

ИНФОРМАТИКА

УДК 621.382

**НЕПРЕРЫВНОЕ И ДИСКРЕТНОЕ ГАРМОНИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ
ДЛЯ ДЕКОМПОЗИЦИИ РЕЧЕВОГО СИГНАЛА НА ПЕРИОДИЧЕСКУЮ И
ШУМОВУЮ КОМПОНЕНТЫ**

И.С. АЗАРОВ, А.А. ПЕТРОВСКИЙ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь**Поступила в редакцию 7 июля 2008*

Показываются особенности сепарации речевого сигнала на периодическую и шумовую компоненты, предлагается точный способ сепарации. Вводится понятие гармонического преобразования и рассматриваются его основные свойства. Описывается алгоритм нахождения дискретного гармонического преобразования, связанный с оценкой изменений во времени частоты основного тона. Приводится пример работы алгоритма для фрагмента речи и демонстрируются преимущества кратковременного гармонического преобразования над кратковременным преобразованием Фурье для нахождения гармонических составляющих речевого сигнала.

Ключевые слова: речевой сигнал, анализ и синтез речевого сигнала, сепарация сигнала на шумовую и периодическую составляющие.

Введение

С задачами декомпозиции сигнала часто сталкиваются при создании систем сжатия и трансформации речи. Речевой сигнал может быть представлен в виде суммы трех основных составляющих: S-N-T синусоидальной (гармонической), шумовой и переходной (от английского Sines–Noise–Transients) [1]. Такая модель позволяет использовать специфику каждой из составляющих и применять соответствующие методы анализа и сжатия для каждой из них. Синусоидальная часть описывает сигнал, который можно описать в виде суммы конечного числа периодических функций (синус и косинус), шумовая — нерегулярный сигнал, который на слух воспринимается как шум, а переходная — непериодические (часто непродолжительные) фрагменты с резким изменением формы сигнала. В общем, можно рассматривать речь как комбинацию двух основных компонент: периодической и непериодической. Данная работа направлена на определение эффективности и точности выделения периодической части сигнала при помощи дискретного гармонического преобразования.

Широко используются модели, различающие только периодическую и шумовую компоненты. В таком случае сигнал обычно представляется в виде [2]

$$s(n) = \sum_{k=1}^K A_k(n) \cos \varphi_k(n) + r(n),$$

где A_k — мгновенная амплитуда k -й гармоники; K — число гармоник, присутствующих в сигнале; $r(n)$ — шумовая компонента; φ_k — мгновенная фаза k -й гармоники определенная как

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

где f_k — мгновенная частота k -й гармоники, F_s — частота дискретизации и $\varphi_k(0)$ — начальная фаза k -й гармоники.

Если для моделирования периодической составляющей используются лишь синусоиды с частотами, кратными частоте основного тона, т.е. $f_k = kf_0$, где f_0 — частота основного тона, то такая модель называется гармонической. Большинство моделей предполагают, что речь может быть вокализованной либо невокализованной. Вокализованная речь может быть представлена при помощи периодических функций, тогда как невокализованная моделируется как белый шум, проходящий через соответствующий фильтр, параметры которого определяются, например, при помощи линейного предсказания [3].

Гармонические модели эффективно используются для кодирования речевых сигналов и обеспечивают высокий коэффициент сжатия [3, 4]. В [5] была представлена модель, которая рассматривает вокализованную речь как сумму гармонически связанных синусоид с амплитудами и фазами, вычисляемыми непосредственно из спектра кратковременного преобразования Фурье (КВПФ), а невокализованная речь рассматривалась как сумма случайно распределенных синусоид со случайными начальными фазами. Позже данная модель была значительно усовершенствована [6].

Однако КВПФ имеет определенные ограничения и не может быть эффективно использовано для нахождения параметров гармонических компонентов в случае частотно-модулированного сигнала. Для более точного определения этих параметров используется гармоническое преобразование (ГП) [7].

1. Гармоническое преобразование

Как известно, преобразование Фурье (ПФ), представляющее сигнал $f(t)$ в виде синусоид различной частоты, определяется следующим образом:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt. \quad (1)$$

Исходный сигнал может быть восстановлен при помощи обратного преобразования Фурье (ОПФ):

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega. \quad (2)$$

ПФ может быть легко использовано для гармонического анализа сигналов, состоящих из фиксированных частотных компонент. Однако ПФ сложно применить в случае, когда частота сигнала изменяется во времени.

Одним из способов локализации гармонических компонентов в частотно-модулированном сигнале является преобразование Фурье с масштабированием. Сигнал из временной оси t отображается на масштабную ось u при помощи некоторой масштабной функции $p(t)$. При этом исходный сигнал растягивается или сжимается во времени для того, чтобы перед вычислением ПФ обеспечить стац-

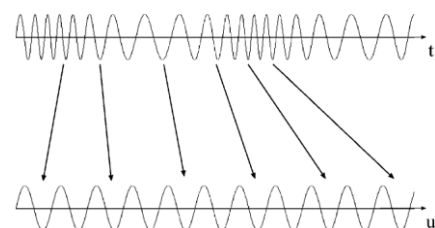


Рис. 1. Временное масштабирование сигнала

онарность частоты [8] рис. 1. Для речевого сигнала это означает стационарность частоты основного тона.

