частоты основного тона.

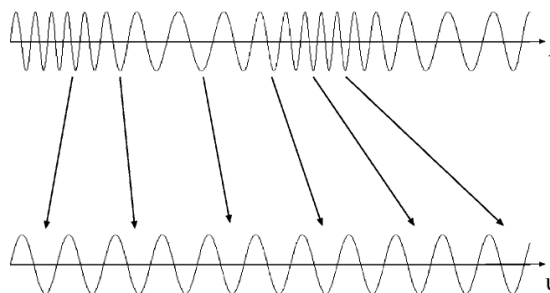


Рисунок 2.1 – Временное масштабирование ЧМ-сигнала

Таким образом, удается получить более четкую локализацию гармонических составляющих в выходном спектре. Представленная идея легла в основу фан-чирп (fan-chirp) [31] и гармонического [32] преобразований, интегрирующих функцию масштабирования времени в преобразование Фурье. В данной трактовке гармоническое преобразование (ГП) непрерывного сигнала $s(t)$, согласно [32], определяется как

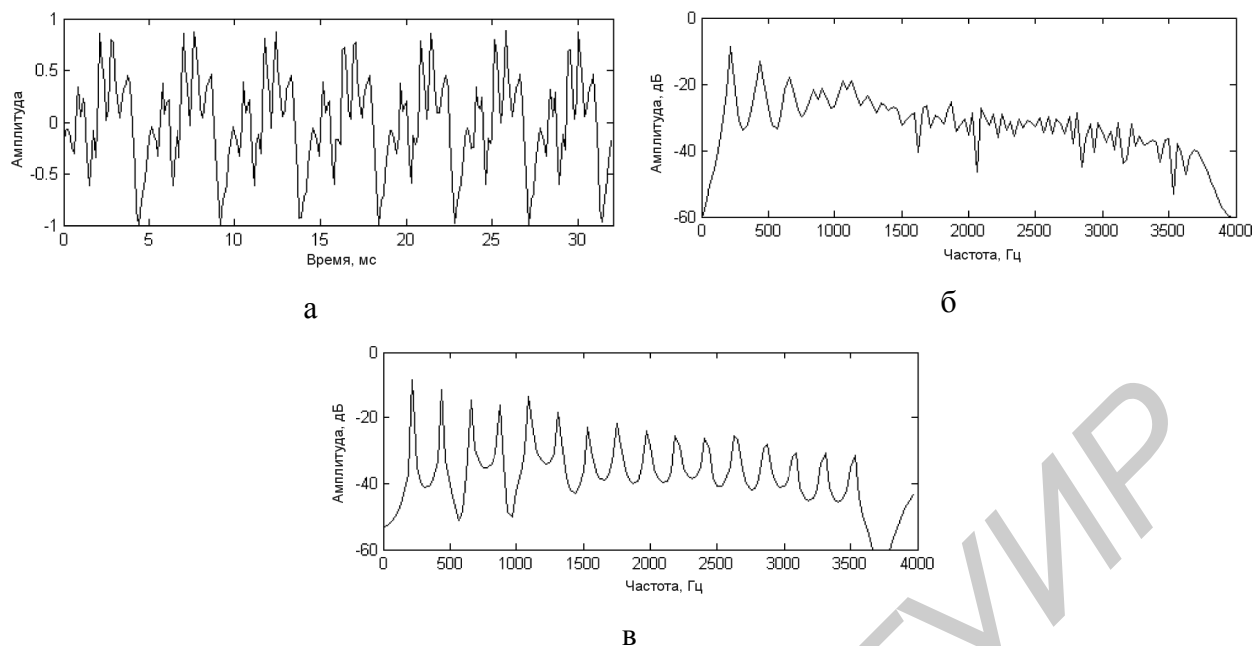
$$S_{\varphi_u(t)}(f) = \int_{-\infty}^{+\infty} s(t) \varphi'_u(t) e^{-jf\varphi_u(t)} dt, \quad (2.9)$$

где $\varphi_u(t)$ – функция основного тона, деленная на ее мгновенную частоту;
 $\varphi'_u(t)$ – производная от $\varphi_u(t)$.

Фазовая функция должна быть дифференцируема и обратима на интервале $-\infty; +\infty$:

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{\varphi_u(t)}(f) e^{jf\varphi_u(t)} df. \quad (2.10)$$

Очевидно, что ГП превращается в ПФ при $\varphi_u(t) = t$. На рисунке 2.2 показан пример преобразования сигнала из временной области в частотную при помощи преобразования Фурье и гармонического преобразования. На рисунке 2.3 схематично показано, как соотносятся ПФ и ГП. Нестационарный гармонический сигнал (рисунок 2.3, а) с использованием ГП выглядит в форме четких пиков в частотной области (рисунок 2.3, в), тогда как при применении ПФ (рисунок 2.3, б) получается пологий спектр, где все гармоники смешаны.



а – исходный вокализованный сегмент речи; б – преобразование Фурье;

в – гармоническое преобразование

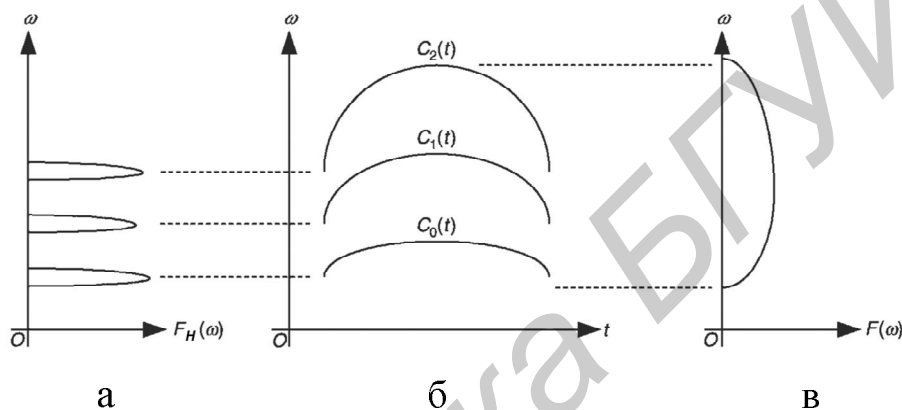
Рисунок 2.2 – Временное масштабирование ЧМ-сигнала

Следует отметить, что фазовая функция $\varphi_u(t)$ должна быть определена перед вычислением ГП. Очевидно, что для заданного сигнала $s(t)$ существует единственная такая функция, действенным методом поиска которой может быть перебор. Функция $\varphi_u(t)$ в таком случае определяется как набор некоторых неизвестных коэффициентов. Для всех возможных комбинаций ГП вычисляется внутри допустимого диапазона значений каждого из них. Теоретически наилучшая концентрация спектра в частотно-временной области определит набор, который и должен использоваться в преобразовании.

Так как перебор является трудоемким способом, существует возможность упрощения процесса поиска с использованием предварительных знаний о сигнале. Очевидно, что поведение $\varphi_u(t)$ в большей степени зависит от локального текущего фрейма и в меньшей – от всего сигнала в целом. Если гармонический анализатор оперирует достаточно короткими фреймами, справедливо предположить, что частота каждого из них изменяется линейно. Это допущение ведет к существенному снижению вычислительной сложности. Знание частотной по-

лосы основной или другой гармоники также может помочь сузить область поиска и упростить процесс вычислений.

Результат гармонического преобразования и функция $\varphi_u(t)$ должны быть сопоставлены с имеющейся информацией о сигнале. Это необходимо для того, чтобы исключить возможность ошибок (удвоение, деление или потеря алгоритмом основного тона, а также его ошибочное определение в невокализованных сегментах или паузах).



а – спектр гармоническое преобразования; б – частотно-временное представление;
 в – преобразование Фурье

Рисунок 2.3 – Схематическое представление преобразования Фурье и гармонического преобразования вокализованной речи

Основные свойства ГП перечислены в таблице 2.1. Последняя колонка приводит условия, которым должна удовлетворять $\varphi_u(t)$. Большинство из них аналогичны свойствам преобразования Фурье. Свойство линейности, в частности, может быть легко получено из соответствующих определений. Для доказательства большинства свойств можно использовать взаимосвязь между ПФ и ГП.

Таблица 2.1 – Основные свойства гармонического преобразования

Свойство	Временная область	Частотная область	$\varphi_u(t)$
Линейность	$ax(t) + by(t)$	$aX_{\varphi_u(t)}(\omega) + bY_{\varphi_u(t)}(\omega)$	
Декомпозиция	$x_e(t) + x_o(t)$, где $x_e(t) = \frac{1}{2}[x(t) + x(-t)]$ $x_o(t) = \frac{1}{2}[x(t) - x(-t)]$	$X_e(\omega) + X_o(\omega)$, где $X_e(\omega) = \text{Re}[X_{\varphi_u(t)}(\omega)]$ $X_o(\omega) = j \text{Im}[X_{\varphi_u(t)}(\omega)]$	Нечетная
Временная инверсия	$x(-t)$	$X_{\varphi_u(t)}(-\omega)$	Нечетная
Объединение	$x^*(t)$	$X_{\varphi_u(t)}^*(\omega)$	
Свертка	$x(t) * y(t)$	$X_{\varphi_u(t)}(\omega)Y_{\varphi_u(t)}(\omega)$	
	$x(t)y(t)$	$X_{\varphi_u(t)}(\omega) * Y_{\varphi_u(t)}(\omega)$	
Симметрия	$\frac{x_{\varphi_u(t)}(t)}{\varphi'_u(t)}$	$X(-\omega)$	
Дельта-функция	$\delta(t - t_0)$ $e^{jw_0\varphi_u(t)}$	$\varphi'_u(t_0)e^{-j\omega\varphi_u(t_0)}$ $\delta(\omega - \omega_0)$	
Энергия	$\int_{-\infty}^{+\infty} x(t)\varphi'_u(t) ^2 dt$	$\int_{-\infty}^{+\infty} X_{\varphi_u(t)}(\omega) ^2$	

Определение и способ вычисления дискретного гармонического преобразования для локализованных речевых сигналов предложены в работе [33]. В речевом сигнале частота основного тона изменяется медленно, поэтому можно допустить, что внутри небольшого временного интервала зависимость частоты от времени линейна.

Мгновенная фаза $\varphi(t)$ синусоиды с линейным изменением частоты определяется известной формулой (для простоты начальная фаза опущена)

$$\varphi(t) = 2\pi \left(f_0 t + \frac{\varepsilon t^2}{2} \right), \quad (2.11)$$

где f_0 – начальная частота;

$\varepsilon = (\Delta f_0/T)$ – изменение частоты основного тона, деленное на время, соответствующее анализируемому фрейму.

Переходя к дискретным сигналам для сегмента длиной в N отсчетов ($T = N/F_s$), где F_s – частота дискретизации, формула (2.11) принимает вид

$$\varphi(n) = 2\pi \left(\frac{f_0 n}{F_s} + \frac{\Delta f_0 n^2}{2NF_s} \right). \quad (2.12)$$

Начальная частота основного тона внутри данного сегмента может быть записана в виде $f_0 = f_c - \frac{af_c}{2}$, где $a = \frac{\Delta f_0}{f_c}$ (f_c – среднее значение частоты основного тона в сегменте длиной N). Подставляя f_0 и Δf_0 в (2.12), получим

$$\varphi(n) = \frac{2\pi f_c}{F_s} \alpha_a(n); \quad \alpha_a(n) = n \left(1 - \frac{a}{2} + \frac{an}{2N} \right). \quad (2.13)$$

Рассмотрим дискретное гармоническое преобразование (ДГП) для сигналов с линейным изменением частоты основного тона. Частоты спектральных линий дискретного преобразования Фурье определяются следующим образом:

$$f_c = \frac{F_s}{N}. \quad (2.14)$$

В ГП средние частоты спектральных линий приравняем к частотам спектральных линий ДПФ. Получим следующее выражение:

$$\varphi(n) = \frac{2\pi}{N} \alpha_a(n). \quad (2.15)$$

Следовательно, можно определить ДГП для сигналов с линейным изменением частоты основного тона [33] в виде

$$S_{\alpha(n)}(f) = \sum_{n=0}^{N-1} s(n)\alpha'(n)e^{-j\frac{2\pi f}{N}\alpha(n)}, \quad (2.16)$$

где $s(n)$ – дискретный сигнал;

$$\alpha'(n) \text{ определяется как } \alpha'(n) = 1 - \frac{a}{2} + \frac{an}{N}.$$

Обратное ДГП будет определяться следующей формулой:

$$s(n) = \frac{1}{N} \sum_{f=0}^{N-1} S_{\alpha(n)}(f)e^{j\frac{2\pi k}{N}\alpha(n)}. \quad (2.17)$$

Основные свойства ДГП приведены в таблице 2.2.

