УДК 534.4 : 621.391 : 681.513.67

С. В. Перченко, А. В. Омельченко

Харьковский национальный университет радиоэлектроники

РОБАСТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РЕЧЕВЫХ СИГНАЛОВ В РАМКАХ МОДЕЛИ ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ

Рассмотрены методы оценивания параметров речевых сигналов в рамках модели линейного предсказания. Показано, что для вокализованных речевых сигналов оценки коэффициентов формирующего фильтра по критерию минимума суммы квадратов ошибок линейного предсказания не являются наилучшими. Предложен метод оценивания параметров формирующего фильтра линейного предсказания, учитывающий импульсный характер возбуждения вокализованной речи.

модель авторегрессии, линейное предсказание, лестничный фильтр, функция потерь

Введение

К настоящему времени метод линейного предсказания (ЛП) достаточно глубоко разработан и широко применяется в задачах кодирования и распознавания речи, идентификации и верификации дикторов [3, 6]. Однако, несмотря на свою эффективность и простоту, этот метод не учитывает в полной мере особенности формирования речевых сигналов, что показано в этой статье. Сегодня, с развитием вычислительной техники стала экономически оправданной реализация модифицированных методов линейного предсказания, в большей мере учитывающих особенности речевых сигналов.

Целью исследования является разработка методов оценивания параметров модели линейного предсказания, учитывающих импульсный характер возбуждения вокализованной речи.

Методы возбуждения формирующего фильтра в современных вокодерах

Практически все современные вокодеры строятся с использованием метода линейного предсказания. При этом улучшение качества речи обеспечивается, в основном, за счет совершенствования методов возбуждения декодеров и процедур кодирования параметров. В то же время собственно метод линейного предсказания практически не претерпел изменений с момента своего появления. К настоящему времени методы возбуждения формирующего фильтра можно разделить на три больших класса: тоно-шумовое возбуждение, возбуждение от сигнала ошибки предсказания (RPE-LTP кодек, использующийся в стандарте GSM) и возбуждение специальным образом сформированной последовательностью импульсов (кодовое возбуждение, многоимпульсное возбуждение)[3].

Классический метод линейного предсказания основан на предположении, что порождающий процесс является белым шумом [5], и, соответственно, алгоритмы анализа сигнала, основанные на этом методе, стремятся свести ошибку предсказания к белому шуму с минимальной мощностью.

Однако реальный сигнал голосового возбуждения нельзя представить исключительно в виде белого шума. Вокализованная речь возбуждается периодической последовательностью импульсов, которые имеют определенную форму и, как считается, индивидуальны для каждого человека. Таким образом, актуальной задачей является разработка метода оценивания параметров модели линейного предсказания, который бы учитывал импульсный характер возбуждения речевого сигнала. Этот метод позволит сформировать остаток предсказания, более соответствующий реальному сигналу возбуждения и более адекватно оценить коэффициенты линейного предсказания.

Классический метод линейного предсказания

Метод линейного предсказания речи основывается на представлении речевого сигнала в виде процесса авторегрессии порядка p. Согласно этой модели сигнал $s[n]$ формируется путем пропускания дискретного белого шума $u[n]$ через рекурсивный формирующий фильтр порядка p (рис. 1, а). Частотная характеристика этого фильтра определяет огибающую спектра сигнала. Текущий отсчет сигнала формируется следующим образом [5]:

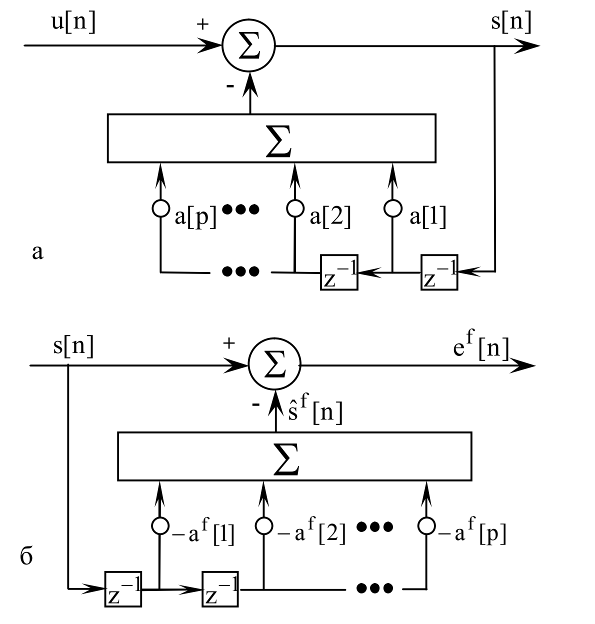

Рис. 1. Формирующий (а) и инверсный (б) фильтры

$$s[n] = -\sum_{k=1}^{p} a[k]s[n-k] + u[n], \qquad (1)$$

где $a[k]$ – коэффициенты модели авторегрессии; p – порядок модели авторегрессии; $u[n]$ – отсчеты сигнала возбуждения (белого шума). Таким образом, для формирования сигнала $s[n]$ необходимо определить коэффициенты формирующего фильтра $\{a[k]\}$ и дисперсию сигнала возбуждения σ_u^2.