Другой реализацией данной операции может быть интеграция масштабирующей функции в само преобразование Фурье. Такое преобразование называется гармоническим преобразованием (ГП).

Изменяющийся во времени гармонический сигнал содержит гармоники, чьи мгновенные частоты можно описать следующим образом [9]:

$$c_k(t) = (k+1)c_0(t), \quad k = 1, 2, 3, \dots,$$

где $c_0(t)$ — частота основного тона и $c_k(t)$ — частота k -й гармоники. Вокализованная речь является одним из типичных примеров изменяющегося во времени гармонического сигнала, который может быть синтезирован при помощи линейной фильтрации периодической последовательности импульсов. Мгновенная частота основного тона может быть определена как мгновенная частота последовательности импульсов, которая изменяется во времени.

ГП функции $f(t)$ определяется как [10]

$$F_{\varphi_u(t)}(\omega) = \int_{-\infty}^{+\infty} f(t) \varphi'_u(t) e^{-j\omega\varphi_u(t)} dt, \quad (3)$$

где $\varphi_u(t)$ является фазовой функцией основного тона, деленной на ее мгновенную частоту; $\varphi'_u(t)$ — первая производная от $\varphi_u(t)$. $\varphi_u(t)$ должна быть дифференцируема и обратима на интервале $(-\infty; +\infty)$. Обратное гармоническое преобразование (ОГП) это:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_{\varphi_u(t)}(\omega) e^{j\omega\varphi_u(t)} d\omega. \quad (4)$$

Следует отметить, что ГП превращается в ПФ при $\varphi_u(t) = t$.

На рис. 2 показан пример преобразования сигнала из временной области в частотную при помощи преобразования Фурье и гармонического преобразования.

На рис. 3 схематично показано как соотносятся преобразование Фурье и гармоническое преобразование. Нестационарный гармонический сигнал (рис. 3, *b*) с использованием ГП выглядит в форме четких импульсов в частотной области (рис. 3, *a*), тогда как при применении ПФ в частотной области не наблюдается каких-либо пиков и все гармоники смешаны (рис. 3, *c*).

Рассмотрим условия существования ГП. Пусть $\varphi_u^{-1}(t)$ является обратной функцией $\varphi_u(t)$. Обозначим $z = \varphi_u(t)$ и $t = \varphi_u^{-1}(z)$. Подстановка $t = \varphi_u^{-1}(z)$ и $z = \varphi_u(t)$ в (1) дает

$$F_{\varphi_u(t)}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega\varphi_u(t)} d\varphi_u(t) = \int_{-\infty}^{+\infty} f(\varphi_u^{-1}(z)) e^{-j\omega z} dz. \quad (5)$$

Подставляя $t = \varphi_u^{-1}(z)$ и $z = \varphi_u(t)$ в (4), получим:

$$f(\varphi_u^{-1}(z)) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_{\varphi_u(t)}(\omega) e^{j\omega\varphi_u(t)} d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_{\varphi_u(t)}(\omega) e^{j\omega z} d\omega. \quad (6)$$

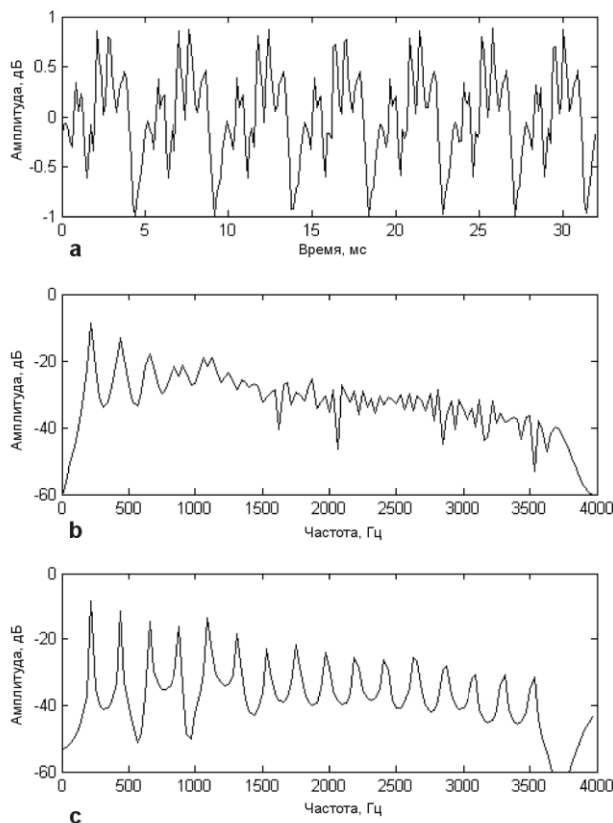


Рис. 2. Преобразование Фурье и гармоническое преобразование: *a* — вокализованный сегмент речи; *b* — преобразование Фурье; *c* — гармоническое преобразование

Сравнивая (5) и (6) с (1) и (2) соответственно, можно установить, что (5) и (6) являются ПФ и ОПФ $f(\varphi_u^{-1}(z))$ соответственно. Существование ПФ и ОПФ $f(\varphi_u^{-1}(z))$ гарантирует существование ГП и ОГП соответственно.