Таблица 2.2 – Основные свойства дискретного гармонического преобразования

Свойство	Временная область	Частотная область
Линейность	$ax(n) + by(n)$	$aX_{\alpha(n)}(k) + bY_{\alpha(n)}(k)$
Свертка	$x(n) * y(n)$	$X_{\alpha(n)}(k)Y_{\alpha(n)}(k)$
	$x(n)y(n)$	$X_{\alpha(n)}(k) * Y_{\alpha(n)}(k)$
Энергия	$\sum_{n=0}^{N-1} x(n)\alpha'(n) ^2$	$\sum_{k=0}^{N-1} X_{\alpha(n)}(k) ^2$

Преобразования (2.16) и (2.17) позволяют анализировать сигнал в гармонической области в том случае, когда известна зависимость частоты основного тона от времени. Однако при анализе речи значение частоты и его приращение неизвестны. Поэтому для определения частоты основного тона используется специальный алгоритм, представленный на рисунке 2.4.

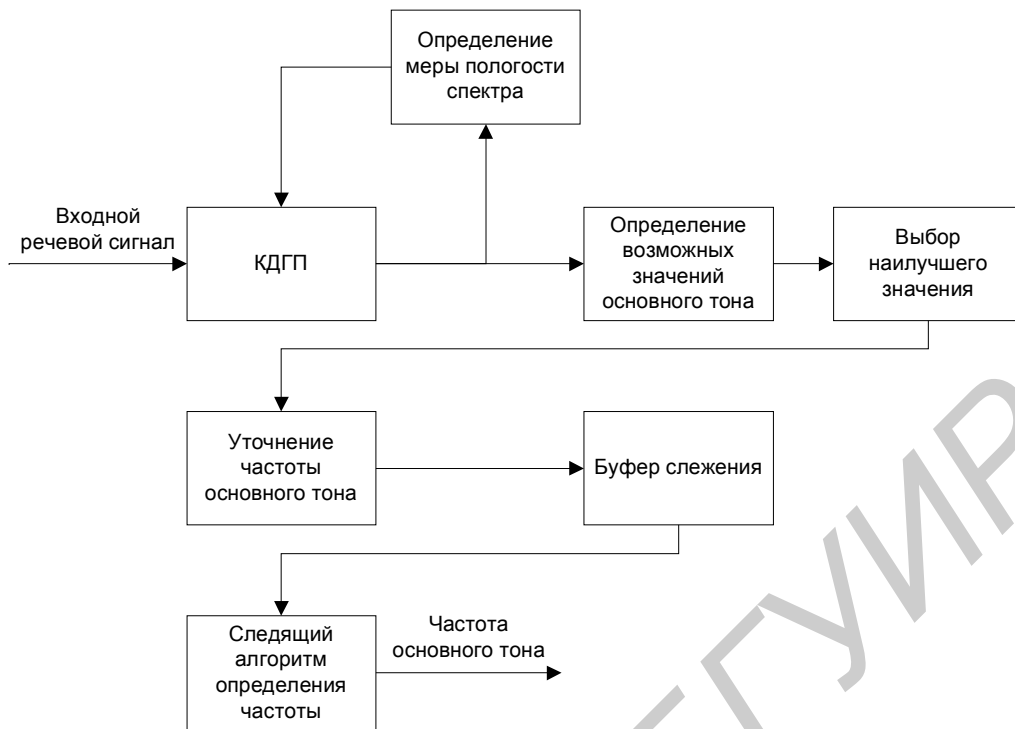


Рисунок 2.4 – Алгоритм определения частоты основного тона

Алгоритм анализирует сигнал фрейм за фреймом, для каждого из которых выполняется кратковременное дискретное гармоническое преобразование (КДГП). Приращение частоты основного тона выполняется путем анализа КДГП для различных фазовых функций, т. е. используя набор фазовых функций с различным параметром a . Его оптимальное значение определяется как минимум меры пологости спектра SFM :

$$\underset{a}{\operatorname{argmin}} SFM(a) = \frac{\sqrt[N]{\prod_{k=0}^{N-1} |\operatorname{КДГП}(a, k)|}}{\frac{1}{N} \sum_{k=0}^{N-1} |\operatorname{КДГП}(a, k)|}, \quad (2.18)$$

где $\operatorname{КДГП}(a, k)$ – гармонический спектр данного сегмента речи, полученный для определенного коэффициента a .

Минимальная пологость спектра означает наивысшую его локализацию в точках, соответствующих гармоникам основного тона, что в рассматриваемом случае обеспечивает максимальное приближение к существующему изменению

частоты. Сама частота основного тона определяется после нахождения изменения.

На первом шаге алгоритм определяет «кандидатов», т. е. возможные значения частоты основного тона f_i , путем определения значений локальных максимумов КДГП на основании алгоритма, предложенного в работе [33]. При этом учитывается, что средняя частота основного тона должна находиться между 50 и 450 Гц. Для каждой возможной частоты основного тона алгоритм ищет ее гармоники. В том случае когда это не удастся, данный «кандидат» отбрасывается. Для того чтобы избежать ошибочного выбора и не принять в качестве частоты основного тона одну из ее гармоник, рассчитывается следующая характеристика:

$$r = \frac{|\sum_{n=1}^{n_{h \max}} |a_n|^2|^2}{n_{h \max}}, \quad (2.19)$$

где a_n – амплитуда n -й гармоники частоты основного тона;

$n_{h \max}$ – максимально возможный номер гармоники для выбранного «кандидата».

В качестве частоты основного тона выбирается «кандидат» с максимальным значением r . Окончательная частота основного тона рассчитывается при помощи следующей формулы:

$$f_r = \frac{\sum_{n=1}^{n_{h \max}} \frac{f_n}{n}}{n_{h \max}}, \quad (2.20)$$

где f_r – частота n -й гармоники выбранного «кандидата».

Описанная процедура определяет среднюю частоту основного тона для одного сегмента. В дальнейшем, для того чтобы обеспечить верный выбор, используется буфер хранения значений частоты основного тона соседних фрей-

мов сигнала. Окончательная оценка частоты выполняется для сегмента, находящегося в центре буфера.

Таким образом, вычисление частоты основного тона происходит с задержкой, равной суммарной длительности нескольких фреймов. В качестве следящего алгоритма используется медианная фильтрация, которая достаточно проста в использовании и обеспечивает устойчивость алгоритма к грубым ошибкам.

При использовании ДГП в гибридных системах обработки сигнала производится разделение сигнала на периодическую и остаточную компоненты. Декомпозиция выполняется во временной области. Периодическая компонента вычисляется из полученных синусоидальных параметров, а шумовая находится как результат вычитания из исходного сигнала.

Для получения всех гармонических параметров (амплитуда, частота и начальная фаза) непосредственно из спектрального представления используется специальный случай ГП. Вариант ДГП, дающий спектр, масштабированный в соответствии с частотой основного тона, определяется как

$$S_{\alpha(n)}(k) = \sum_{n=0}^{N-1} s(n)\alpha'(n)e^{-j\frac{2\pi k f_r}{F_s}\alpha(n)}, \quad (2.21)$$

где f_r – уточненная частота основного тона;

$k = 1 \dots K$;

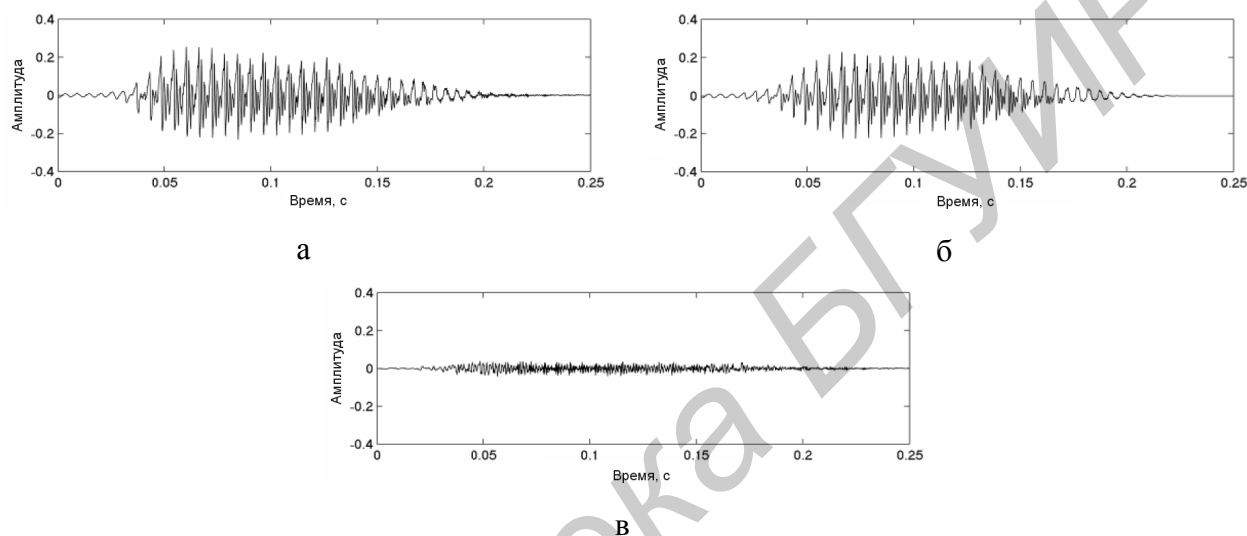
K – число гармоник основного тона.

Амплитуды и фазы могут быть вычислены непосредственно из коэффициентов $S_{\alpha(n)}(k)$:

$$MAG_k[S(k)] = \sqrt{\text{Re}S(k)^2 + \text{Im}S(k)^2}; \quad (2.22)$$

$$\varphi_k [S(k)] = -\arctan \frac{\text{Im}S(k)}{\text{Re}S(k)}. \quad (2.23)$$

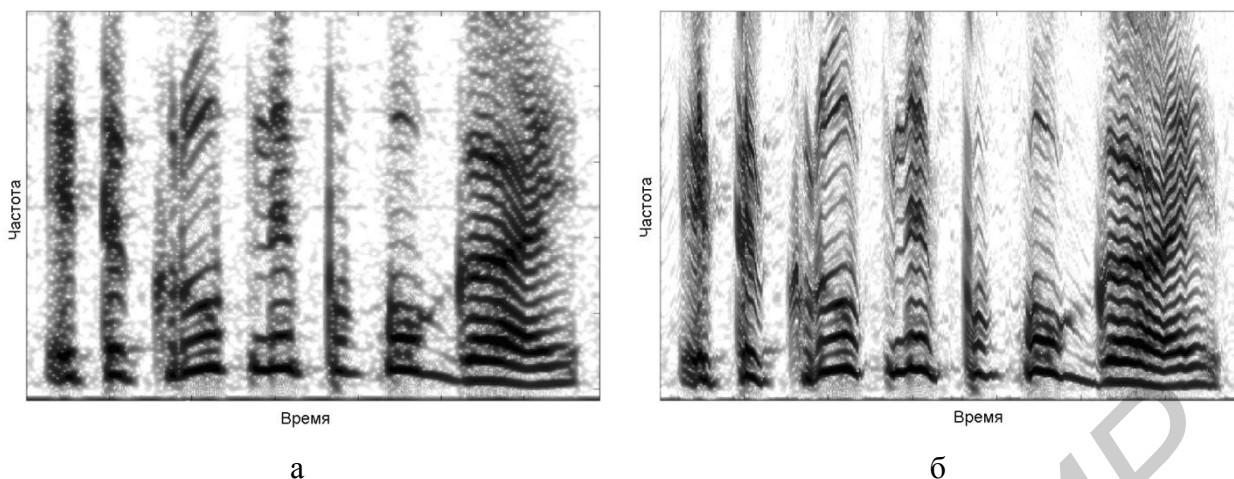
Периодическая компонента сигнала синтезируется при помощи формулы (1.1), а затем для получения остатка вычитается из исходного сигнала. Пример такой сепарации показан на рисунке 2.5.



а – исходный речевой сигнал; б – синтезированная периодическая часть;
в – остаток

Рисунок 2.5 – Пример сепарации речевого сигнала при помощи дискретного гармонического преобразования

Как было сказано выше, использование гармонического преобразования позволяет с большей точностью выделять параметры гармонических ЧМ-сигналов по сравнению с преобразованием Фурье, что можно наглядно продемонстрировать при помощи сравнительных спектрограмм, представленных на рисунке 2.6.



