

УДК 621.382

И.С. Азаров, А.А. Петровский

**АНАЛИЗ РЕЧИ НА ОСНОВЕ
МГНОВЕННОГО ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ**

Рассматривается новый способ анализа вокализованного речевого сигнала, основанный на мгновенном линейном предсказании (МЛП). Параметры модели линейного предсказания (ЛП) вычисляются из мгновенных параметров гармонической модели (ГМ) при помощи специального способа конверсии. Показаны методы описания сигнала средствами ГМ и ее связь с МЛП. Исследуется применение рассматриваемого способа для анализа речевого сигнала. Предлагается способ вычисления коэффициентов ЛП на основе быстрого преобразования Фурье. Приводятся полученные экспериментальные результаты.

Введение

В настоящее время модель ЛП является одной из самых используемых техник в задачах цифровой обработки речи. Несмотря на то, что модель применяется уже достаточно давно (самой первой работой о применении математического аппарата предсказания случайных процессов к речевым сигналам считается статья Н.Н. Акинфиева, опубликованная в 1957 г. [1]), она, как и способы ее использования, постоянно совершенствуется усилиями современных ученых. Перечислим основные свойства модели, которые объясняют ее популярность:

- описание процесса речеобразования адекватным способом, который может быть физически интерпретирован;
- определение частотной огибающей вокального (голосового) тракта;
- определение параметров (формантных частот), характеризующих индивидуальность диктора и фонетические признаки;
- наличие альтернативного представления параметров ЛП в виде линейных спектральных частот, которое устойчиво к векторному квантованию;
- обеспечение разделения голосового тракта и источника;
- простота в использовании, низкая вычислительная сложность.

В настоящее время существует большое количество работ, посвященных способам вычисления параметров предсказания, получения спектральных оценок с их помощью и определения формантных траекторий.

ЛП используется в задачах распознавания речи, поскольку оно обеспечивает возможность получения фонетической оценки речевого сегмента. По этой же самой причине ЛП широко применяется в задачах конверсии голоса и синтеза речи по тексту.

Параметры модели ЛП (далее коэффициенты предсказания) могут быть описаны через коэффициенты отражения секций акустической трубы, что обеспечивает возможность определения параметров голосового тракта диктора. Это свойство может применяться для диагностики различных заболеваний.

Существует два основных метода определения коэффициентов предсказания: автокорреляционный и ковариационный. Оба метода используют представление сигнала во временной области. Коэффициенты предсказания определяют частотную характеристику фильтра, характеризующего состояние голосового тракта в определенный момент времени. Однако, с одной стороны, данный момент не может быть точно локализован, с другой стороны, точность сильно зависит от стационарности исследуемого сигнала. Другими словами, данные методы вычисления обеспечивают получение некоторой средней оценки анализируемого участка сигнала во временно-частотной области.

Идея, положенная в основу данной работы, заключается в использовании ГМ для вычисления коэффициентов предсказания. Согласно этой модели [2] вокализованный сигнал может быть описан в виде суммы периодических компонент, каждая из которых имеет мгновенную

частоту, амплитуду и фазу. ГМ обеспечивает точную локализацию энергии сигнала во временной и частотной областях. Каждому набору мгновенных гармонических параметров можно поставить в соответствие набор коэффициентов предсказания, наиболее точно аппроксимирующий спектральную огибающую. Таким образом, локализованный сигнал можно описать при помощи МЛП с мгновенными коэффициентами предсказания. Это позволит получить ряд преимуществ, в частности значительно улучшить временное разрешение анализа.

Нахождение набора коэффициентов предсказания, наиболее близкого набору мгновенных параметров ГМ, далее будем называть конверсией параметров ГМ в коэффициенты предсказания.

1. Модель линейного предсказания

Модель ЛП основывается на предположении, что любой отсчет речевого сигнала $s(n)$ можно приближенно оценить линейной комбинацией некоторого числа p предшествующих ему отсчетов, что приводит к следующему соотношению:

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n),$$

где a_1, a_2, \dots, a_p – коэффициенты предсказания; $u(n)$ – нормализованная последовательность возбуждения (ошибка предсказания) и G – коэффициент усиления [3]. В z -области коэффициенты предсказания задают передаточную функцию

$$H(z) = \frac{G}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{G}{A(z)}.$$

Ошибка предсказания $e(n)$ определяется как разность между исходными и приближенно вычисленными (предсказанными) отсчетами:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k).$$

Основная задача ЛП – определение набора коэффициентов предсказания, которые минимизируют $e(n)$. Как было сказано выше, существует два основных решения, которые называются автокорреляционным и ковариационным методами соответственно.