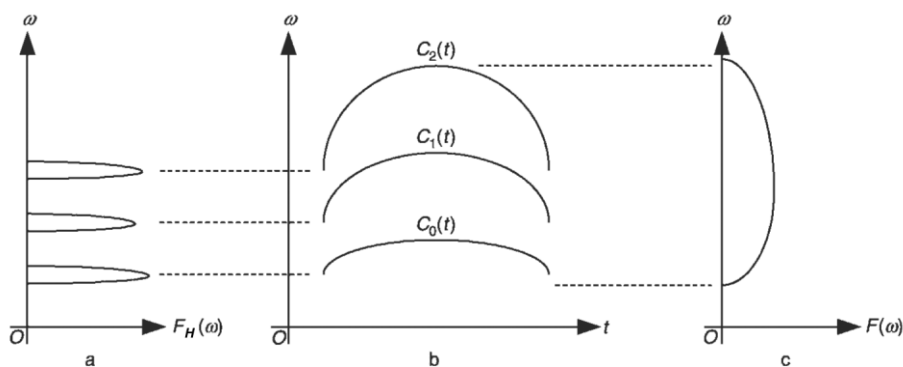


Рис. 3. Гармоническое преобразование и преобразование Фурье: *a* — гармоническое преобразование; *b* — частотно-временное представление; *c* — преобразование Фурье

Следует отметить, что $\varphi_u(t)$ должна быть известна перед вычислением ГП. Очевидно, что для данного сигнала функция $\varphi_u(t)$ единственна. Действенным методом поиска данной функции может быть поиск при помощи перебора. Весь сигнал делится на набор перекрывающихся сегментов. Для каждого сегмента выражение $\varphi_u(t)$ определяется как набор некоторых неизвестных коэффициентов. Внутри допустимого диапазона значений каждого из коэффициентов ГП сегмента вычисляется для всех возможных комбинаций. Теоретически наилучшая концентрация спектра в частотно-временной области определит наиболее желаемый набор. Та-

ким образом, набор коэффициентов, который позволяет достичь наилучшей концентрации, должен использоваться в функции $\varphi_u(t)$, которая в свою очередь может быть найдена путем сравнения концентрации всех гармонических преобразований. $\varphi_u(t)$ может быть найдена для каждого сегмента одним и тем же способом.

Поиск перебором ведет к большой вычислительной сложности, тем не менее существуют некоторые способы, для того чтобы упростить данный процесс, используя знания о сигнале. Вообще говоря, значение $\varphi_u(t)$ в определенный момент времени сильно зависит от одного локального текущего сегмента и не зависит от всего сигнала в целом. Если все сегменты являются достаточно короткими, справедливо предположить, что $\varphi_u(t)$ для каждого из них линейна. Это допущение ведет к существенному снижению вычислительной сложности. Так же знание частотной полосы основной или другой гармоники может помочь сузить область поиска и упростить вычисления.

Результат гармонического преобразования и функция $\varphi_u(t)$ должны быть сопоставлены с данными, имеющимися о сигнале, для того чтобы исключить возможные ошибки, связанные с неправильным использованием гармоники в качестве основной или потере алгоритмом основного тона в невокализованных сегментах или при наличии пауз в сигнале.

2. Свойства гармонического преобразования

Несколько свойств ГП перечислено в табл. 1, последняя колонка приводит условия, которым должна удовлетворять $\varphi_u(t)$. Большинство этих свойств, за исключением некоторых, аналогичны свойствам преобразования Фурье. Свойство линейности может быть легко получено из определения данного в (3) и (4). Для выполнения доказательств можно использовать взаимосвязь между ПФ и ГП, описанную в (5) и (6). Ниже приведены некоторые из них [7].

Таблица 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 $X_e(\omega) = \text{Re}[X_{\varphi_u(t)}(\omega)]$ $X_o(\omega) = j \text{Im}[X_{\varphi_u(t)}(\omega)]$	Нечетн.
Временная инверсия	$x(-t)$	$X_{\varphi_u(t)}(-\omega)$	Нечетн.
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	$\int_{-\infty}^{+\infty} x(t)\varphi'_u(t) ^2 dt$	$\int_{-\infty}^{+\infty} X_{\varphi_u(t)}(\omega) ^2 d\omega$	

2.1. Условие существования. Условие существования ГП:

$$\int_{-\infty}^{+\infty} |f(t)\varphi'_u(t)| dt < M < +\infty,$$

где M — положительная константа. Свойство может быть доказано через условие существования ПФ:

$$\int_{-\infty}^{+\infty} |f(t)| dt < M < +\infty.$$

Применение этого условия существования к (5) дает условие существования ГП:

$$\int_{-\infty}^{+\infty} |f(\varphi_u^{-1}(t))| dz = \int_{-\infty}^{+\infty} |f(t)\varphi'_u(t)| dt < M < +\infty.$$

2.2. Свертка во временной области. Свойство ПФ — свертка во временной области справедливо и для ГП. Таким образом, гармоническое преобразование свертки двух сигналов $x(t)$ и $y(t)$ является произведением их гармонических преобразований: $X_{\varphi_u(t)}(\omega)$ и $Y_{\varphi_u(t)}(\omega)$.