В классическом методе линейного предсказания для определения параметров АР модели производится минимизация мощности ошибки линейного предсказания сигнала. Ошибка предсказания вперед вычисляется следующим образом:

$$e^f[n] = s[n] - \hat{s}^f[n], \qquad (2)$$

где $\hat{s}^f[n]$ представляет собой линейное предсказание сигнала вперед сигнала $s[n]$ (рис. 1, б) и вычисляется следующим образом:

$$\hat{s}^f[n] = -\sum_{k=1}^{p} a^f[k]s[n-k], \qquad (3)$$

где $a^f[k]$ – коэффициенты линейного предсказания вперед.

Оценивание коэффициентов формирующего АР фильтра по отсчетам сигнала $s[n]$ в классическом методе линейного предсказания производятся алгоритмами, оптимальными по критерию минимума суммы квадратов ошибок предсказания. Решение задачи минимизации энергии сигнала ошибки предсказания, приводит к системе уравнений Юла–Уолкера, решение которой относительно $\{a^f[k]\}$ позволяет получить минимально возможное значение энергии сигнала ошибки [5].

С учетом моделей формирования сигнала (1) и линейного предсказания (2), (3) можем представить z-преобразование сигнала ошибки ЛП таким образом:

$$E(z) = U(z)\left(1+\sum_{1}^{p} a^{f}[k]z^{-k}\right)\Bigg/\left(1+\sum_{1}^{p} a[k]z^{-k}\right), \quad (4)$$

где U(z) – z-преобразование сигнала возбуждения. Отношение двух полиномов в (4) можно в общем случае записать в виде z-преобразования импульсной характеристики эквивалентного фильтра [7]:

$$E(z) = U(z)\left(1+\sum_{k=1}^{\infty} h_{экв}[k]\right),$$

а сам сигнал ошибки будет представлять собой результат преобразования порождающего процесса эквивалентным фильтром $\{h_{экв}[k]\}$:

$$e^{f}[n] = u[n] + \sum_{k=1}^{\infty} h_{экв}[k]u[n-k]. \quad (5)$$

Если отсчеты порождающего процесса $u[n]$ не коррелированы, то дисперсия сигнала $e^f[n]$ выражается следующим образом:

$$\sigma_e^2 = \sigma_u^2\left(1+\sum_{k=1}^{\infty}\left(h_{экв}[k]\right)^2\right). \quad (6)$$

Из (6) видно, что $\sigma_e^2 \geq \sigma_u^2$. Критерием оптимальности в классическом методе ЛП является минимум мощности сигнала ошибки σ_e^2, что согласно (6), означает равенство нулю $h_{экв}[k]$ для всех k. С учетом (4) это возможно, когда наборы коэффициентов {a[k]} и $\{a^f[k]\}$ совпадают [7]. Таким образом, можно говорить, что критерий минимума σ_e^2 обеспечивает адекватное оценивание коэффициентов модели авторегрессии. Однако данное утверждение в общем может быть неверно для сигналов с коррелированным порождающим процессом. Это показано ниже численным моделированием (табл. 1). В этом случае более подходящим, для оценивания параметров АР модели, может оказаться критерий, отличный от минимума суммы квадратов ошибок предсказания.

Другим множеством параметров, эквивалентным коэффициентам линейного предсказания, является множество коэффициентов отражения $\{k_m\}$. Они могут быть рекурсивно вычислены по коэффициентам $\{a^f[k]\}$. Коэффициенты отражения также могут вычисляться напрямую, через отчеты сигнала:

$$f^{m}[n] = f^{m-1}[n] + k_m b^{m-1}[n-1]; \quad (7)$$

$$b^{m}[n] = b^{m-1}[n-1] + k_m f^{m-1}[n]. \quad (8)$$

Фильтр лестничной структуры, определяемый выражениями (7) и (8), показан на рис. 2. Он состоит из последовательно соединенных звеньев 1-го порядка, а выражения (7) и (8) вычисляются рекурсивно через результаты для предыдущего звена.