Автокорреляционный метод предполагает, что $s(n) = 0$ вне заданного сегмента сигнала $0 \leq n < N$, и минимизирует ошибку предсказания на бесконечном интервале. Задача сводится к решению системы

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

где $j = 1, 2, \dots, p$ и $r(l) = \sum_{n=0}^{N-1-l} s(n)s(n+l)$ – автокорреляционная функция ($l \geq 0$).

Ковариационный метод минимизирует ошибку предсказания на заданном интервале и может быть описан системой [1]

$$\sum_{i=1}^p a_i c_{ij} = -c_{0j}, \quad (1)$$

где $j = 1, 2, \dots, p$ и $c_{ij} = \sum_{n=p}^{N-1} s(n-i)s(n-j)$.

Автокорреляционный метод гарантирует получение стабильного фильтра $H(z) = \frac{G}{A(z)}$

(т. е. все корни $A(z)$ лежат внутри единичной окружности), в то время как ковариационный метод не может этого гарантировать и, как следствие, реже применяется в обработке речи, хотя может давать более точные результаты [1].

Поведение фильтра-предсказателя $A(z)$ в частотной области можно определить, вычислив передаточную функцию на единичной окружности [4]:

$$H(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^p a_k e^{-j\omega k}} = \frac{G}{A(e^{j\omega})}.$$

2. Гармоническая модель

ГМ основана на предположении, что сигнал может быть представлен в виде суммы периодических функций (синусоид), каждая из которых имеет свою амплитуду, частоту и фазу. На практике часто сигнал разделяют на периодическую часть и шумовую. Таким образом, он может быть записан в виде [5]

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

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

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

Здесь 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. Различают вокализованные и невокализованные сегменты речи. Вокализованная речь может быть эффективно представлена при помощи периодических функций, тогда как невокализованная моделируется как белый шум, проходящий через соответствующий фильтр.

ГМ позволяет эффективно описать вокализованный речевой сигнал и часто является оптимальной формой его представления [6]. Вычисление параметров ГМ – сложная задача. На сегодняшний день существует большое количество подходов к ее решению. Отметим некоторые из них:

- непрерывное и дискретное гармоническое преобразование [7, 8];
- фан-чирп-преобразование [9];
- методы, использующие фильтр или банки фильтров [10].

В данной работе метод гармонического анализа речи основан на следящем фильтре [11, 12]. Сигнал рассматривается как сумма спектральных компонентов, которые можно получить, используя дискретное преобразование Фурье (ДПФ). С помощью ДПФ сигнал представим в виде

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

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

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

где F_s – частота дискретизации сигнала.

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

$$s_{F_1, F_2}(n) = s(n) * \int_{F_1}^{F_2} \cos\left(\frac{2\pi}{F_s} n f\right) df. \quad (3)$$

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 (3). 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:

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

После соответствующих преобразований выражение принимает вид

$$s_{F_c, F_\Delta}(n) = A(n) \cos\left(\frac{2\pi}{F_s} n F_c\right) + B(n) \sin\left(\frac{2\pi}{F_s} n F_c\right), \quad (4)$$

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

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

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

Можно показать, что выражение (4) 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 отсчета n :

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

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} исходя из характеристик анализируемого сигнала, легко получить мгновенные параметры гармоник основного тона.

Выражение (4) позволяет учесть операцию временного масштабирования. Для этого следует согласовать центр частотного диапазона с частотой основного тона:

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

где $\varphi_k(n) = \left(\sum_{i=0}^n F_0(n) - \sum_{i=0}^{N/2} F_0(n)\right)k$;

$F_0(n)$ – мгновенная частота основного тона;

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

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

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

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

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.