Поскольку $t = \varphi_u^{-1}(z)$ и $z = \varphi_u(t)$, имеем:

$$\begin{aligned} \int_{-\infty}^{+\infty} x(t)y(t)\varphi'_u(t)e^{-j\omega\varphi_u(t)} dt &= \int_{-\infty}^{+\infty} x(\varphi_u^{-1}(z))y(\varphi_u^{-1}(z))e^{-j\omega z} dz = \\ &= \int_{-\infty}^{+\infty} x(\varphi_u^{-1}(z))e^{-j\omega z} dz \int_{-\infty}^{+\infty} y(\varphi_u^{-1}(z))e^{-j\omega z} dz = \int_{-\infty}^{+\infty} x(t)\varphi'_u(t)e^{-j\omega\varphi_u(t)} dt \int_{-\infty}^{+\infty} y(t)\varphi'_u(t)e^{-j\omega\varphi_u(t)} dt \\ & \qquad \qquad \qquad t = \varphi_u^{-1}(\varphi_u(t)) \\ &= X_{\varphi_u(t)}(\omega)Y_{\varphi_u(t)}(\omega). \end{aligned}$$

2.3. Свойство энергии. ГП не сохраняет энергию сигнала. Для того чтобы это продемонстрировать, можно использовать теорему Парсеваля, справедливую для ПФ:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega,$$

где $X(\omega)$ — ПФ сигнала $x(t)$. Применяя теорему Парсеваля к (5), получим:

$$\int_{-\infty}^{+\infty} |F_{\varphi_u(t)}(\omega)|^2 d\omega = \int_{-\infty}^{+\infty} |f(\varphi_u^{-1}(z))|^2 dz = \int_{-\infty}^{+\infty} |f(\varphi_u^{-1}(\varphi_u(t)))|^2 d\varphi_u(t) = \int_{-\infty}^{+\infty} |f^2(t)\varphi'_u(t)| dt.$$

Таким образом, ГП не удовлетворяет теореме Парсеваля и не сохраняет энергию сигнала.

3. Способ вычисления гармонического преобразования вокализованного речевого сигнала при помощи его декомпозиции в частотной области

Как указывалось, ГП можно вычислить, при помощи ПФ, предварительно выполнив масштабирование речевого сигнала во временной области, обеспечивая стационарность частоты основного тона [8]. Для сохранения динамического диапазона при вычислении масштабированного сигнала можно использовать интерполирующий фильтр в области гармоник основного тона. Фильтр пропускает ограниченное число гармоник основного тона, необходимое для анализа, и отбрасывает гармоники более высоких порядков, находящихся вне интересующей области. Такой фильтр (фильтр низких порядков) [9] имеет в области гармоник основного тона постоянную частотную характеристику, во временной же области его частотная характеристика изменяется для каждого вычисляемого отсчета. Таким образом, каждый отсчет масштабированного сигнала может быть вычислен как сумма взвешенных импульсной характеристикой

отсчетов исходного сигнала. С помощью этого метода можно получить динамический диапазон более 100 дБ [9].

На основе данного подхода можно предложить способ вычисления гармонического преобразования при помощи декомпозиции сигнала в частотной области. Сигнал рассматривается как сумма спектральных компонентов, которые можно получить, используя дискретное преобразование Фурье (ДПФ). С помощью ДПФ сигнал представим в виде

$$s(n) = \sum_{k=0}^{N/2} \text{MAG}[S(k)] \cos\left(\frac{2k\pi n}{N} + \varphi[S(k)]\right), \quad (7)$$

где MAG — амплитуда; k — частота; φ — начальная фаза, вычисляемые при помощи соответствующих соотношений:

$$S(k) = \frac{1}{N} \sum_{n=0}^{N-1} s(n) e^{-j2\pi nk/N},$$

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

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

Выражение (7) можно записать в виде суммы сверток сигнала с функциями косинуса соответствующих частот f :

$$s(n) = \sum_{k=0}^{N/2} s(n) \cdot \cos\left(\frac{2\pi}{F_s} n f_k\right),$$

где F_s — частота дискретизации сигнала.

Если заменить фиксированную частоту f_k частотным диапазоном (от F_1 до F_2), то получим следующее выражение для одного компонента:

$$s_{F_1, F_2}(n) = s(n) \cdot \int_{F_1}^{F_2} \cos\left(\frac{2\pi}{F_s} n f\right) df. \quad (8)$$

Имея произвольный вектор разбиения частотного диапазона $\bar{F} = (F_1, F_2, \dots, F_L)$, удовлетворяющий условию $F_1 \leq F_2 \leq \dots \leq F_L$, где $F_1 = 0$, F_L — частота Найквиста и $L-1$ — число компонент, можно представить исходный сигнал в виде

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

Рассмотрим более подробно выражение (8). Вычислим соответствующий интеграл:

$$\int_{F_1}^{F_2} \cos\left(\frac{2\pi}{F_s} n f\right) df = \begin{cases} F_2 - F_1, & n = 0 \\ \frac{F_s}{n\pi} \cos\left(\frac{n\pi}{F_s}(F_2 + F_1)\right) \sin\left(\frac{n\pi}{F_s}(F_2 - F_1)\right), & n \neq 0. \end{cases}$$