а – ДПФ; б – ДГП

Рис. 2.6 – Сравнительные спектрограммы вокализованного отрывка речи, построенные при помощи различных преобразований

Спектрограмма, построенная при помощи ДГП, позволяет проследить частотную траекторию гармоник более высоких порядков по сравнению со спектрограммой, построенной при помощи ДПФ. Гармоническое преобразование трансформирует исходный сигнал из временной области в область гармоник, т. е. в специальный гармонический спектр, позволяющий с большой точностью выделить параметры периодической части исходного сигнала.

Основные преимущества описанного ДГП перед ДПФ заключаются в следующих возможностях:

- определение изменений частоты основного тона и адаптация преобразования к ним;
- получение спектра сигнала, в котором каждая из гармоник (в том числе и гармоники высоких порядков) имеет четкую локализацию, даже если частота основного тона не является стационарной.

К недостаткам дискретного гармонического преобразования можно отнести следующие ограничения:

- допускается только линейное изменение частоты основного тона и его гармоник внутри фрейма анализа, поэтому для достижения высокой точности

требуется использование коротких фреймов, что ухудшает частотное разрешение преобразования;

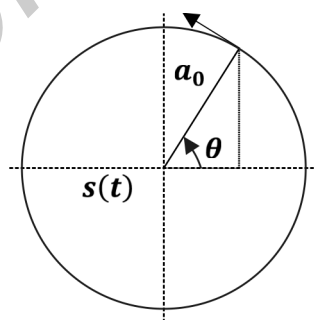
- частоты гармоник принимаются строго кратными частоте основного тона, что может приводить к ошибкам в оценке гармоник высоких порядков;

- как и в преобразовании Фурье, амплитуда считается неизменной на всем фрейме анализа, что отрицательно сказывается на точности оценки неустойчивых гармонических компонент;

- преобразование неприменимо для негармонических синусоидальных сигналов (без выраженной частоты основного тона) и для сложных сигналов, состоящих из нескольких гармонических составляющих (например несколько дикторов, говорящих одновременно).

2.3 Мгновенные гармонические параметры

Понятие мгновенной частоты связано с механикой и может трактоваться через движение по круговой траектории. Если тело движется с постоянной скоростью по кругу, находящемуся в центре координат, то проекции его движения на оси будут простыми гармоническими колебаниями (рисунок 2.7).



$$s(t) = a_0 \cos \theta(t); \quad \theta(t) = \omega_0 t$$

Рисунок 2.7 – Простое гармоническое колебание $s(t)$

В момент времени t проекция $s(t)$ характеризуется смещением, постоянной скоростью и ускорением, которые определяются формулами (2.24), (2.25) и (2.26) соответственно:

$$s(t) = a_0 \cos \theta = a_0 \cos \omega t ; \quad (2.24)$$

$$s'(t) = -a_0 \omega \sin \omega t ; \quad (2.25)$$

$$s''(t) = a_0 \omega^2 \cos \omega t = -\omega^2 s(t). \quad (2.26)$$

Частоту $f = \frac{\omega}{2\pi}$ можно соотнести со смещением, решив дифференциальное уравнение (2.29). Решением является $z(t)$, определяемое как

$$z(t) = \alpha e^{j2\pi f t}, \quad (2.27)$$

где $\omega = 2\pi f$ – постоянная угловая скорость;

α – произвольная константа.

Понятие мгновенной частоты было введено в 1949 году [5] на основе выражения простого гармонического колебания:

$$s(t) = a \cos (2\pi f t + \theta), \quad (2.28)$$

где a – амплитуда;

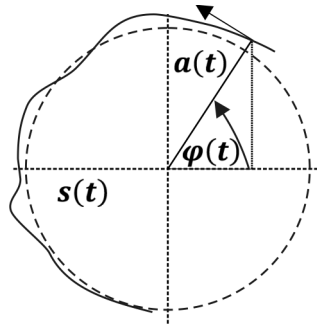
f – частота колебаний;

θ – фазовая константа.

Аргумент функции косинуса $(2\pi f t + \theta)$ является фазой $\varphi(t)$. Амплитудные модуляции (АМ) были определены изменениями a как функции от t :

$$a(t) = a_0 [1 + \mu g(t)], \quad (2.29)$$

где $g(t)$ – модулирующий сигнал (рисунок 2.8).



сигнал определен как $s(t) = a(t)\cos\varphi(t)$;

$a(t)$ – амплитудная огибающая; $\varphi(t)$ – фаза

Рисунок 2.8 – Амплитудная огибающая и мгновенная фаза
АМ-ЧМ-сигнала

Аналогично фазовые модуляции были определены как

$$\theta(t) = \theta_0 [1 + \mu_\theta(t)]. \quad (2.30)$$

Таким образом, фаза, которая является аргументом косинуса в (2.28), принимает вид $\varphi(t) = 2\pi f t + \theta(t)$. Для того чтобы получить выражение для сигнала с частотной модуляцией, которая определялась соответствующим образом как функция мгновенной частоты от времени:

$$f_i(t) = f_0 [1 + \mu_f(t)], \quad (2.31)$$

выражение (2.28) было представлено в следующем виде:

$$s(t) = a \cos \left[\int_0^t 2\pi f_i(t) dt + \theta \right], \quad (2.32)$$

где весь аргумент косинуса представляет собой фазу $\varphi(t)$. Запись сигнала $s(t)$ в таком виде позволяет дать мгновенной частоте следующее определение:

$$f_i(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt}. \quad (2.33)$$

Для определения мгновенных гармонических параметров часто используют понятие аналитического сигнала. В работе [9] предложен способ формирования уникального комплексного сигнала $z(t)$ для заданного действительного сигнала $s(t)$. Он заключается в вычислении ПФ действительного сигнала с последующим приравниванием нулю амплитуд, принадлежащих отрицательным частотам, и удваиванием амплитуд положительных частот. Операцию можно выполнить и без перехода в частотную область:

$$z(t) = s(t) + jH[s(t)] = a(t)e^{j\varphi(t)}, \quad (2.34)$$

где H – преобразование Гильберта, определенное как

$$H[s(t)] = p. v. \int_{-\infty}^{+\infty} \frac{s(t - \tau)}{\pi\tau} d\tau. \quad (2.35)$$

Здесь $p. v.$ обозначает главное значение интеграла по Коши.

Функции $s(t)$ и $H[s(t)]$ теоретически отличаются по фазе на $\frac{\pi}{2}$. Сигнал $z(t)$ называют аналитическим сигналом или комплексным сигналом Габора. Поэтому мгновенную амплитуду $a(t)$ и мгновенную фазу $\varphi(t)$ можно вычислить следующим образом:

$$a(t) = \sqrt{s^2(t) + H^2[s(t)]}; \quad (2.36)$$

$$\varphi(t) = \arctan \left(\frac{H[s(t)]}{s(t)} \right). \quad (2.37)$$

Используя средние меры Габора, в работе [29] показано, что средняя частота спектра сигнала равна его средней мгновенной частоте:

$$\langle f \rangle = \langle f_i \rangle, \quad (2.38)$$

$$\text{где } \langle f \rangle = \frac{\int_{-\infty}^{+\infty} f |Z(f)|^2 df}{\int_{-\infty}^{+\infty} |Z(f)|^2 df}; \quad (2.39)$$

$$\langle f_i \rangle = \frac{\int_{-\infty}^{+\infty} f_i(t) |z(t)|^2 dt}{\int_{-\infty}^{+\infty} |z(t)|^2 dt}; \quad (2.40)$$

$Z(f)$ – преобразование Фурье-сигнала $z(t)$.

Аналитический сигнал для действительного дискретного сигнала имеет ту же интерпретацию и генерируется обычно при помощи ДПФ.

Кроме методов оценки, основанных на использовании аналитического сигнала, существует альтернативный подход с использованием оператора энергии [16], который для дискретного сигнала $s(n)$ вычисляется как

$$\Psi[s(n)] = s^2(n) - s(n-1)s(n+1), \quad (2.41)$$

где дифференцирование выполняется численно, как симметрическая разность.

В работе [16] предложен дискретный алгоритм сепарации энергии (DESA), который оценивает мгновенную амплитуду $MAG(n)$ и мгновенную частоту $f(n)$ сигнала $s(n)$ при помощи следующих выражений:

$$\text{MAG}(n) = \frac{2\Psi[s(n)]}{\sqrt{\Psi[s(n+1) - s(n-1)]}}; \quad (2.42)$$

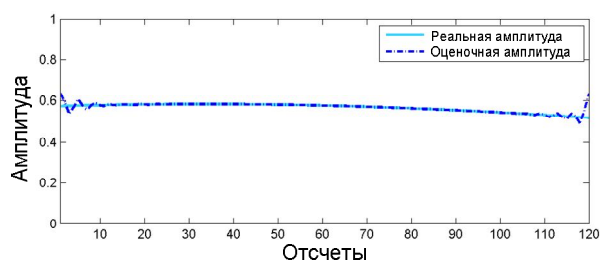
$$f(n) = \arcsin \sqrt{\frac{\Psi[s(n+1) - s(n-1)]}{4\Psi[s(n)]}}. \quad (2.43)$$

Алгоритм оценки требует всего несколько отсчетов анализируемого сигнала для вычисления гармонических параметров, что делает его весьма эффективным в использовании.

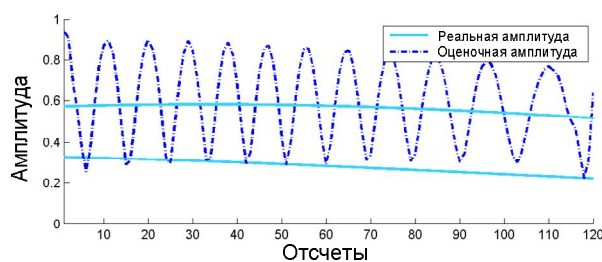
Вышеизложенные способы оценки мгновенных гармонических параметров применимы только для однокомпонентных периодических сигналов, т. к. однозначная функция мгновенной частоты теряет смысл тогда, когда сигнал состоит из нескольких компонент (рисунок 2.9). Поэтому для использования описанных способов в системах обработки цифрового звука предварительно применяют узкополосную фильтрацию [2, 12, 16], позволяющую разделить исходный сигнал на компоненты с ограниченным спектром.

Выбор фильтров является сложной задачей, т. к. во многих случаях он зависит от приложения. Однако очевидно, что для применения мгновенного гармонического анализа речи и звука должна быть обеспечена фильтрация в произвольной полосе частот, что затрудняет использование готовых банков фильтров.

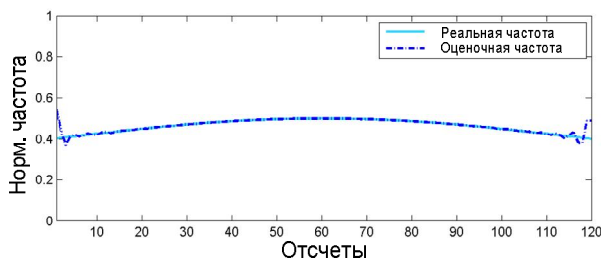
Применение методов мгновенного гармонического анализа позволяет достичь максимально высокого разрешения во временной и частотной областях одновременно, однако необходимость предварительного применения узкополосной фильтрации накладывает некоторые ограничения, которые перечислены ниже:



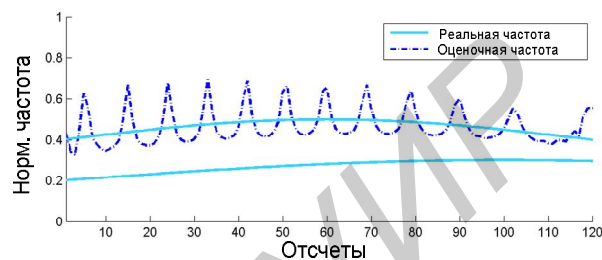
а



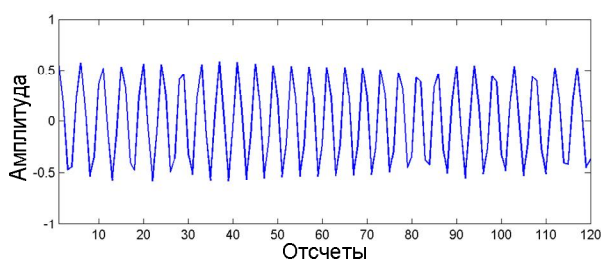
г



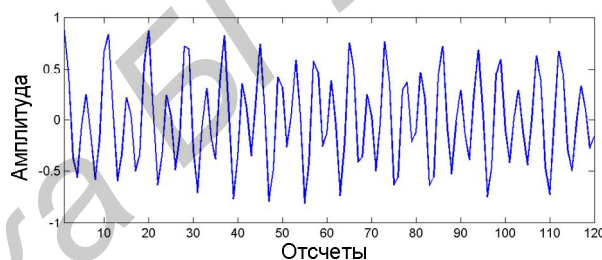
б



д



в



е

а – мгновенная амплитуда однокомпонентного сигнала;

б – мгновенная частота однокомпонентного сигнала; в – однокомпонентный сигнал;

г – мгновенные амплитуды двухкомпонентного сигнала;

д – мгновенные частоты двухкомпонентного сигнала; е – двухкомпонентный сигнал

Рисунок 2.9 – Анализ однокомпонентного и двухкомпонентного сигналов при помощи комплексного сигнала Габора

– необходимо знать примерную частоту анализируемого компонента, для того чтобы использовать нужную полосу пропускания, что требует предварительного анализа сигнала;

– фильтр должен изменять полосу пропускания от фрейма к фрейму, что в зависимости от реализации синтеза фильтра может оказаться сложным для вычисления;

– с одной стороны, фильтр должен иметь как можно более высокий порядок и как можно более узкую полосу пропускания (для точного выделения анализируемого компонента), с другой – полоса пропускания должна быть максимально большой (для пропуска компоненты с частотной модуляцией).