Коэффициенты отражения в классическом методе линейного предсказания оцениваются, подобно коэффициентам ЛП, путем совместной минимизации ошибок предсказания вперед и назад методом наименьших квадратов. Существует несколько модификаций этого метода:

– гармонический:

$$\hat{k}_m = \frac{2\sum\limits_{n=0}^{N-1} f^{m-1}[n]b^{m-1}[n-1]}{\sum\limits_{n=0}^{N-1}\left(f^{m-1}[n]\right)^2 + \sum\limits_{t=0}^{N-1}\left(b^{m-1}[n-1]\right)^2}; \quad (9)$$

– геометрический (метод Берга):

$$\hat{k}_m = \frac{\sum\limits_{n=0}^{N-1} f^{m-1}[n]b^{m-1}[n-1]}{\sqrt{\sum\limits_{n=0}^{N-1}\left(f^{m-1}[n]\right)^2 \cdot \sum\limits_{n=0}^{N-1}\left(b^{m-1}[n-1]\right)^2}}, \quad (10)$$

и другие [5].

Модифицированный метод оценивания

Как известно, в самом общем случае, все речевые сегменты можно разделить на два типа: вокализованные и невокализованные. Это приводит к двоичной, тоно-шумовой модели возбуждения: невокализованная речь возбуждается источником белого шума, а вокализованная – генератором импульсов с периодом следования, равным периоду основного тона речевого сигнала. Но, как было показано выше, метод оценивания параметров модели авторегрессии, основанный на построении соответствующего фильтра линейного предсказания с оцениванием его параметров по критерию минимума суммы квадратов ошибок предсказания, предполагает возбуждение этой модели белым шумом. При анализе же вокализованного сигнала алгоритмами, оптимальными по критерию минимума суммы квадратов ошибок линейного предсказания, энергия сигнала ошибки распределяется по всему периоду сигнала. Можно говорить о том, что алгоритмы минимизации, основанные на методе наименьших квадратов, «стремятся размазать» энергию сигнала ошибки по временному отрезку анализа как можно более равномерно с целью получения на выходе сигнала как можно более близкого к белому шуму. Но это противоречит модели импульсного возбуждения, для которой вся энергия возбуждения сосредоточена на небольших временных отрезках, соответствующих импульсам голосового возбуждения.

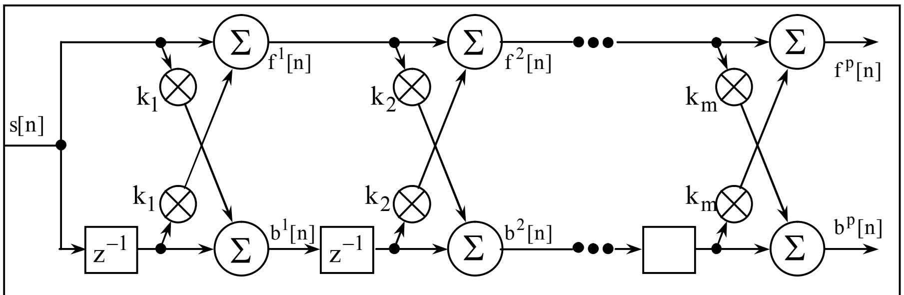

Рис. 2. Лестничная реализация фильтра линейного предсказания

Для модификации метода линейного предсказания с целью сделать его адекватным вокализованному сигналу предлагается использовать следующий комплекс мер:

1 При оценивании коэффициентов фильтра применять робастные методы оценивания и, в частности, применять функции потерь (рабочие функции минимизации) отличные от квадратичной [1, 2].

2 В качестве модели голосового тракта использовать лестничный фильтр и при этом оценивать его параметры с использованием метода адаптивного градиентного обучения [4].

Речевой сигнал может быть представлен в виде свертки сигнала возбуждения с импульсной характеристикой фильтра. Учитывая, что для вокализованной речи энергия сигнала возбуждения распределена во времени неравномерно, можно сделать вывод о том, что в процессе оценивания коэффициентов формирующего фильтра целесообразно ослабить влияние речевого сигнала на значения оценок в те промежутки времени, в течение которых на формирующий фильтр воздействует импульс возбуждения.

Как известно, целью адаптации линейной системы является нахождение вектора коэффициентов этой системы, минимизирующего некоторую функцию, которая является критерием качества адаптации системы. Для системы, использующей лестничный фильтр, адаптация производится путем совместной минимизации ошибок предсказания вперед и назад (7), (8). В этом случае критерий адаптации запишется следующим образом:

$$E_{fb}\left(\vec{k}\right)=\frac{1}{2}\left(\sum_n \rho\left(f_k^p[n]\right)+\sum_n \rho\left(b_k^p[n]\right)\right), \qquad (11)$$

где $\rho(x)$ – функция потерь.