Таким образом, частотный компонент, выделенный из сигнала в диапазоне (F_1, F_2) можно представить в аналитической форме:

$$s_{F_1, F_2}(n) = \sum_{i=0}^{N-1} \frac{s(i) F_s}{(n-i)\pi} \cos\left(\frac{(n-i)\pi}{F_s}(F_2 + F_1)\right) \cdot \sin\left(\frac{(n-i)\pi}{F_s}(F_2 - F_1)\right),$$

что можно записать следующим образом:

$$s_{F_c, F_\Delta}(n) = A(n) \cos\left(\frac{2\pi}{F_s} n F_c\right) + B(n) \sin\left(\frac{2\pi}{F_s} n F_c\right), \quad (9)$$

где

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

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

$$F_c = \frac{F_2 + F_1}{2}, \quad F_\Delta = \frac{F_2 - F_1}{2}.$$

Выражение (9) представляет собой синусоиду с частотной и амплитудной модуляцией:

$$s(n)_{F_c, F_\Delta} = C(n) \cos\left(\frac{2\pi}{F_s} F_c n + \alpha(n)\right),$$

где $C(n) = \sqrt{A^2(n) + B^2(n)}$; $\alpha(n) = \arctan\left(-\frac{B(n)}{A(n)}\right)$.

Из последнего выражения вычислим значения мгновенной частоты F , амплитуды MAG и фазы ϕ синусоиды для любого отсчета n :

$$F_{F_c, F_\Delta}(n) = \frac{\alpha(n + \Delta t) - \alpha(n)}{2\pi \Delta t} F_s + F_c, \quad MAG_{F_c, F_\Delta}(n) = C(n), \quad \phi_{F_c, F_\Delta}(n) = 2\pi F_c n + \alpha(n).$$

Таким образом, определив произвольный частотный диапазон (F_1, F_2) , можно представить часть сигнала, находящуюся в этом диапазоне, в виде синусоиды с определенными мгновенными параметрами (частотой, амплитудой и фазой). Выбирая вектор разбиения частотного диапазона \bar{F} соответствующим образом, исходя из характеристик анализируемого сигнала, легко получить мгновенные параметры гармоник основного тона.

Выражение (9) позволяет учесть операцию временного масштабирования. Для этого следует согласовать центр частотного диапазона с частотой основного тона:

$$s_{F_c, F_\Delta}(n) = A(n) \cos\left(\frac{2\pi}{F_s} \phi_k(n)\right) + B(n) \sin\left(\frac{2\pi}{F_s} \phi_k(n)\right),$$

где $\phi_k(n) = \left(\sum_{i=0}^n F_0(n) - \sum_{i=0}^{N/2} F_0(n)\right)k$; $F_0(n)$ — мгновенная частота основного тона;

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

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

Соответствующие выражения для частоты F , амплитуды MAG и фазы ϕ будут выглядеть следующим образом:

$$F_{F_c, F_\Delta}(n) = \frac{\alpha(n + \Delta t) - \alpha(n)}{2\pi \cdot \Delta t} F_s + F_0 \cdot k,$$

$$MAG_{F_c, F_\Delta}(n) = C(n),$$

$$\phi_{F_c, F_\Delta}(n) = 2\pi F_0 kn + \alpha(n),$$

$$\text{где } C(n) = \sqrt{A^2(n) + B^2(n)}, \alpha(n) = \arctan\left(-\frac{B(n)}{A(n)}\right).$$

Полученные выражения для определения мгновенных гармонических параметров являются непрерывными функциями и позволяют определять гармонические параметры сигнала в произвольные моменты времени, не ограничиваясь моментами, которые соответствуют дискретным отсчетам сигнала.

4. Дискретное гармоническое преобразование

Определение и способ вычисления дискретного гармонического преобразования (ДГП) были предложены в [10]. Поскольку в речевом сигнале частота основного тона изменяется достаточно медленно, целесообразно допустить, что внутри небольшого временного интервала зависимость частоты основного тона от времени линейна. Мгновенная фаза $\varphi(t)$ синусоиды с линейным изменением частоты определяется известной формулой (для простоты начальная фаза опущена):

$$\varphi(t) = 2\pi \left(f_0 t + \frac{\varepsilon t^2}{2} \right), \quad (10)$$

где f_0 — начальная частота и $\varepsilon = (\Delta f_0 / T)$ — изменение частоты основного тона, деленное на длину сегмента (т.е. на время, за которое происходит данное изменение). Переходя к дискретным сигналам, для сегмента длиной в N отсчетов ($T = N / F_s$), где F_s — частота дискретизации, формула (10) может быть переписана в виде

$$\varphi(n) = 2\pi \left(\frac{f_0 n}{F_s} + \frac{\Delta f_0 n^2}{2NF_s} \right). \quad (11)$$