Задача оценки гармонических параметров сводится к определению численных значений переменных выражения (1.1), а именно: мгновенной амплитуды $MAG_k(n)$, мгновенной частоты $f_k(n)$ и фазы $\varphi_k(n)$ для заданного дискретного момента времени из анализируемого фрейма сигнала $s(n)$, $0 \leq n < N$. Оценка параметров должна проводиться исходя из предположения, что периодические компоненты анализируемого фрейма могут быть частотно-модулированы и их гармонические параметры меняются с каждым отсчетом.

Оценка мгновенных параметров возможна путем генерации аналитических сигналов с предварительной узкополосной фильтрацией. Однако для решения поставленной задачи требуется синтез специального фильтра анализа, имеющего следующие свойства:

- фильтрация в заданной полосе частот;
- импульсная характеристика фильтра должна быть описана в виде аналитического выражения (как непрерывная функция от двух параметров – граничных частот полосы пропускания фильтра);
- оценка мгновенных параметров непосредственно из выходного сигнала (выходной сигнал может быть интерпретирован в виде аналитического сигнала);
- модулирование импульсной характеристики в соответствии с возможными частотными модуляциями анализируемого компонента (неявное временное масштабирование);
- непрерывные и гладкие траектории выходных параметров $MAG_k(n)$ и $f_k(n)$.

Таким образом, узкополосная фильтрация, с одной стороны, обеспечит разделение анализируемого сигнала на периодические компоненты и получение

мгновенных гармонических параметров, с другой – позволит правильно обрабатывать компоненты с частотной модуляцией. Данный подход объединяет в себе элементы преобразований со встроенным временным масштабированием (фан-чирп-преобразование и ДГП) и методов оценки на основе аналитических сигналов.

Кроме синтеза фильтра анализа с указанными выше свойствами необходимо также использовать алгоритм локализации периодических компонент в частотной области и алгоритм оценки основного тона для гармонических сигналов. Для использования метода анализа в системах мультимедиа также необходимо применять средства постобработки получаемых мгновенных параметров, обеспечивающих корректное разделение сигнала на периодическую и остаточную части.

2.4 Синтез стационарного фильтра анализа

Фильтр анализа может быть синтезирован, используя N -точечное преобразование Фурье, которое можно рассматривать как фильтр с конечной импульсной характеристикой (КИХ), формирующий на выходе стационарную синусоиду $\bar{s}(n)$ заданной нормализованной частоты f :

$$\bar{s}(n) = \text{MAG}(S(f)) \cos\left(\frac{2\pi n f}{N} + \varphi(S(f))\right), \quad (2.44)$$

где постоянная амплитуда $\text{MAG}(S(f))$ и начальная фаза $\varphi(S(f))$ может быть вычислена следующим образом:

$$\text{MAG}(S(f)) = \sqrt{\text{Re}S(f)^2 + \text{Im}S(f)^2}; \quad (2.45)$$

$$\varphi(S(f)) = -\arctan \frac{\text{Im}S(f)}{\text{Re}S(f)}. \quad (2.46)$$

Импульсная характеристика этого фильтра $h(n)$ может быть записана в виде непрерывной аналитической функции для частоты f в герцах:

$$h(n) = \cos\left(\frac{2\pi}{F_s}nf\right). \quad (2.47)$$

Используя (2.47), можно получить выражение для импульсной характеристики фильтра, формирующего на выходе синусоидальный компонент с ограниченным частотным диапазоном:

$$h(n) = \frac{\int_{F_1}^{F_2} \cos\left(\frac{2\pi}{F_s}nf\right) df}{F_2 - F_1}, \quad (2.48)$$

где F_1 и F_2 – границы полосы пропускания ($F_1 < F_2$).

Интегрирование выражения (2.48) с раскрытием неопределенности при $n = 0$ ведет к импульсной характеристике фильтра в следующем виде:

$$h(n) = \begin{cases} 1, & n = 0 \\ \frac{F_s}{n\pi} \cos\left(\frac{2\pi n}{F_s}F_c\right) \sin\left(\frac{2\pi n}{F_s}F_\Delta\right) & n \neq 0, \end{cases} \quad (2.49)$$

где $F_c = (F_1 + F_2)/2$;

$F_\Delta = (F_2 - F_1)/2$.

Параметры F_c и F_Δ соответствуют центральной частоте полосы пропускания и половине ее ширины соответственно. Сигнал на выходе фильтра $s_{F_c, F_\Delta}(n)$ может быть вычислен как свертка $s(n)$ с $h(n)$ и записан в виде следующей суммы:

$$s_{F_c, F_\Delta}(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \cos\left(\frac{2\pi(n-i)}{F_s} F_c\right) \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right). \quad (2.50)$$

Запишем полученное выражение следующим образом:

$$s_{F_c, F_\Delta}(n) = A(n) \cos(0n) + B(n) \sin(0n), \quad (2.51)$$

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

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

Таким образом, сигнал на выходе фильтра (2.50) представляет собой АМ- и ЧМ-функцию косинуса

$$s_{F_c, F_\Delta}(n) = \text{MAG}(n) \cos(\varphi(n)) \quad (2.52)$$

с мгновенной амплитудой $\text{MAG}(n)$, фазой $\varphi(n)$ и частотой $f(n)$, которые могут быть определены по следующим формулам:

$$\text{MAG}(n) = \sqrt{A^2(n) + B^2(n)}; \quad (2.53)$$

$$\varphi(n) = \arctan\left(\frac{-B(n)}{A(n)}\right); \quad (2.54)$$

$$f(n) = \frac{\varphi(n+1) - \varphi(n)}{2\pi} F_s. \quad (2.55)$$

Так как фаза компонент $A(n)$ и $B(n)$ отличается на $\frac{\pi}{2}$, выход фильтра анализа $s_{F_c, F_\Delta}(n)$ может быть записан в виде аналитического сигнала $s_{F_c, F_\Delta}^a(n)$:

$$s_{F_c, F_\Delta}^a(n) = A(n) + jB(n). \quad (2.56)$$

Можно показать, что при разбиении всего частотного диапазона сигнала произвольными частотами F_1, F_2, \dots, F_L такими, что $F_1 = 0$; $F_L = F_s/2$; $F_1 \leq F_2 \leq \dots \leq F_L$ и $L - 1$, исходный сигнал $s(n)$ равняется сумме выходов фильтров анализа:

$$s(n) = \sum_{k=1}^{L-1} s_{F_k, F_{k+1}}(n). \quad (2.57)$$

Сложение соответствующих аналитических сигналов $s_{F_k, F_k}^a(n)$ дает в результате комплексный сигнал Габора, соответствующий исходному сигналу $s(n)$, который вычисляется при помощи соотношения (2.34):

$$z(n) = \sum_{k=1}^{L-1} s_{F_k, F_k}^a(n) = s(n) + jH[s(n)]. \quad (2.58)$$

Мгновенные гармонические параметры выходного сигнала могут быть рассчитаны в любой момент времени, принадлежащий анализируемому фрейму сигнала, причем этот момент не ограничивается дискретными отсчетами сигнала, т. к. выход фильтра записан в виде непрерывных функций. Очевидно, что полоса пропускания фильтра, задаваемая параметрами F_1 и F_2 (или F_c и F_Δ), должна содержать анализируемую компоненту и в случае наличия частотной модуляции полоса может оказаться достаточно большой (рисунок 2.10).

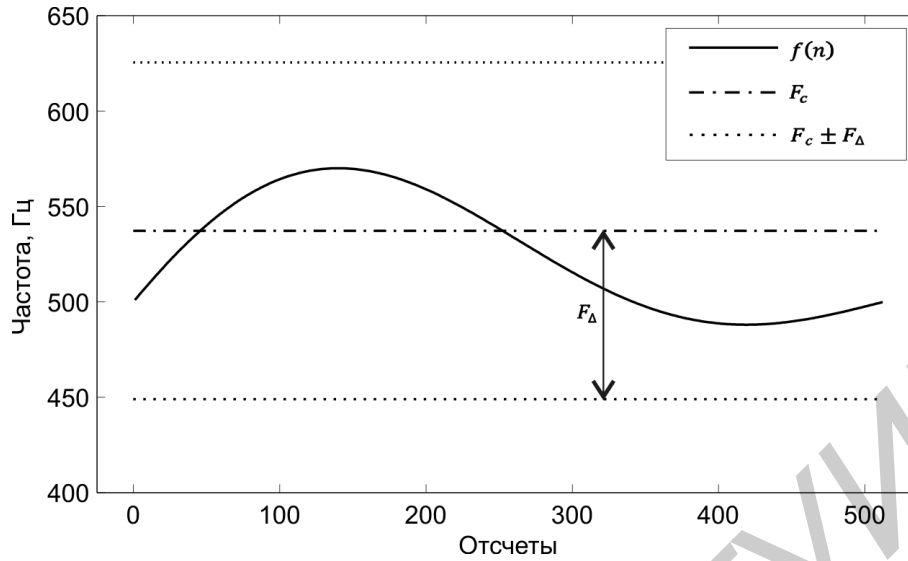


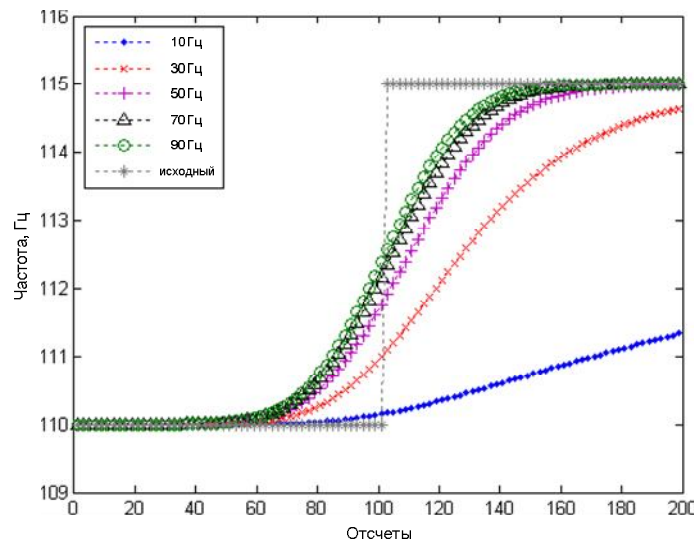
Рисунок 2.10 – Кратковременный гармонический анализ с использованием стационарного фильтра

На рисунке 2.11 показано, как расширение полосы пропускания фильтра влияет на точность оценки мгновенной частоты. Анализируемая гармоническая компонента синтезирована с разрывом контура частоты, что дает возможность определить скорость реакции фильтров анализа.