В классическом методе наименьших квадратов критерием качества является среднеквадратическое отклонение (СКО), а соответствующей ему функцией потерь – квадратичная функция ошибки предсказания:

$$\rho(x)=x^2/2\,. \qquad (12)$$

Для анализа вокализованных речевых сигналов целесообразно использовать функции потерь, отличные от квадратичной, которые уменьшают влияние максимальных значений сигнала на результаты оценок. Предложено большое количество таких функций. Наиболее распространенными и простыми являются функция Хубера:

$$\rho(x)=\begin{cases} x^2/2, & |x|\le a;\\ a|x|-a^2/2, & |x|>a, \end{cases} \qquad (13)$$

и функция Андрюса [1]:

$$\rho(x)=\begin{cases} -a\cdot\cos(x/a)+a, & |x|\le a\pi;\\ 2a, & |x|>a\pi. \end{cases} \qquad (14)$$

В данной работе также исследовалась модифицированная функция Хубера, позволяющая полностью исключить из участия в оценивании отсчеты сигнала, превышающие пороговый уровень:

$$\rho(x)=\begin{cases} x^2/2, & |x|\le a;\\ a^2/2, & |x|>a. \end{cases} \qquad (15)$$

Для нахождения оптимальных оценок весовых коэффициентов необходимо минимизировать функционал ошибки (11). Это производится путем решения системы уравнений:

$$\frac{\partial E_{fb}\left(\vec{k}\right)}{\partial k_i}=0, \quad i=\overline{1,p}\,. \qquad (16)$$

Для лестничного фильтра, учитывая (11), эти уравнения можно записать в виде:

$$\sum_n \psi\left(f^p[n]\right)\frac{\partial f^p[n]}{\partial k_i}+\sum_n \psi\left(b^p[n]\right)\frac{\partial b^p[n]}{\partial k_i}=0\,, \quad i=\overline{1,p}\,, \qquad (17)$$

где $\psi(x)$ является производной функции потерь и называется функцией влияния. Следует заметить, что для классического МНК с квадратичной функцией потерь функция влияния вырождается в $\psi(x)=x$. Ниже приведены функции влияния, соответствующие функциям потерь Андрюса

$$\psi(x)=\begin{cases} \sin(x/a), & |x|\le a\pi;\\ 0, & |x|>a\pi \end{cases} \qquad (18)$$

и модифицированной Хубера

$$\psi(x)=\begin{cases} x, & |x|\le a;\\ 0, & |x|>a. \end{cases} \qquad (19)$$

На рис. 3 представлены рассмотренные функции потерь и соответствующие функции влияния.

Для того, чтобы эффективность оценивания коэффициентов с использованием неквадратичных функций потерь была инвариантна относительно среднеквадратического значения сигнала ошибки, аргумент функции влияния в (17) нужно масштабировать: $\psi(x/s)$, где s – робастная оценка параметра масштаба [1]:

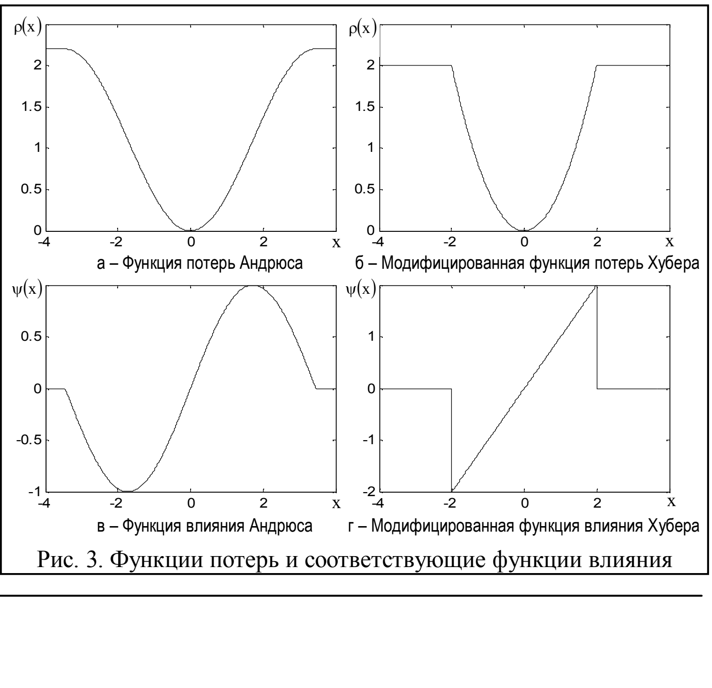

Рис. 3. Функции потерь и соответствующие функции влияния

$$s_f = \mathrm{med}\left|f^p[n]\right| / 0{,}6745;$$
$$s_b = \mathrm{med}\left|b^p[n]\right| / 0{,}6745. \quad (20)$$