Начальная частота основного тона внутри данного сегмента может быть выражена как

$$f_0 = f_c - \frac{af_c}{2}, \quad a = \frac{\Delta f_0}{f_c}, \quad (12)$$

где f_c — среднее значение частоты основного тона в данном сегменте длиной N . Подставляя f_0 и Δf_0 в (11), из формул (12) получим

$$\varphi(n) = \frac{2\pi f_c}{F_s} \alpha_a(n), \quad \alpha_a(n) = n \left(1 - \frac{a}{2} + \frac{an}{2N} \right). \quad (13)$$

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

$$f_c = \frac{F_s}{N}. \quad (14)$$

В ГП средние частоты спектральных линий приравняем к частотам спектральных линий ДПФ. Используя (14) и (13), получим

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

Таким образом, мы можем определить ДГП для сигналов с линейным изменением частоты основного тона [10]:

$$S(k) = \sum_{n=0}^{N-1} s(n) \alpha'(n) e^{-j \frac{2\pi k}{N} \alpha(n)} \quad (15)$$

где $s(n)$ — дискретный сигнал, а $\alpha'(n)$ определяется как

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

Тогда обратным ДГП будет

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{j \frac{2\pi k}{N} \alpha(n)} \quad (16)$$

Основные свойства ДГП приведены в табл. 2.

Таблица 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	$\sum_{n=0}^{N-1} x(n) \alpha'(n) ^2$	$\sum_{k=0}^{N-1} X_{\alpha(n)}(k) ^2$

5. Декомпозиция речевого сигнала на периодическую и шумовую компоненты

5.1. Определение частоты основного тона. Преобразования, данные в (15) и (16), позволяют анализировать сигнал в гармонической области в случае, когда известна зависимость частоты основного тона от времени. Однако при анализе речи и средняя частота основного тона и его приращение неизвестны. Блок-схема алгоритма определения частоты основного тона показана на рис. 4 [10].

Прежде всего, алгоритм находит приращение частоты основного тона, путем анализа КДГП для различных фазовых функций, т.е. фазовых функций с различным параметром a . Оптимальное значение определяется как минимум меры пологости спектра:

$$\arg \min_a SFM(a) = \frac{\sqrt{\prod_{k=0}^{N-1} |STHT(a, k)|}}{\frac{1}{N} \sum_{k=0}^{N-1} |STHT(a, k)|},$$

где $STHT(a, k)$ — гармонический спектр данного сегмента речи, полученный для определенного коэффициента a . Минимальная пологость спектра означает наивысшую его концентрацию, что в рассматриваемом случае обеспечивает максимальное приближение к существующему изменению частоты основного тона. После нахождения изменения определяется сама частота основного тона.

На первом шаге алгоритм определяет "кандидатов" — возможные значения частоты основного тона f_i путем определения пиков (значений локальных максимумов) КДГП на основании алгоритма, предложенного в [10], учитывая, что средняя частота основного тона должна находиться между 50 и 450 Гц. Для каждой возможной частоты основного тона алгоритм ищет ее гармоники. В 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:

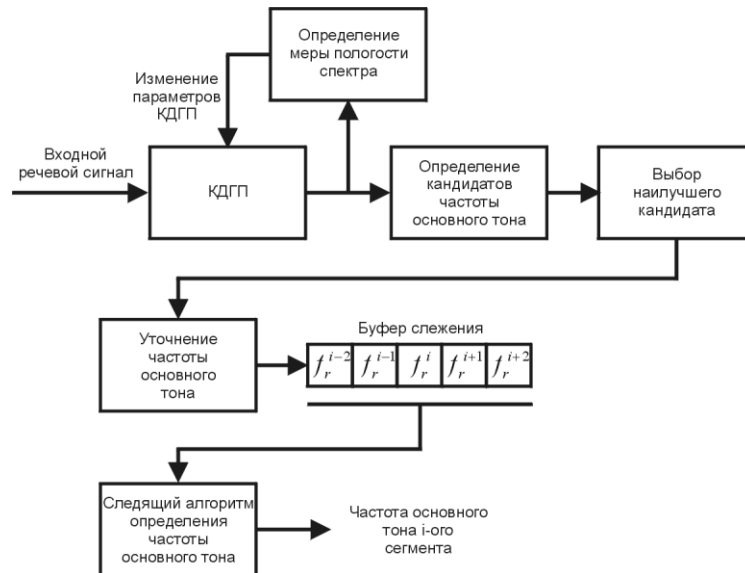


Рис. 4. Алгоритм определения частоты основного тона

$$r = \frac{\left| \sum_{n=1}^{n_{h \max}} |a_n|^2 \right|}{n_{h \max}},$$

где a_n — амплитуда n -й гармоники частоты основного тона; $n_{h \max}$ — максимально возможный номер гармоники для выбранного "кандидата". Таким образом, в качестве частоты основного тона выбирается "кандидат" с максимальным значением r .

Окончательно частота основного тона уточняется при помощи следующей формулы:

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

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 для хранения значений частоты основного тона соседних сегментов сигнала. Окончательная оценка частоты основного тона выполняется для сегмента, находящегося в центре буфера. Таким образом, вычисление частоты основного тона происходит с задержкой на несколько сегментов. В качестве следящего алгоритма используется медианная фильтрация, которая достаточно проста в использовании и обеспечивает устойчивость алгоритма против грубых ошибок.