Как было отмечено ранее, расширение полосы пропускания нежелательно, т. к. это приводит к попаданию дополнительных составляющих в выходной сигнал и, как следствие, к ухудшению точности анализа. На рисунке 2.12 показано, как ширина полосы пропускания фильтра влияет на точность оценки мгновенной частоты, если анализируемый сигнал содержит белый шум. Здесь используется отношение гармоники/шум (HNR) как мера содержания шума в сигнале:

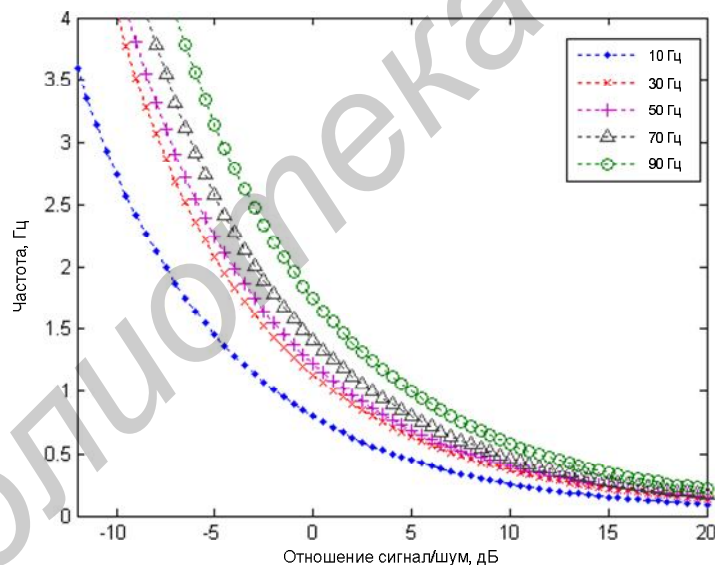
$$HNR = 10 \lg \frac{E_h}{E_r}, \quad (2.59)$$

где E_h и E_r являются энергиями гармонической и шумовой составляющих соответственно.



Ширина полосы пропускания от 10 до 90 Гц

Рисунок 2.11 – Оценка параметров частотно-модулированной гармонической компоненты при помощи фильтров анализа с разной полосой пропускания



Ширина полосы пропускания от 10 до 90 Гц

Рисунок 2.12 – Средняя ошибка оценки мгновенной частоты в зависимости от уровня белого шума в сигнале для фильтров анализа с разной полосой пропускания

В любом случае ширина полосы пропускания фильтра должна быть достаточно большой для того, чтобы выполнять адекватную оценку гармонических амплитуд. В том случае когда полоса слишком узкая, оценочная амплитуда бу-

дет меньше реальной. Минимальная ширина полосы пропускания определяется таким образом, чтобы не занижать оцениваемый амплитудный параметр более чем на указанный порог. Она зависит от длины окна анализа (продолжительность анализируемого фрейма) и используемой оконной функции (рисунок 2.12).

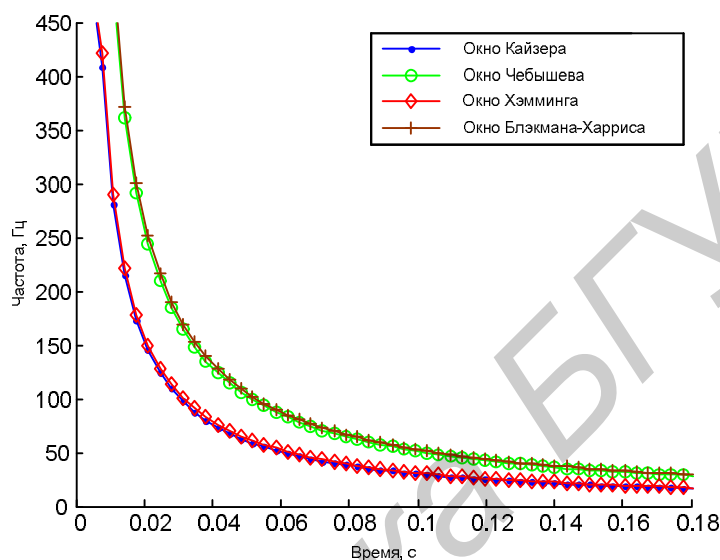


Рисунок 2.13 – Зависимость минимальной ширины полосы пропускания от продолжительности анализируемого фрейма при допустимом понижении амплитуды не более чем на -20дБ

Из рисунка 2.12 следует, что необходимая ширина полосы уменьшается с увеличением продолжительности анализируемого фрейма. Следовательно, для наиболее точной оценки следует использовать продолжительные фреймы, что не всегда допустимо из-за частотной модуляции гармонических компонент.

Таким образом, в отдельных приложениях (вокализованная речь) для точного анализа частотно-модулированных компонент нужно использовать специальный фильтр, адаптирующийся к изменяющемуся частотному контуру.

2.5 Синтез частотно-модулированного фильтра анализа

Для точной оценки синусоидальных параметров компонент с сильной частотной модуляцией должен быть использован частотно-модулированный

фильтр [24, 25], импульсная характеристика которого модулируется в соответствии с частотным контуром анализируемой компоненты. Данный подход хорош для анализа вокализованной речи, т. к. приблизительные частотные траектории гармоник могут быть получены из контура частоты основного тона. Непрерывность импульсной характеристики синтезированного стационарного фильтра позволяет достаточно просто адаптировать его к частотным модуляциям. Рассматривая центральную частоту полосы пропускания как функцию от времени $F_c(n)$, выражение импульсной характеристики фильтра анализа (2.49) может быть представлено в следующем виде:

$$h(n) = \begin{cases} 1, & n = 0 \\ \frac{F_s \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right) \sin\left(\frac{2\pi n}{F_s} F_\Delta\right)}{2F_\Delta} & n \neq 0, \end{cases} \quad (2.60)$$

где $\varphi_c(n, i) = \begin{cases} \sum_{j=n}^i F_c(j), & n < i, \\ -\sum_{j=i}^n F_c(j), & n > i, \\ 0, & n = i. \end{cases}$

Тогда выходной сигнал фильтра будет иметь следующий вид:

$$s_{F_c, F_\Delta}(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right) \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right), \quad (2.61)$$

что можно записать как

$$s_{F_c, F_\Delta}(n) = A(n) \cos(0n) + B(n) \sin(0n), \quad (2.62)$$

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

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

Требуемые мгновенные параметры (амплитуда, частота, фаза) так же, как и в случае стационарного фильтра, могут быть вычислены при помощи выражений (2.53) – (2.55).

Частотно-модулированный фильтр имеет масштабированную в частотной области полосу пропускания, задаваемую частотным контуром $F_c(n)$, что обеспечивает адекватный анализ периодических компонент с быстрым изменением частоты. Этот подход является альтернативой временному масштабированию, которое применяется в анализе речи [2, 31].

Таким образом, анализируемая частотно-модулированная компонента может быть выделена при помощи узкой масштабированной полосы пропускания (рисунок 2.14). Сужение полосы обеспечивает более точную локализацию и соответственно делает оценки параметров более точными. Это также позволяет применять мгновенный гармонический анализ к гармоникам высокого порядка речевого сигнала (рисунок 2.15). Чем выше номер гармоники, тем больше изменение ее частоты и импульсная характеристика ЧМ-фильтра изменяется соответствующим образом.

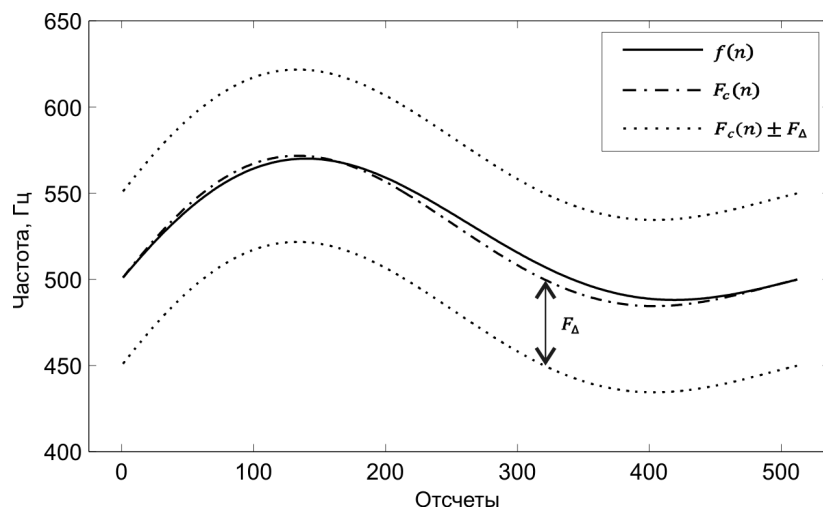
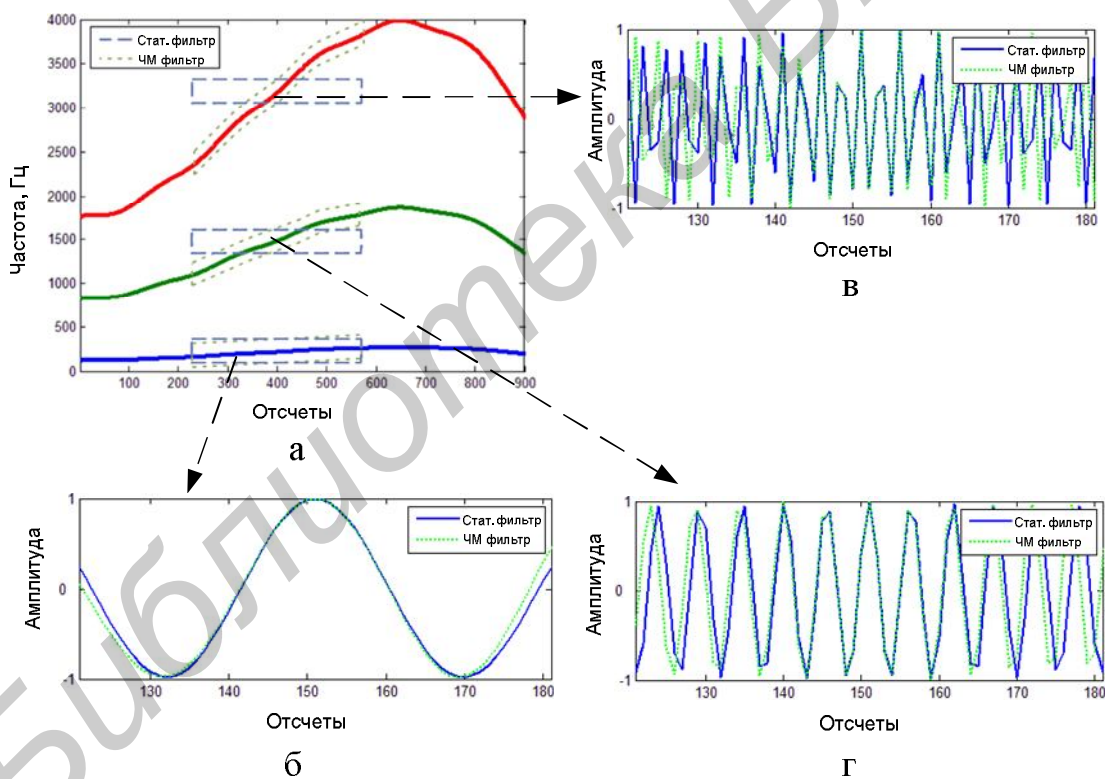


Рисунок 2.14 – Кратковременный гармонический анализ с использованием частотно-модулированного фильтра



а – полосы пропускания фильтров и частотные контура гармоник;
 б – г – импульсные характеристики фильтров для оценки основного тона
 7-й и 15-й гармоник соответственно

Рисунок 2.15 – Применение стационарного и ЧМ-фильтра для анализа вокализованной речи

Для кодирования и обработки звуковых сигналов, как это было показано выше, используются гибридные модели, разделяющие сигнал на периодическую, транзиентную и шумовую компоненты. Первичное разделение синусоидальной (периодической) части и остатка может выполняться путем гармонического анализа на основе фильтров. На рисунке 2.16 показана структура такой системы анализа, использующая гибридную модель параметризации [22].

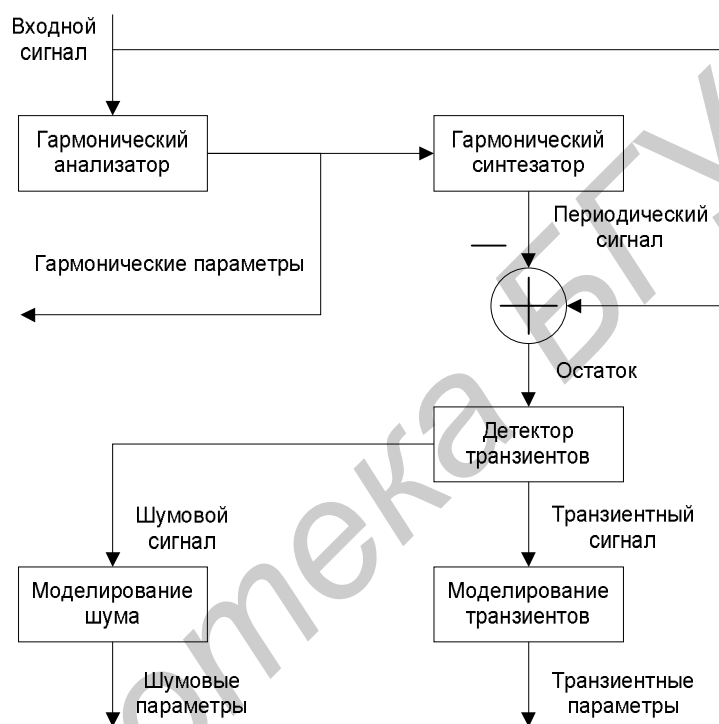


Рисунок 2.16 – Система гибридного анализа для кодирования звука и речи

Вначале синусоидальный анализатор выполняет оценку параметров с постобработкой для выделения продолжительных периодических компонент. Затем выделенная периодическая часть анализируемого сигнала синтезируется синусоидальным синтезатором и вычитается из исходного сигнала. Остаток попадает на вход детектора транзиентов, который выполняет его дополнительное разделение.