3. Конверсия параметров гармонической модели в коэффициенты предсказания

Задача конверсии параметров гармонической модели в коэффициенты предсказания может быть сформулирована следующим образом: для набора гармонических параметров A_1, \dots, A_K и f_1, \dots, f_K найти набор коэффициентов предсказания a_1, \dots, a_p , минимизирующий ошибку предсказания для сигнала

$$s(n) = \sum_{k=1}^K A_k \cos\left[\frac{2n\pi f_k}{F_s} + \varphi_k(0)\right]. \quad (5)$$

Очевидным решением является синтез сигнала некоторой конечной длины по формуле (5), а затем применение автокорреляционного или ковариационного метода. Однако данное решение не самое эффективное с точки зрения вычислительной сложности и, кроме того, не самое точное. Используя знания о гармонической структуре данного сигнала, можно предложить более эффективное решение.

Рассмотрим синусоиду с постоянной амплитудой A , постоянной частотой f и нулевой начальной фазой:

$$s(n) = A \cos(fn).$$

Ошибка предсказания представляет собой синусоиду и может быть записана следующим образом:

$$\begin{aligned} e(n) &= s(n) - \tilde{s}(n) = A \cos(fn) - \sum_{i=1}^p A a_i \cos(f(n-i)) = \\ &= A \left[1 - \sum_{i=1}^p a_i \cos(fi) \right] \cos(fn) - A \sum_{i=1}^p a_i \sin(fi) \sin(fn). \end{aligned}$$

Минимизирование ошибки предсказания в данном случае может быть достигнуто минимизацией амплитуды синусоиды. Можно минимизировать ее квадрат на бесконечном интервале:

$$A_e^2 = A^2 \left(\left(1 - \sum_{i=1}^p a_i \cos(fi) \right)^2 + \left(\sum_{i=1}^p a_i \sin(fi) \right)^2 \right).$$

Если дан вектор амплитуд A_1, A_2, \dots, A_K и вектор частот f_1, f_2, \dots, f_K , можно записать сумму квадратов амплитуд сигнала-ошибки как

$$A_e^2 = \sum_{k=1}^K A_k^2 \left(\left(1 - \sum_{i=1}^p a_i \cos(f_k i) \right)^2 + \left(\sum_{i=1}^p a_i \sin(f_k i) \right)^2 \right). \quad (6)$$

Чтобы минимизировать это значение, можно использовать хорошо известный способ, а именно нахождение частных производных для всех переменных a_i с последующим решением системы линейных уравнений. Получим следующую систему:

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

где $j=1, 2, \dots, p$ и $q(l) = \sum_{k=1}^K A_k \cos(f_k l)$ ($l \geq 0$).

Система (7) идентична системе в автокорреляционном методе и может быть решена соответствующим образом. Существует множество хорошо изученных подходов, обеспечивающих получение решения простыми и эффективными способами [1, 3, 4].

Ранг матрицы

$$Q = \begin{pmatrix} q(0) & q(1) & \dots & q(p-1) \\ q(1) & q(0) & \dots & q(p-2) \\ \dots & \dots & \dots & \dots \\ q(p-1) & q(p-2) & \dots & q(0) \end{pmatrix}$$

не превышает $2K$, следовательно, нет нужды выбирать $p > 2K$. Тем не менее, если p по условиям решаемой задачи должно быть больше $2K$, то систему можно уменьшить и после ее решения добавить $p - 2K$ нулей к полученным коэффициентам предсказания.

Рассмотрим, как предлагаемый (гармонический) способ оценки коэффициентов предсказания соотносится с другими методами (рис. 1). Анализируемый сигнал представляет собой синтезированный гармонический сигнал, состоящий из двух синусоид $A_1 = 1$, $A_2 = 1$ и $f_1 = 200$, $f_2 = 400$ Гц с частотой дискретизации 8 кГц. Анализ проводился автокорреляционным, ковариационным и предлагаемым методами с порядком предсказания $p = 8$. Для того чтобы обеспечить наиболее точные результаты, для автокорреляционного метода использовалось 400 отсчетов (50 мс), в то время как для ковариационного – 40 отсчетов (5 мс), что точно соответствует одному периоду частоты основного тона. Полученные коэффициенты приведены в таблице.