5.2. Вычисление периодической и шумовой компоненты. Декомпозиция сигнала выполняется во временной области. Периодическая компонента вычисляется при помощи КДГП,

а шумовая находится как результат вычитания из исходного сигнала полученной периодической компоненты.

Периодическую часть сигнала можно записать следующим образом:

$$h(n) = \sum_{k=1}^K A_k \cos(k\varphi(n) + \varphi_k(0)), \quad (17)$$

где A_k — амплитуда k -й гармоники; $\varphi(n)$ — мгновенная фаза k -й гармоники, определенная в (13) со средней частотой основного тона f_c ; $\varphi_k(0)$ — начальная фаза k -й гармоники.

Для получения всех гармонических параметров (амплитуды, частоты и начальной фазы) непосредственно из спектрального представления используется специальный случай ГП [11]. Вариант ДГП, дающий спектр, масштабированный в соответствии с частотой основного тона, определяется как

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

где f_r — уточненная частота основного тона, $k = 1..K$, а K — число гармоник основного тона. Амплитуды и фазы могут быть вычислены непосредственно из коэффициентов $S(k)$:

$$A_k = \sqrt{\operatorname{Re} S(k)^2 + \operatorname{Im} S(k)^2},$$

$$\varphi_k(0) = -\arctan \frac{\operatorname{Im} S(k)}{\operatorname{Re} S(k)},$$

где Re и Im означают действительную и мнимую части $S(k)$ соответственно. Периодическая компонента генерируется при помощи формулы (17), а шумовая компонента вычисляется следующим образом:

$$r(n) = s(n) - h(n).$$

Пример сепарации сигнала показан на рис. 5.

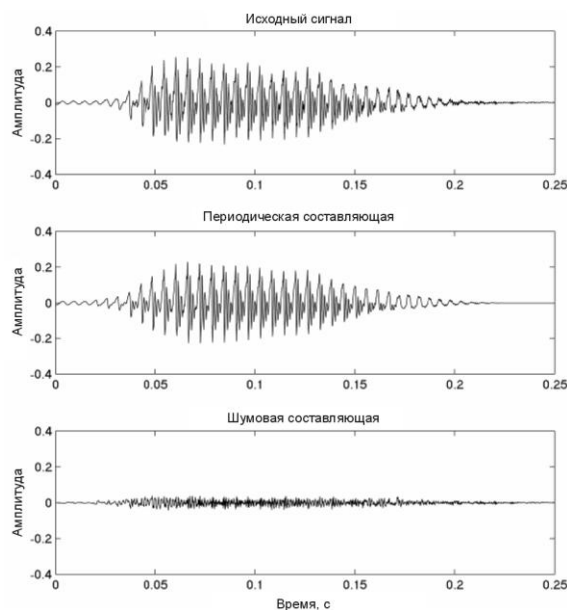


Рис. 5. Пример сепарации речевого сигнала

5.3. Экспериментальные результаты. Как было сказано, использование гармонического преобразования позволяет с большей точностью выделять параметры частотно-модулированных сигналов. Это можно наглядно продемонстрировать при помощи сравнительных спектрограмм (рис. 6).

Спектрограмма, построенная при помощи КДГП (рис. 6,*b*) позволяет проследить частотную траекторию гармоник более высоких порядков, по сравнению со спектрограммой, построенной при помощи КВПФ (рис. 6,*a*).

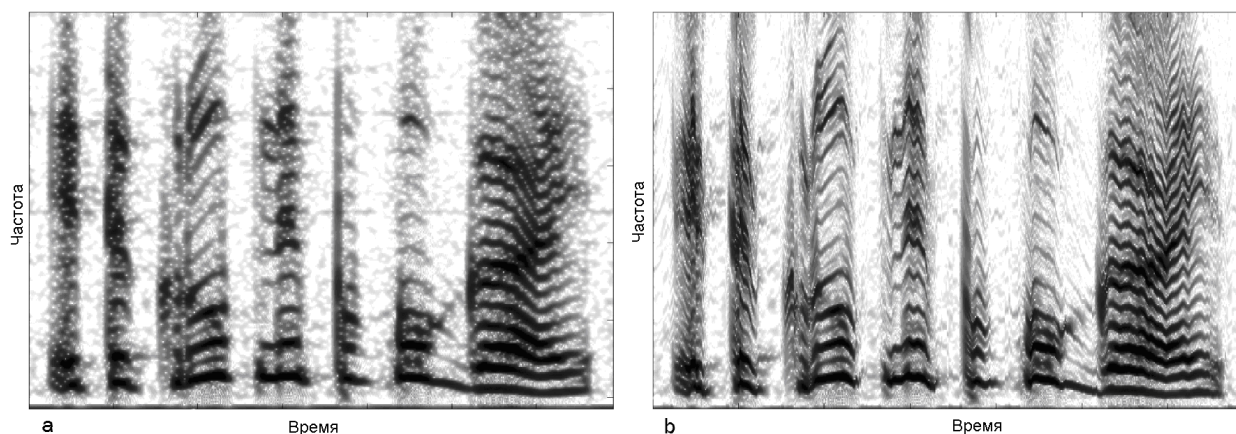


Рис. 6. Сравнительные спектрограммы вокализованного сегмента речи: *a* — спектрограмма, построенная при помощи КВПФ; *b* — спектрограмма, построенная при помощи КДГП