Для локализации периодических компонент в частотной области в синусоидальном анализаторе используется итеративный перерасчет оцениваемых

частот. На каждом шаге полоса пропускания каждого фильтра перемещается в соответствии с полученными оценками, для того чтобы поместить пик энергии оцениваемого компонента в центре полосы (рисунок 2.17).

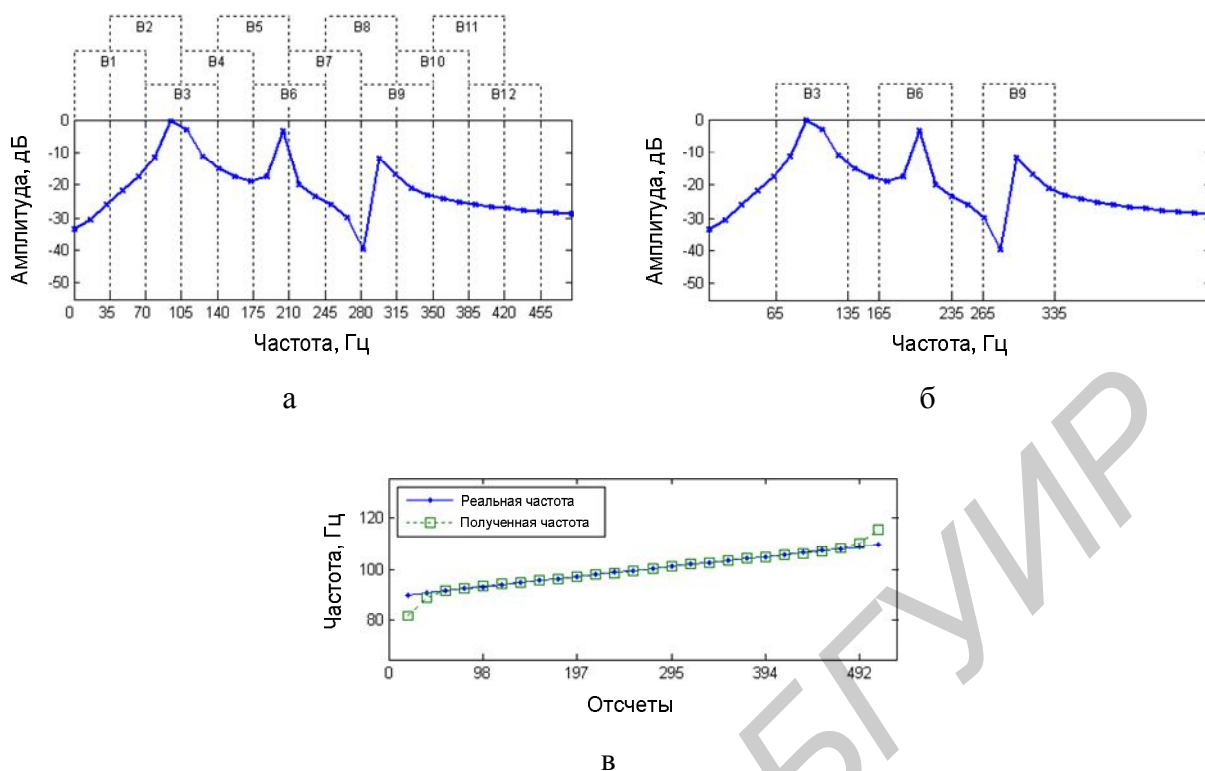
На первом шаге весь частотный диапазон анализируемого сигнала разбивается на полосы B_1, \dots, B_h (h – число полос) с центральными частотами $F_C^{B_1}, \dots, F_C^{B_h}$ соответственно. На каждом шаге соответствующие оценочные мгновенные частоты $f^{B_1}(n_c), \dots, f^{B_h}(n_c)$ определяются в момент времени, соответствующий центральному отсчету окна анализа n_c . Центральные частоты фильтров изменяются в соответствии с ними: $F_C^{B_x} = f^{B_x}(n_c)$.

После того как все энергетические максимумы локализованы, можно определить требуемые синусоидальные параметры (амплитудная огибающая, мгновенная частота и фаза). Параметры определяются при помощи выражений (2.51), (2.53)–(2.55). Так как в процессе анализа несколько фильтров могут локализовать одну и ту же компоненту, с целью исключения дублирования параметров выполняется сравнение центральных частот $F_C^{B_1}, \dots, F_C^{B_h}$.

Чтобы отбросить короткие и нестабильные компоненты, которые, очевидно, являются составляющими транзиентов или шума, выполняется слежение за полученными синусоидальными параметрами. Для выделения только достаточно продолжительных и стабильных значений сравниваются частоты и амплитуды компонент соседних фреймов.

Частотно-модулированные фильтры хорошо применимы к анализу речевых сигналов, т. к. вокализованная речь имеет гармоническую структуру. Предлагаемая процедура оценки гармонических параметров состоит из двух основных этапов:

- оценка частоты основного тона и классификация фрейма как вокализованного или невокализованного;
- оценка гармонических параметров с уточнением частоты основного тона.



а – первоначальное разбиение частотного диапазона; б – локализация компонент после второй итерации; в – оценка мгновенной частоты в полосе B_3

Рисунок 2.17 – Кратковременный анализ звукового сигнала с итеративной локализацией периодических компонент ($N = 512$), ($B_1 - B_{12}$) – полосы пропускания фильтров

Частота основного тона определяется предварительно, чтобы найти приблизительные частотные траектории гармоник, необходимые для синтеза ЧМ-фильтров. Задача оценки частоты основного тона сводится к нахождению тональной компоненты с минимальной частотой и энергией выше определенного амплитудного порога. Диапазон изменения частоты основного тона в речевых приложениях обычно находится в интервале 60 - 420 Гц. Значения частоты периодических компонент речевого сигнала в этом диапазоне изменяются довольно медленно, что делает возможным их оценку с использованием стационарных фильтров. Поэтому для определения контуров частоты основного тона целесообразно применение способа оценки мгновенных синусоидальных параметров с последующим выбором из их числа наиболее подходящих путем сравнения полученных параметров. 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 удастся выделить продолжительную синусоидальную компоненту, соответствующие фреймы сигнала классифицируются как невокализованные (рисунок 2.18).

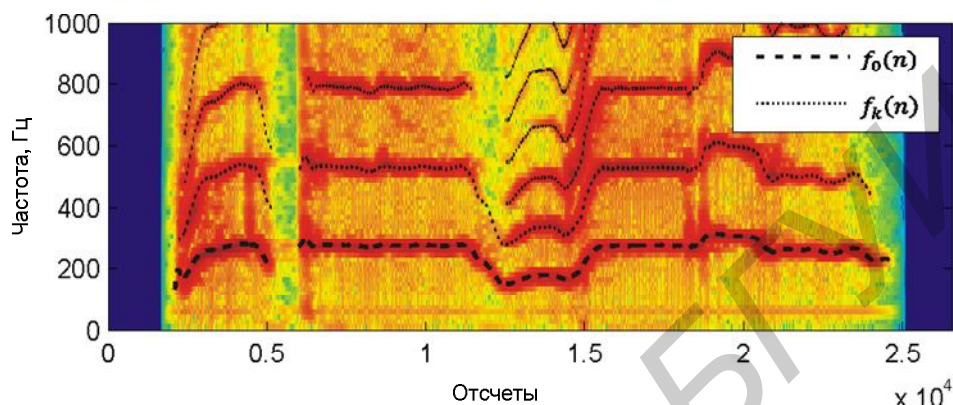


Рисунок 2.18 – Оценка контура частоты основного тона

ЧМ-фильтры после оценки частоты основного тона можно использовать для определения гармонических параметров вокализованных фреймов сигнала. Необходимые траектории центральных частот полос пропускания фильтров $F_c(n)$ вычисляются как мгновенная частота основного тона, умноженная на номер k соответствующей гармоники: $F_c^k(n) = k f_0(n)$. Процедура оценки проводится последовательно, начиная с первой гармоники и заканчивая последней. После оценки каждой следующей гармоники контур частоты основного тона уточняется с учетом полученных параметров по следующей формуле перерасчета:

$$f_0(n) = \sum_{i=0}^k \frac{f_i(n) \text{MAG}_i(n)}{(i+1) \sum_{j=0}^k \text{MAG}_j(n)}. \quad (2.63)$$

Таким образом, при оценке гармоник высокого порядка значения частоты основного тона становятся более точными, что позволяет правильно синтезировать частотно-модулированный фильтр анализа. Параметры гармоник вычисляются по соответствующим формулам (2.62), (2.53)–(2.55).

3 Моделирование переходных фрагментов сигнала методом согласованной подгонки

Алгоритмы согласованной подгонки используются в некоторых системах параметрического кодирования для компактного представления переходной составляющей различных сигналов [15, 27]. Основными задачами согласованной подгонки являются: поиск метода классификации, выбор наиболее значимых компонент сигнала, а также определение базисной функции для обеспечения эффективного описания входного сигнала с минимальной ошибкой.

Процесс оптимизации согласованной подгонки может быть основан на знании психоакустических свойств восприятия звуков человеком. Это позволяет адаптировать размер словаря в соответствии с особенностями человеческого слуха. Для того чтобы элементы словаря оптимальным образом поставить в соответствие отдельным сегментам сигнала (с точки зрения величины вносимых искажений), используется адаптивный психоакустический критерий. Такого рода подходы успешно применяются для моделирования сигнала затухающими синусоидами и вейвлетами [27].

3.1 Согласованная подгонка с динамически оптимизируемым словарем на основе психоакустики

Пусть $\{\varphi_n(t): n \in Z\}$ является множеством вейвлетных функций и $E \in \{(l, n): 0 \leq l \leq L, 0 \leq n \leq 2^l\}$ – узлы древовидной структуры. Тогда интервал $[0, 1)$ разделяется на двухэлементные интервалы $I_{l,n} = [n2^{-l}, (n+1)2^{-l}]$,

которые соответствуют определенному множеству узлов $(l, n) \in E$. В частности $\{\varphi_{l,n,k}(t): (l, n) \in E, k \in Z\}$, где $\varphi_{l,n,k}(t) \triangleq 2^{-l/2} X_n(2^{-l}t - k)$ – базовая форма в пространстве сигнала $span\{\varphi_0(t - k): k \in Z\}$. Каждый узел $(l, n) \in E$ дерева соответствует своей частотной полосе. В соответствии с двухэлементной древовидной структурой сигнал разделяется на полосы, которые близки к критическим [23]: $(l, n) \in E_{CB}, l = \{0, b\}$, где E_{CB} описывает ограниченную древовидную структуру вейвлет-пакета, b является максимальным номером уровней вейвлет-декомпозиции и зависит от частотного диапазона. Так, для обработки звуковых сигналов b равняется 8. В соответствии с E_{CB} частотная полоса $[0-22,05$ кГц] делится на 25 полос. Коренной узел $(l, n) = (0,0)$ данного дерева соответствует всему частотному диапазону сигнала.

Общий алгоритм может быть описан, как аппроксимация анализируемого сигнала $x(n)$ линейным пространством с атомами g_γ , выбираемыми из словаря вейвлет-пакета D [27]. Каждый вектор $g_\gamma \in D$ имеет индекс $\gamma = (l, n, k)$, причем $0 < l < \log_2(N)$, $0 < n < 2^l$, $0 < k < 2^{-l}N$, где N длина фрейма сигнала. Данные вектора имеют такие же свойства частотно-временной локализации, как и дискретная оконная функция, расширенная на 2^l и имеющая центр в $2^l(k + 1/2)$.