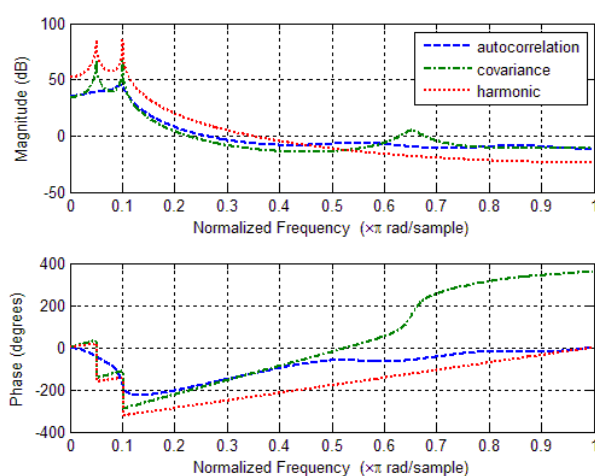


Рис.1. Сравнительный спектральный анализ гармонического сигнала, выполненный различными методами

Сравнительная оценка коэффициентов предсказания для гармонического сигнала

Автокорреляционный метод	Ковариационный метод	Гармонический метод
Коэффициенты предсказания		
-1,9816	-1,7037	-3,8775
0,6289	0	5,7574
0,4239	0	-3,8775
0,2187	1,7242	1,0000
0,0145	0	0
-0,1881	-1,4448	0
-0,3884	0	0
0,2889	0,4437	0
Сигнал-остаток, дБ		
-40,36	-278,99	-246,79

Видно, что ковариационный метод имеет наименьший сигнал-остаток, однако он отражает гармоническую структуру анализируемого сигнала не так точно, как предлагаемый метод. Автокорреляционный метод показывает сильно сглаженный спектр и требует увеличения длительности интервала анализа для получения более точных результатов.

4. Стабильность фильтра предсказателя

Практически все речевые приложения, использующие линейное предсказание, требуют, чтобы фильтр $\frac{1}{A(z)}$ был стабильным. Стабильность гарантируется, если все корни $A(z)$ полинома находятся внутри единичной окружности. Предлагаемый способ определения коэффициентов предсказания в некоторых случаях (когда $p \geq 2K$) может дать в результате полином $A(z)$ с корнями, лежащими на единичной окружности. Для получения стабильного фильтра требуется изменить параметры системы (7). Зная, что автокорреляционный метод обеспечивает получение устойчивого фильтра, можно использовать это свойство.

Рассмотрим сигнал $s(n) = t(n)$, $1 \leq n \leq 2p$, который удовлетворяет следующим условиям:

$\sum_{n=1}^p |t(n)| > 0$ и $\sum_{n=p+1}^{2p} |t(n)| = 0$. Можно обеспечить стабильность фильтра, если искать коэффициенты предсказания, одновременно минимизирующие ошибку для сигнала $t(n)$ и выражение (6).

Для минимизации ошибки предсказания можно воспользоваться (1), в результате получим следующую систему:

$$\sum_{i=1}^p a_i c_{ij} = 0, \quad (8)$$

где $j = 1, 2, \dots, p$ и $c_{ij} = \sum_{n=p}^{N-1} s(n-i)s(n-j)$.

Объединяя (7) и (8), найдем искомое решение, обеспечивающее получение стабильного фильтра:

$$\sum_{i=1}^p a_i [c_{ij} + q(|i-j|)] = -q(j),$$

где $j = 1, 2, \dots, p$.

Выбор $t(n)$ может быть произвольным, однако существует такой набор значений, который не приводит к изменению всей системы (7), вместо этого изменяя лишь один ее элемент $q(0)$.

Рассмотрим $t(n)$ следующего вида:

$$t(n) = \begin{cases} 0, n \neq p; \\ \Delta d, n = p. \end{cases}$$

Тогда c_{ij} (8) примет вид

$$c_{ij} = \begin{cases} 0, i \neq j; \\ \Delta d, i = j. \end{cases}$$

В результате система (7) может быть записана как

$$\sum_{i=1}^p a_i q(|i-j|) = -q(j), \quad (9)$$

где $j = 1, 2, \dots, p$;

$$q(l) = \begin{cases} \sum_{k=1}^K A_k \cos(f_k l) + \Delta d, l = 0; \\ \sum_{k=1}^K A_k \cos(f_k l), l > 0. \end{cases}$$

Значение Δd определяет, насколько близко корни $A(z)$ будут находиться к единичной окружности. Чем ближе Δd к нулю, тем ближе фильтр к нестабильности, тем не менее, данный параметр не следует выбирать слишком большим, поскольку это существенно ухудшит точность предсказания.