Для оценки описанного алгоритма нахождения КДГП проводился ряд экспериментов на синтетических сигналах. Процедура тестирования заключается в следующем: было создано два речевых набора (один для мужского голоса со средней частотой 120 Гц, другой для женского со средней частотой 200 Гц). Для того чтобы проверить работу КДГП, были использованы различные изменения частоты основного тона в обоих случаях. Приращение частоты основного тона выбирались случайным образом, однако оно составляло не более 30% от средней частоты основного тона в заданном сегменте. Тестирование проводилось для сигналов с разными соотношениями гармоники/шум путем различного зашумления сигнала. Результаты эксперимента приведены в табл. 3.

Таблица 3. Результат сепарации синтезированных сигналов

Средняя частота основного тона	Исходное отношение гармоники/шум, дБ	Полученное отношение гармоники/шум, дБ	Сепарированная периодическая часть сигнала сигнал/шум, дБ
120	∞	59,6	59,6
120	30	29,2	33,9
120	10	10,6	20,9
120	5	5,7	16,1
120	0	1,05	11,3
200	∞	68,3	68,3
200	30	30,3	38,9
200	10	10,6	21,5
200	5	5,54	16,3
200	0	1,06	11,2

В колонке "Исходное отношение" приведено отношение в исходном сигнале. После определения периодической и шумовой части сигнала вычислялось отношение гармоники/шум. Среднее значение этого измерения показано в колонке "Полученное отношение". Затем качество полученной периодической части определялось вычислением отношения сигнал/шум, которое определено как отношение энергии сигнала к энергии сигнала ошибки. Сигнал ошибки определялся как разность между исходной и полученной периодическими компонентами.

6. Выводы

Предложен новый способ сепарации речевого сигнала, основанный на гармоническом преобразовании. Гармоническое преобразование трансформирует исходный сигнал, рассматриваемый как сумму частоты основного тона и его гармоник, в спектр, позволяющий с большой точностью выделить параметры гармонической части исходного сигнала. Введено понятие гармонического преобразования и приведены его основные свойства. Введено понятие дискретного гармонического преобразования для сигналов с линейным изменением частоты основного тона.

Основные преимущества использования КДГП заключаются в следующем: данное преобразование позволяет определить изменения частоты основного тона, причем сама частота основного тона может быть неизвестна, а также спектр сигнала в таком виде, что каждая из гармоник (в том числе и гармоники высоких порядков) имеет четкую форму, даже если частота основного тона не является стационарной.

CONTINUOUS AND DISCRETE HARMONIC TRANSFORM FOR PERIODIC / NOISE DECOMPOSITION OF SPEECH SIGNAL

I.S. AZAROV, A.A. PETROVSKY

Abstract

An accurate method for signal harmonic/noise separation is proposed. Some peculiarities of the signal separation are shown. The definition of harmonic transform along with its basic features is introduced. The algorithm for harmonic transform evaluating, involving pitch track estimation is described. The example of harmonic estimation is given and advantages of the harmonic transform over Fourier transform are demonstrated.

Литература

1. *Levine S., Verma T. and Smith J.O.* // Proc. of the International Conference on Acoustic, Speech, and Signal Processing. Seattle, 1998.
2. *Kondoz A.M.* // Digital speech: coding for low bit rate communication systems. New York, 1996.
3. *Маркел Д., Грей А.* Линейное предсказание речи. М., 1980.
4. *Spanias A.S.* // Proc. IEEE. 1994. Vol. 82, No 10. P. 1541–1582.
5. *McAulay R.J., Quatieri T.F.* // "Sinusoidal Coding" in Speech Coding and Synthesis (W. Klein and K. Palival, eds.). Amsterdam: Elsevier Science Publishers, 1995.
6. *George E.B., Smith M.J.T.* // IEEE Trans. on Speech and Audio Processing. 1997. Vol. 5, No. 5. P. 389–406.
7. *Zhang F., Bi G., Chen Y.Q.* // IEEE Trans. on Vis. Image Signal Processing. August 2004. Vol. 151, No. 4. P. 257–264.
8. *Abe T., Honda M.* // IEEE Trans. on Audio, Speech, and Language processing. July 2006. Vol. 14, No. 4. P. 1292–1300.
9. *Petrovsky A., Stankevich A., Balunowski J.* // Proc. of the 6th Intern. congress "On sound and vibration", ICSV'99. Copenhagen, Denmark, 1999. P. 2985–2992.
10. *Zubrycki P., Petrovsky A.* // Proc. of the 15th European Signal Process. Conf., (EUSIPCO-2007). Poznan, 2007. P. 2336–2340.
11. *Sercov V., Petrovsky A.* // Proc. of the 9th European Signal processing conference, EUSIPCO '98, Vol. II, Sep. 8–11, 1998. Rhodes, Greece. P. 1137–1140.
12. *Sercov V., Petrovsky A.* // Proc. of the 6th European Conf. on Speech Communication and Technology EUROSPEECH '99, Budapest, Hungary, 1999. P. 1479-1482.