Метод моделирования переходных сегментов на основе согласованной подгонки с динамически оптимизируемым словарем при помощи психоакустического критерия состоит из двух основных этапов. Первый представляет собой пофреймную оптимизацию вейвлет-пакета, основанную на весовой функции энтропии входного сигнала $x(n)$ [23]; второй является алгоритмом согласованной подгонки с перцептуальной оценкой (рисунок 3.1). Результатами первого этапа являются: дерево вейвлет-пакета E_j , оптимизированное для текущего фрейма входного сигнала $x(n)$; вычисленный порог маскирования $T_{l,n}$ и порог временного маскера $F_{l,n}$ в узлах древовидной структуры вейвлет-пакета E_j [23]; скалограмма аудиторного возбуждения для входного сигнала $x(n)$, исполь-

зующая $T_{l,n}$ и $F_{l,n}$ для всех узлов. При первой итерации процедуры согласованной подгонки входной сигнал $x(n)$ разделяется при помощи банка фильтров, который реализует психоакустическое адаптивное дерево вейвлет-пакета для одного фрейма.

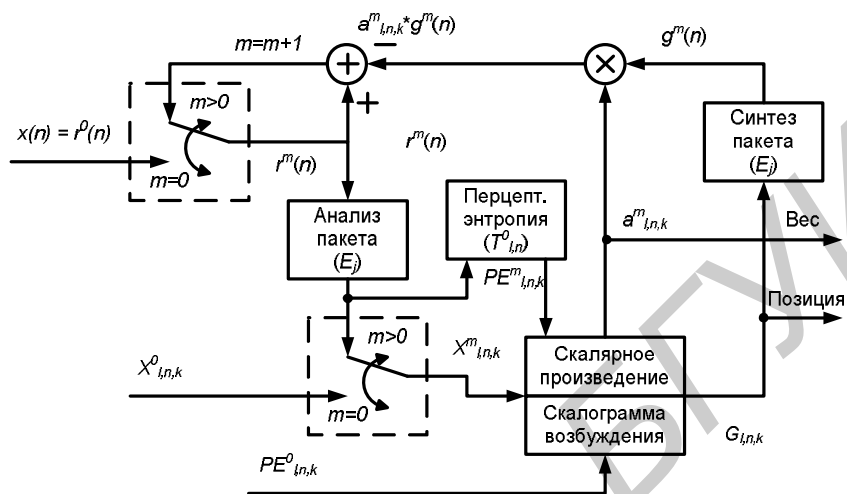


Рисунок 3.1 – Блок-схема согласованной подгонки с динамически оптимизируемым словарем на основе психоакустического критерия

Каждый вейвлет-коэффициент соответствует скалярному произведению входного сигнала и атома g_γ из словаря. Наиболее значимые компоненты можно определить при помощи классификации выбранных, перцептально значимых коэффициентов вейвлет-пакета, которые выбираются следующим образом. Каждый новый добавляемый коэффициент обеспечивает максимальный прирост в соответствии между скалограммами аудиторного возбуждения $G_{l,n,k}$, соответствующими исходному и полученному сигналам. Скалограммы сигналов строятся с использованием маскирующих порогов $T_{l,n}$ в вейвлет-области. Выбирается коэффициент с максимальным по модулю значением. Вклад вектора – $\alpha_\gamma^m \cdot g_\gamma^m(n)$, затем вычитается из сигнала $x(n)$ и процесс повторяется для остатка $r(n)$. После итерации t остаток $r^t(n)$ представляет собой:

$$r^m(n) = \begin{cases} x(n), & m = 0, \\ r^{m+1}(n) + \alpha_\gamma^m \cdot g_\gamma^m(n), & m \neq 0, \end{cases} \quad (3.1)$$

где α_γ^m – вес, соответствующий вектору оптимума $g_\gamma^m(n)$ итерации m ;

γ^m – индекс словаря вейвлет-пакета итерации m .

Вектор оптимума – это вектор с максимальным векторным произведением и сигналом остатка $\langle r^m, g_\gamma^m \rangle$. Для того чтобы найти компактное представление сигнала, каждый коэффициент вейвлет-пакета, имеющий наибольший вес возбуждения, добавляется в параметрическое описание. Вес возбуждения соответствует разнице между исходной и моделируемой скалограммами коэффициентов возбуждения.

Алгоритм согласованной подгонки может быть реализован так, как показано ниже.

Входные данные: оптимизированная древовидная структура вейвлет-пакета E_j для сигнала $x(n)$; порог маскирования $T_{l,n}$; временный маскер $F_{l,n}$ в узлах E_j ; скалограмма аудиторного возбуждения $G_{l,n,k}$, соответствующая сигналу $x(n)$.

Алгоритм согласованной подгонки:

- установить номер итерации $m = 0$;
- NEXT
- разместить $G_{l,n,k}$ и установить $G_{l,n,k} = 0$ для всех l, n, k в соответствии с древовидной структурой вейвлет-пакета E_j ;
- вычислить $PE_{l,n}^m$ для всех узлов (l, n) , используя $T_{l,n}$ (10);
- если $PE_{l,n}^m == 0 \forall (l, n, k) \in E_j$ СТОП;
- если $PE_{l,n}^m == 0$, то $X_{l,n,k}^m = 0$ для $k = \{0, K_{l,n} - 1\}$ узла (l, n) ;
- выбрать из $X_{l,n,k}^m$ значимые коэффициенты $X_{l,n,k}^{*m}$, имеющие наибольший вес возбуждения;

- создать скалограмму аудиторного возбуждения, соответствующую моделируемому сигналу, используя $T_{l,n}$ и $F_{l,n}^{m-1}$ для выполненной итерации и каждого нового значимого коэффициента $X_{l,n,k}^{*m}$;
- выбрать вес $\alpha_{l,n,k}^m = X_{l,n,k}^{*m}$, который улучшает соответствие между скалограммами (исходной и моделируемой);
- получить позицию коэффициента вейвлет-пакета: $l^* = l$, $n^* = n$, $k^* = k$;
- установить 1 на позиции (l^*, n^*, k^*) : $G_{l^*,n^*,k^*} = 1$;
- синтезировать атом $g^m(n)$ из G_{l^*,n^*,k^*} , используя обратный вейвлет-пакет с древовидной структурой E_j , соответствующей словарю вейвлет-пакета;
- вычислить остаточный сигнал $r^m(n)$ из $g^m(n)$ и $\alpha_{l,n,k}^m$;
- применить оптимизированный на основе текущего фрейма вейвлет-пакет с древовидной структурой E_j к остаточному сигналу $r^m(n)$;
- увеличить номер итерации $m = m + 1$;
- перейти к NEXT.

Преимуществом алгоритма является минимизация перцептуального искажения, определяемого при помощи перцептуально оптимизированного временно-частотного плана соответствующей декомпозиции вейвлет-пакета для выбора оптимального атома для каждой итерации подгонки. Более того, существует точный психоакустический критерий остановки описанной процедуры. Число итерации алгоритма согласованной подгонки для анализируемого фрейма определяется числом перцептуально важных коэффициентов вейвлет-пакета.

Сравнительная сходимость трех различных методов согласованной подгонки показана на рисунке 3.2. Описанная процедура обеспечивает наименьшее среднеквадратичное отклонение по сравнению с методами на основе затухающих синусоид и вейвлет-пакетов с избыточным словарем.

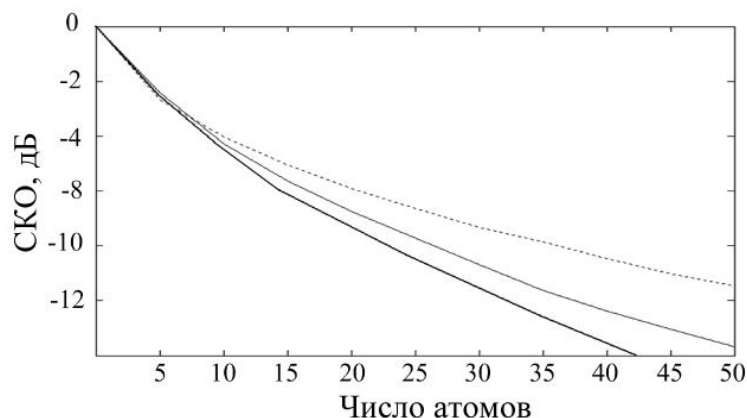


Рисунок 3.2 – Среднеквадратичное отклонение: затухающие синусоиды (пунктирная линия); вейвлет-пакет с избыточным словарем (тонкая сплошная линия); вейвлет-пакет, оптимизированный для каждого фрейма на основе психоакустики (толстая сплошная линия)

3.2 Определение переходных фреймов и применение метода согласованной подгонки

Способ определения переходных фреймов основан на предположении, что энергия остаточного сигнала после синусоидального моделирования (переходная плюс шумовая составляющие) резко изменяется при наличии выраженной переходной составляющей [27]. Изменения энергии могут быть общими либо представлять собой перераспределение между разными частотными полосами. Остаточный сигнал переводится в область вейвлетов, используя декомпозицию при помощи вейвлет-пакета второго уровня. Алгоритм вычисляет энергию вейвлет-коэффициентов в каждой полосе. Затем энергия каждой полосы фрейма i делится на энергию соседних фреймов $(i - 1)$ и $(i + 1)$, после чего полученный коэффициент сравнивается с порогом. Порог задается в зависимости от синусоидальных амплитуд, полученных на стадии гармонического анализа, для того чтобы отбросить маскированные переходные составляющие.

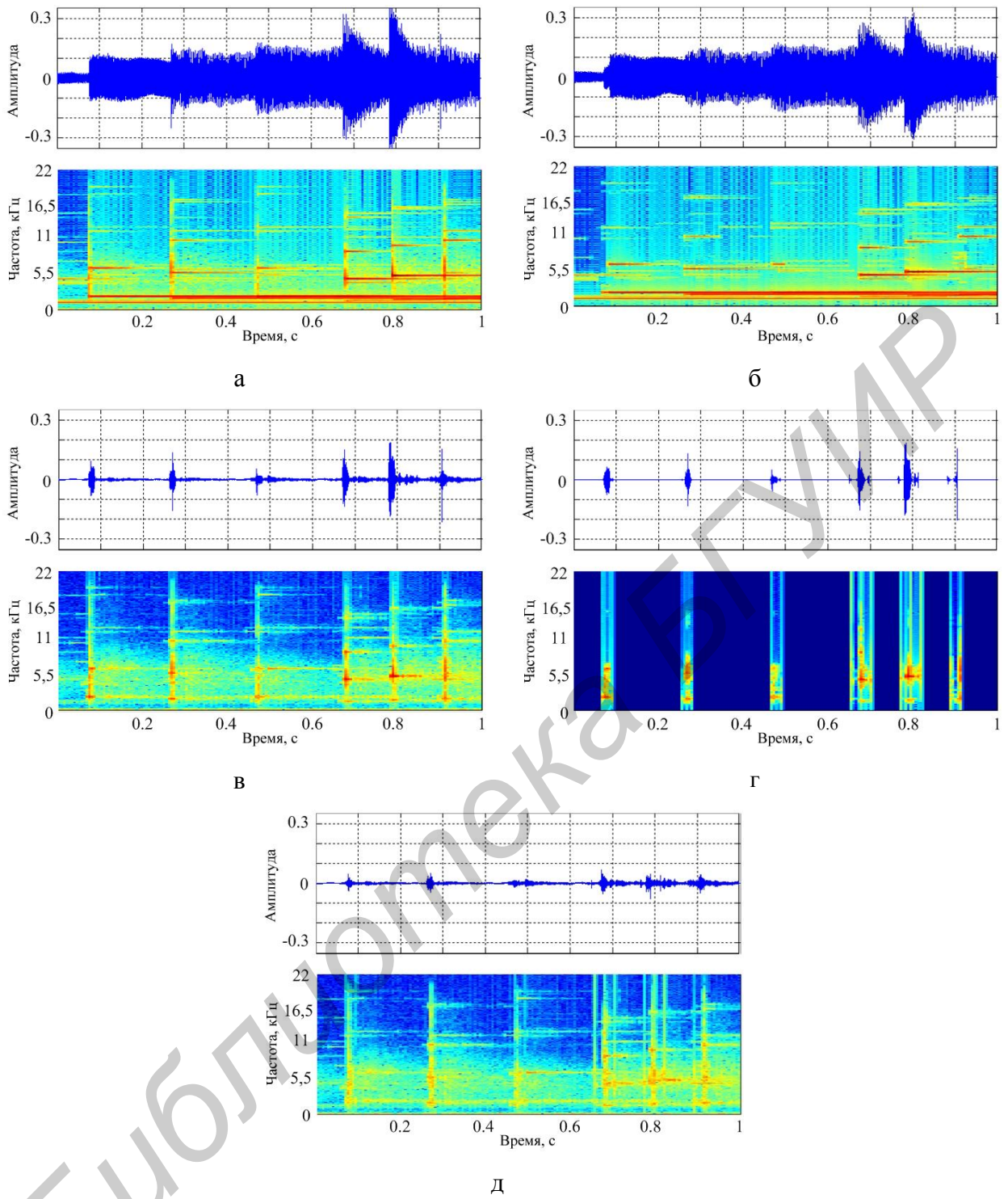
Далее приведена полная сепарация аудиосигнала на синусоидальную, переходную и шумовую составляющие. Каждая составляющая показана в виде

волновой формы и спектрограммы для качественной демонстрации, дающей наиболее полное представление о гибридном подходе.