5. Быстрое вычисление коэффициентов предсказания

Одним из применений предлагаемого метода конверсии параметров может быть получение коэффициентов предсказания непосредственно из ДПФ. Для этого можно воспользоваться системой (7) или (9). Необходимые гармонические параметры могут быть получены из ДПФ следующим образом:

$$A_k(f_k) = \left| \frac{1}{N} \sum_{n=0}^{N-1} s(n) e^{-j2\pi n f_k / N} \right|.$$

По своим результатам метод очень близок к автокорреляционному (рис. 2). Однако, если спектр сигнала не пологий и имеет острые пики, предлагаемый метод может быть точнее.

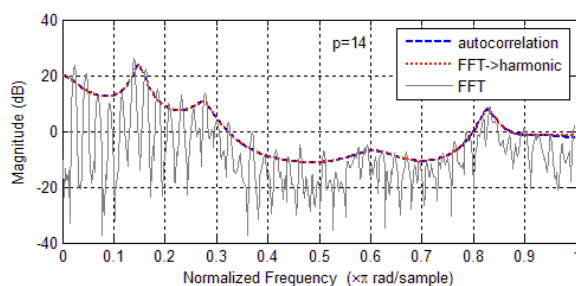


Рис. 2. Спектральный анализ при помощи автокорреляционного метода и метода оценки коэффициентов ЛП из преобразования Фурье: анализируется фрагмент вокализованной речи длиной 512 отсчетов с частотой дискретизации 8 кГц

Получение коэффициентов предсказания из преобразования Фурье может быть использовано во многих приложениях для уменьшения вычислительной сложности и увеличения точности предсказания, если анализируемый сигнал является гармоническим.

6. Экспериментальный анализ речевого сигнала

Для вокализованной речи ГМ располагает мгновенными гармоническими параметрами для любого момента времени. Эти параметры являются непрерывными и неявно отражают мгновенную формантную структуру сигнала и состояние вокального тракта. Для определения траекторий формантных частот широко используются методы на основе линейного предсказания, которые требуют большой длительности анализируемых сегментов и не могут обеспечить получение точных результатов, особенно для нестационарных сегментов сигнала. Комбинируя ГМ и ЛП путем оценки гармонических параметров с последующей их конверсией в коэффициенты предсказания, можно значительно улучшить частотное и временное разрешения. На рис. 3 показана оценка формантной структуры сигнала, полученная разными методами.

Исходный сигнал представлял собой фрагмент вокализованной речи, оцифрованный с частотой дискретизации 8 кГц. Использовался автокорреляционный метод (рис. 3, а) с длиной сегмента в 512 отсчетов, 97%-м перекрытием и окном Хэмминга. Для конверсии гармонических параметров (рис. 3, б) использовались наборы мгновенных гармонических параметров, полученные способом, описанным в разд. 2. Порядок предсказателя $p = 14$ в обоих случаях. Видно, что формантные траектории на рис. 3, а имеют разрывы и не так четко выражены, как на рис. 3, б.

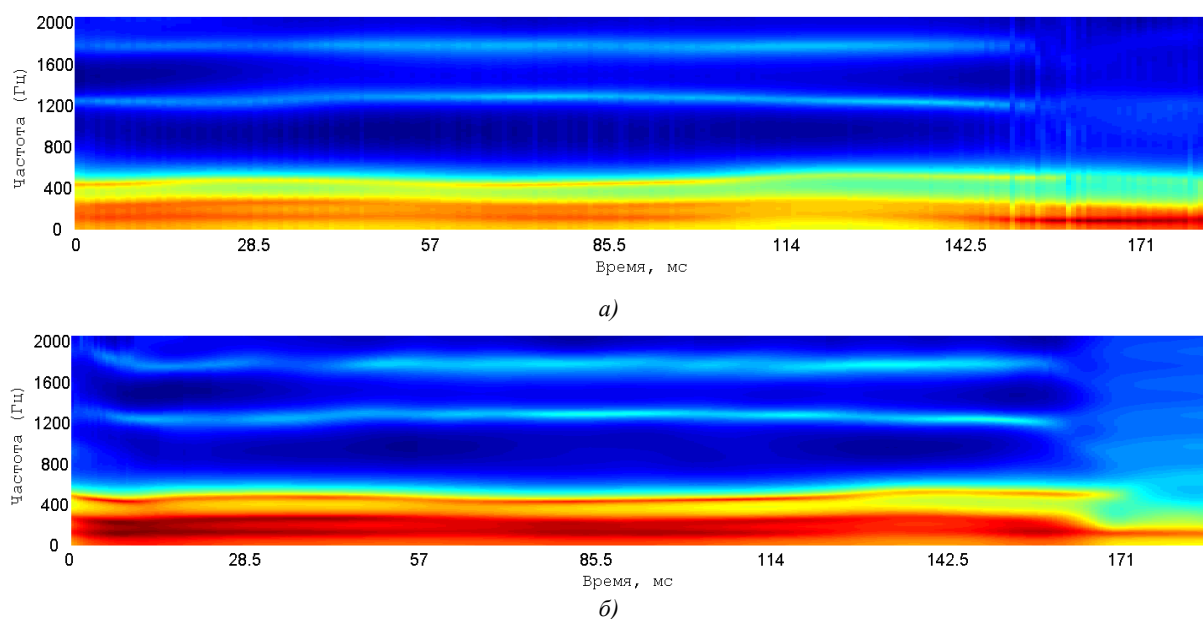


Рис. 3. Формантный анализ вокализованной речи: *а)* автокорреляционный метод; *б)* предлагаемый метод

Рассмотрим использование предлагаемого метода еще на одном примере (рис. 4). Здесь анализируется продолжительный отрезок речи.

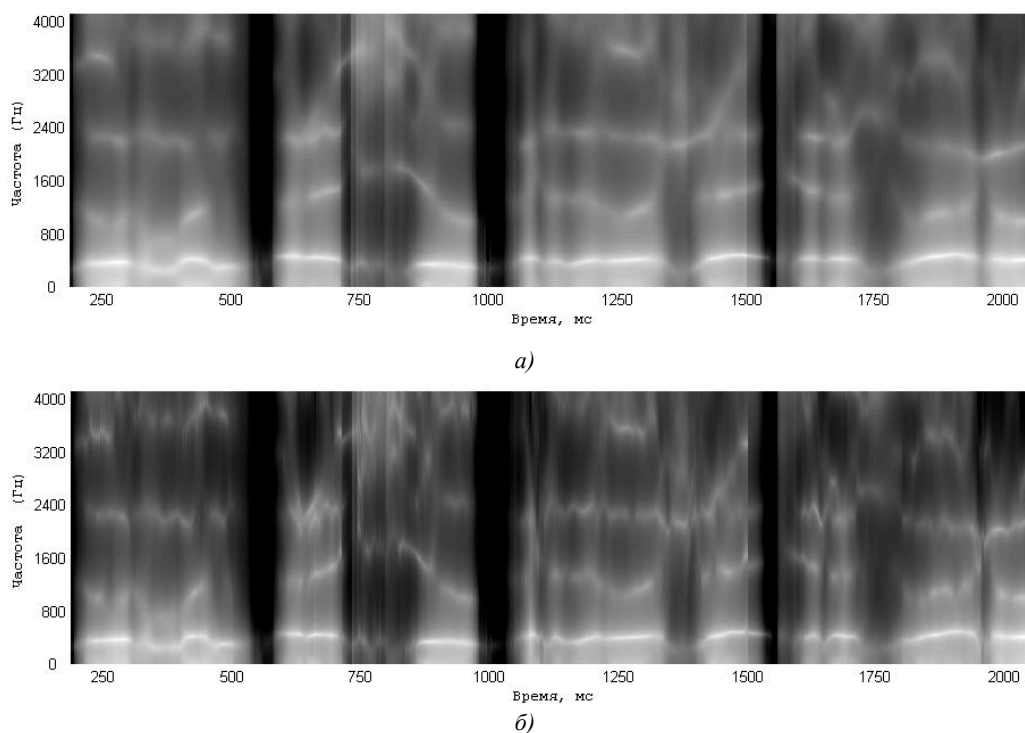


Рис. 4. Формантный анализ продолжительного фрагмента речи:
а) автокорреляционный метод; *б)* предлагаемый метод

В данном примере гармонический анализатор проводил классификацию фреймов речи на вокализованные и невокализованные. Вокализованные фреймы анализировались предлагаемым методом, невокализованные – автокорреляционным (рис. 4, б). В сравнении с чистым автокорреляционным методом (рис. 4, а) предлагаемый метод анализа показывает более контрастную спектрограмму, что свидетельствует о более точной локализации энергии сигнала в частотной области. Также заметно, что на рис. 4, б формантные траектории имеют более сложную форму, хотя повторяют общий контур формантных траекторий на рис. 4, а.