На рисунке 3.3 приведены все сигналы, возникающие на различных этапах анализа. Переходные компоненты сигнала, которые были найдены из остатка синусоидального моделирования, обрабатывались вышеизложенным алгоритмом на основе перцептуально оптимизированных для каждого фрейма вейвлет-пакетов. Входной остаточный сигнал был разделен на фреймы длиной в 1024 отсчетов. В эксперименте были использованы фильтры 40-го порядка.

Воссозданные из параметрического описания переходные сегменты сигнала потребовали 20, 23, 18, 32, 36, 25, 27 и 20 атомов соответственно. В результате сложения синусоидальной, переходной и шумовой компонент получается исходный сигнал.

Гармонический анализ с трекингом параметров позволяет должным образом и без дополнительной обработки разделять входной сигнал на периодическую и остаточную компоненты, оставляя переходные сегменты сигнала в неизменном виде. Использование гармонического анализа в начале схемы параметризации существенно упрощает дальнейшую обработку сигнала (в частности определение переходных сегментов сигнала). Описанная методология для выбора наиболее значимых вейвлет-коэффициентов основывается на максимизации соответствия между аудиорными скалограммами, соответствующими исходному и синтезированному сигналам. Основное достоинство изложенного метода заключается в том, что словарь вейвлет-пакета оптимизирован на основе принципов психоакустики отдельно для каждого фрейма сигнала, что значительно понижает общее число коэффициентов, необходимых для достижения заданного уровня перцептуальных искажений.



а – входной сигнал; б – синтезированная периодическая часть;

в – остаток (переходная составляющая + шум);

г – переходная часть сигнала (моделирование методом согласованной подгонки);

д – остаток (шум)

Рисунок 3.3 – Сепарация сигнала для гибридного параметрического представления

Литература

- 1 Accardi, A. J. A modular approach to speech enhancement with an application to speech coding / A. J. Accardi, R.V. Cox // Acoustics, speech, and signal processing: proceedings of IEEE int. conf., Phoenix, Arizona, USA, March 15–19, 1999. – Phoenix, 1999. – P. 201–204.
- 2 Abe, T. Sinusoidal model based on instantaneous frequency attractors / T. Abe, M. Honda // IEEE Trans. on Audio, Speech, and Language processing. – 2006. – Vol. 14. – №4. – P. 1292–1300.
- 3 Ali, A. M. A. Robust auditory-based speech processing using the average localized synchrony detection / A. M. A. Ali // IEEE Transactions on Speech and Audio Processing. – 2002. – Vol.10. – №5. – P. 279–292.
- 4 Bagchi, S. The Nonuniform Discrete Fourier Transform and Its Applications in Filter Design / S. Bagchi, S. K. Mitra // IEEE Trans. Circuits Systems II. – 1996. – Vol. 43. – №6. – P. 422–433.
- 5 Boashash, B. Estimating and interpreting the instantaneous frequency of a signal / B. Boashash // Proceedings of the IEEE. – 1992. – Vol. 80. – №4. – P. 520–568.
- 6 Das, A. Low-rate multimode multiband spectral coding of speech / A. Das, A. Gersho // Int. J. of speech technology. – 1999. – Vol. 2. – P. 317–327.
- 7 Dutoit, T. An Introduction to Text-to-speech Synthesis / T. Dutoit. – The Netherlands : Kluwer Academic Publishers, 1997. – 326 p.
- 8 Edler, B. Parametric Audio Coding / B. Edler, H. Purnhagen // Signal processing: proceedings of 5-th int. conf. (ICSP'2000), Beijing, China, August 21–25, 2000. Beijing, 2000. – P. 21–24.
- 9 Gabor, D. Theory of communication / D. Gabor // Proc. IEE. – 1946. – Vol. 93. – №3. – P. 429–457.
- 10 George, E. B. Perceptual considerations in a low bit rate sinusoidal vocoder / E. B. George, M. J. T. Smith // Computers and communications: proceedings

of int. conf. (IPCCC'90), Phoenix, USA, March 21–23, 1990. – Phoenix, 1990. – P. 268–275.

11 George, E. B. Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model / E. B. George, M. J. T. Smith // IEEE Trans. on Speech and Audio Process. – 1997. – Vol. 5. – №5. – P. 389-406.

12 Gianfelici, F. Multicomponent AM–FM Representations: An Asymptotically Exact Approach / F. Gianfelici [et al.] // IEEE Transactions on Audio, Speech, and Language Processing. – 2007. – Vol.15, №3. – P. 823–837.

13 Laroche, J. HNS: Speech modification based on a harmonic+noise model / J. Laroche, Y. Stylianou, E. Moulines // Acoustic, Speech, and Signal Processing: proceedings of IEEE International conference ICASSP-93, Minneapolis, USA, 27–30 April 1993 – Minneapolis, 1993. – P. 550–553.

14 Levine, S. A sines+transients+noise audio representation for data compression and time/pitch scale modifications / S. Levine, J. Smith // Signal processing: proceedings of 105th AES convention, San Francisco, USA, 1998. – San Francisco, 1998. – Preprint №4781. – 21 p.

15 Mallat, S. Matching pursuits with time-frequency dictionaries / S. Mallat, Z. Zhang // IEEE. Transactions on Signal Processing. – 1993. – Vol. 41. – №12. – P. 3397–3415.

16 Maragos, P. Energy separation in signal modulations with application to speech analysis / P. Maragos, J. F. Kaiser, T. F. Quatieri // IEEE Trans. On Signal Process. – 1993. – Vol. 41. – №10. – P. 3024–3051.

17 McAulay, R. J. Computationally Efficient Sine-wave Synthesis and Its Application to Sinusoidal Transform Coding / R. J. McAulay, T. F. Quatieri // Acoustics, speech and signal processing: proceedings of IEEE int. conf., New York, USA, April 11–14, 1988. – New York, 1988. – P. 370–373.

18 McAulay, R. J. Pitch estimation and voicing detection based on sinusoidal model / R. J. McAulay, T. F. Quatieri // Acoustics, speech and signal processing :

proceedings of IEEE int. conf. (ICASSP'90), Albuquerque, New Mexico, USA, April 3–6, 1990. – Albuquerque, 1990. – P. 249–252.

19 McAulay, R. J. Sinusoidal coding / R.J. McAulay, T.F. Quatieri // *Speech Coding and Synthesis* / Eds. W.B. Kleijn, K.K. Paliwal. – Amsterdam : Elsevier, 1995. – P. 121–173.

20 McAulay, R. J. Speech analysis/synthesis based on a sinusoidal representation / R. J. McAulay, T. F. Quatieri // *IEEE Trans. on Acoust., Speech and Signal Processing*. – 1986. – Vol. ASSP-34. – P. 744–754.

21 Painter, T. Sinusoidal analysis-synthesis of audio using perceptual criteria / T. Painter, A. Spanias // *EURASIP Journal on Applied Signal Processing*. – 2003. – №1. – P. 15-20.

22 Petrovsky, Al . Hybrid signal decomposition based on instantaneous harmonic parameters and perceptually motivated wavelet packet for scalable audio coding / Al. Petrovsky, E. Azarov, A. Petrovsky // *Signal processing*. – 2011. – №91. – P. 1489-1504.

23 Petrovsky, A. Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfigurable system / A. Petrovsky, A. Krahe and A. Petrovsky // *Signal processing: proceedings of 114-th AES convention, Netherlands, Amsterdam, 2003*. – Amsterdam, 2003. – Preprint №5778. – 22 p.

24 Petrovsky, A. The order tracking front-end algorithms in the rotating machine monitoring systems based on the new digital low order tracking / A. Petrovsky, A. Stankevich, J. Balunowski // *On sound and vibration: proceedings of the 6th intern. congress, Copenhagen, Denmark, 1999*. – Copenhagen, 1999. – P. 2985–2992.

25 Petrovsky, A. A. The synthesis of high order digital band-pass filters with tunable center frequency and bandwidth / A. A. Petrovsky // *Signal processing: proceedings of 8-th europ. conf. (EUSIPCO-96), Trieste, Italy, September 10–13, 1996*. – Trieste, 1996. – P. 1527–1530.

26 Spanias, A. S. Speech coding: a tutorial review / A. S. Spanias // *Proc. IEEE*. – 1994. – Vol. 82. №10. – P. 1541–1582.

- 27 Vera-Candeas, P. Transient modeling by Matching-Pursuits with a wavelet dictionary for parametric audio coding / P. Vera-Candeas, N. Ruiz-Reyes and Roza-Zurera // IEEE Signal Processing Letters. – 2004. – Vol. 11. – №3. – P. 349–352.
- 28 Verma, T. Sinusoidal modeling using frame-based perceptually weighted matching pursuits / T. Verma, T. Meng // Acoustics, speech and signal processing: proceedings of IEEE intern. conf. (ICASSP-99), Phoenix, Arizona, USA, March 15–19, 1999. – Phoenix, 1999. – Vol. 2. – P. 981–984.
- 29 Ville, J. Theorie et application de la notion de signal analytique / J. Ville // Cables et Transmissions. – 1948. – Vol.2A. – №1. – P. 61–74.
- 30 Warped discrete Fourier transform in perceptual speech and audio processing / A. A. Petrovsky [et al.] // New trends in audio and video: proceedings of X int. symposium, Wroclaw, Poland, 16–18 September. – Wroclaw, 2004. – C. 153–159.
- 31 Weruaga, L. The fan-chirp transform for non-stationary harmonic signals / L. Weruaga, M. Kepesi // Signal Processing. – 2007. – Vol.87. - № 6. – P. 1–18.
- 32 Zhang, F. Harmonic transform / F. Zhang, G. Bi, Y.Q. Chen // IEEE Proc.-Vis. Image Signal Process. – 2004. – Vol.151. – №4. – P. 257–264.
- 33 Zubrycki, P. Accurate speech decomposition into periodic and aperiodic components based on discrete harmonic transform / P. Zubrycki, A. Petrovsky // Signal processing : proceedings of 15-th europ. conf. (EUSIPCO'2007), Poznan, Poland, 2007. – Poznan, 2007. – P. 2336–2340.
- 34 Zwicker, E. Psychoacoustics: facts and models / E. Zwicker, H. Fastl. – Berlin: Springer-Verlag, 1990. – 354 p.
- 35 Zwislocki, J. Analysis of some auditory characteristics // Handbook of Mathematical Psychology / J. Zwislocki. – New York : John Wiley and Sons Inc., 1965. – 267 p.

36 Основы психоакустики. Субъективные и объективные методы оценки разборчивости речи [Электронный ресурс] / И. Алдошина. – // Журнал Звукорежиссер, 2002. – Режим доступа: http://rus.625-net.ru/audioproducer/2002/05/theory_3.htm. – Дата доступа: 25.05.2006.

37 Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи : отчет о НИР (промежут.) / рук. темы А.А. Петровский. – Минск: БГУИР, 2002. – 71 с.

Библиотека БГУИР

Содержание

1 Синусоидальное представление сигнала.....	3
2 Частотно-временные преобразования	7
2.1 Преобразование Фурье	7
2.2 Гармоническое преобразование	9
2.3 Мгновенные гармонические параметры.....	21
2.4 Синтез стационарного фильтра анализа.....	29
2.5 Синтез частотно-модулированного фильтра анализа.....	35
3 Моделирование переходных фрагментов сигнала методом согласованной подгонки.....	43
3.1 Согласованная подгонка с динамически оптимизируемым словарем на основе психоакустики	43
3.2 Определение переходных фреймов и применение метода согласованной подгонки	48
Литература	51

Учебное издание

Петровский Александр Александрович
Азаров Илья Сергеевич

ЦИФРОВАЯ ОБРАБОТКА МЕДИАДАННЫХ

Учебно-методическое пособие

Редактор *Н. В. Гриневич*
Корректор *Е. Н. Батурчик*

Подписано в печать 16.05.2012. Формат 60x84 1/16. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 3,49. Уч.- изд. 3,1. Тираж 130 экз. Заказ 329.

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