Заключение

В настоящей работе был предложен способ анализа вокализованного речевого сигнала на основе МЛП. Коэффициенты предсказания вычисляются для каждого момента времени из мгновенных параметров ГМ. Такой подход позволяет получить более высокое частотно-временное разрешение анализа и существенно уменьшить сигнал ошибки предсказания по сравнению с автокорреляционным методом. Ковариационный метод может обеспечить получение минимально возможного сигнала ошибки, однако он не может быть использован для оценки параметров речевого тракта и формантных частот, поскольку искажает гармоническую структуру сигнала в частотной области. Таким образом, для анализа и представления гармонического сигнала при помощи МЛП предлагаемый метод является наиболее подходящим. Конечная точность оценки параметров сигнала зависит от точности представления сигнала ГМ.

В работе также был предложен способ быстрой оценки коэффициентов предсказания из ДПФ. Точность оценки в этом случае очень близка к автокорреляционному методу, однако может быть выше для гармонических сигналов.

Список литературы

1. Маркел, Дж.Д. Линейное предсказание речи / Дж.Д. Маркел, А.Х. Грэй. – М. : Связь, 1980. – 308 с.
2. Almeida, L.B. Nonstationary spectral modeling of voiced speech / L.B. Almeida, J.M. Tribolet // IEEE Trans. on Acoust., Speech and Sig. Proc. – 1983. – Vol. ASSP-31, № 3. – P. 664–678.
3. Rabiner, L. Fundamentals of speech recognition / L. Rabiner, B.H. Juang. – N.-J. : Prentice Hall, 1993. – 507 p.
4. Huang, X. Spoken language processing / X. Huang, A. Acero, H.W. Hon. – N.-J. : Prentice Hall, 2001. – 980 p.
5. Kondo, A.M. Digital speech: coding for low bit rate communication systems / A.M. Kondo. – N.-Y. : John Wiley & Sons, 1996. – 441 p.
6. Spanias, A.S. Speech coding: a tutorial review / A.S. Spanias // IEEE Proc. – 1994. – Vol. 82, № 10. – P. 1541–1582.
7. Zhang, F. Harmonic transform / F. Zhang, G. Bi, Y.Q. Chen // IEEE Proc. Vis. Image Signal Process. – Vol. 151, № 4. – 2004. – P. 257–264.
8. Zubrycki, P. Accurate speech decomposition into periodic and aperiodic components based on discrete harmonic transform / P. Zubrycki, A. Petrovsky // Proc. of the 15th European Signal Process. Conf. (EUSIPCO-2007). – Poznan, 2007. – P. 2336–2340.
9. Weruaga, L. The fan-chirp transform for non-stationary harmonic signals / L. Weruaga, M. Kepesi // Signal Processing. – Vol. 87, № 6. – 2007. – P. 1504–1522.
10. Abe, T. Harmonic tracking and pitch extraction based on instantaneous frequency / T. Abe, T. Kobayashi // Acoustics, Speech, and Signal Processing (ICASSP-95). – 1995. – Vol. 1, № 9. – P. 756–759.
11. Azarov, E. Estimation of the instantaneous harmonic parameters of speech / E. Azarov, A. Petrovsky, M. Parfieniuk // Proc. of the 16th European Signal Processing Conference [Electronic resource]. – Lausanne, 2008. – CD-ROM.
12. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала // И.С. Азаров, А.А. Петровский // Речевые технологии. – 2008. – № 1. – С. 67–77.

Поступила 19.01.09

*Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: palex@bsuir.by*

E.S. Azarov, A.A. Petrovsky

SPEECH ANALYSIS BASED ON INSTANTANEOUS LINEAR PREDICTION

A new method for voiced speech analysis based on instantaneous linear prediction (LP) is proposed. The possibility of evaluating the instantaneous LP coefficients is provided by a conversion technique. It evaluates the coefficients from instantaneous harmonic parameters that can be available using harmonic modeling. Instantaneous LP approach may be very useful in various speech applications due to its high resolution both in time and frequency domain. Applications of proposed method are also examined and reported in the paper. A new method of fast LP coefficients evaluation based on Fast Fourier Transform is proposed. The paper includes corresponding experimental results.