С учетом того, что значения сигналов ошибок предсказания вперед и назад для модели порядка p можно рекурсивно выразить через ошибки предсказания моделей меньших порядков:

$$f^p[n] = s[n] + k_1 b^0[n-1] + \ldots + k_p b^{p-1}[n-1]; \quad (21)$$

$$b^p[n] = s[n-p] + k_1 f^0[n-p+1] + k_2 f^1[n-p+2] + \ldots$$
$$+ k_m f^{m-1}[n-p+m] + \ldots + k_p f^{p-1}[n], \quad (22)$$

можем окончательно записать (17) в следующем виде ($i = \overline{1,p}$):

$$\sum_n \psi(f^p[n]) b^{i-1}[n-1] + \sum_n \psi(b^p[n]) f^{i-1}[n-p+i] = 0. \quad (23)$$

Градиентный адаптивный лестничный метод

Определим значения вектора весовых коэффициентов формирующего фильтра градиентным методом.

Если на конкретном образце сигнала этот метод сходится, то он сходится достаточно быстро [4]. Если же в каком-либо конкретном случае процедура градиентного поиска не сходится, то коэффициенты можно оценивать одним из стандартных методов, например в соответствии с выражениями (9) и (10).

Для рассматриваемого случая с инверсным лестничным фильтром градиентом является набор частных производных (16) по всем коэффициентам фильтра [4]:

$$\nabla = \left[\frac{\partial E_{fb}(\vec{k})}{\partial k_1}; \quad \frac{\partial E_{fb}(\vec{k})}{\partial k_2}; \ldots \quad \frac{\partial E_{fb}(\vec{k})}{\partial k_p} \right]. \quad (24)$$

Каждая компонента градиента определяет степень наклона рабочей функции по оси соответствующего коэффициента, а сам градиент определяет направление на минимум функции потерь. В общем случае можно начинать процесс с произвольного начального значения $\vec{k}_0$ из области допустимых значений, но в нашем случае целесообразно начинать со значений, определенных стандартными методами (гармоническим или методом Берга), что обеспечит более быструю сходимость и меньшую вероятность попадания в локальный минимум функции потерь. Итеративный процесс градиентного поиска можно представить в следующем виде:

$$\vec{k}_{j+1} = \vec{k}_j + \mu(-\nabla_j), \quad (25)$$

где j – номер итерации; $\vec{k}_j$ – текущее значение вектора коэффициентов; $\vec{k}_{j+1}$ – новое значение. Параметр μ определяет устойчивость и скорость сходимости алгоритма. В принципе он может быть как константой, так и зависеть от номера итерации. Необходимо отметить, что выбор наиболее подходящих значений μ сопряжен со значительными трудностями. Если предположить, что входной сигнал является гауссовским процессом с нулевым средним значением, то сходимость процесса адаптации будет гарантироваться до тех пор, пока μ будет иметь положительное значение, лежащее в интервале от 0 до $1/\lambda_{max}$, где λ_{max} – максимальное собственное значение автокорреляционной матрицы сигнала [4, 5]. В отношении же вокализованного речевого сигнала такое предположение не верно. Поэтому в разработанном алгоритме приемлемое значение μ выбирается адаптивно.

Разработка алгоритма и испытание на моделях речевых сигналов

На основе вышеизложенных теоретических предпосылок разработан алгоритм помехоустойчивого оценивания параметров АР модели. Он состоит из следующих шагов:

1. Для входного сигнала производится оценивание параметров линейного предсказания классическим геометрическим методом (10), и фильтрация производится в соответствии с выражениями (7), (8).
2. По полученной в п.1 ошибке предсказания вычисляется параметр масштаба s аргумента функции влияния. Для функций Хубера и Андрюса он вычисляется в соответствии с (20), для модифицированной функции Хубера он вычисляется как доля от максимального значения сигнала ошибки. Наиболее подходящие значения параметра масштаба для модифицированной функции Хубера лежат в пределах 0,6…0,75.
3. Выбирается начальное значение параметра скорости адаптации $\mu = 0{,}05$ и начинается процедура градиентного поиска согласно (25). В качестве начальных значений коэффициентов используются значения, полученные в п.1.
4. Производится фильтрация текущего отсчета сигнала с использованием текущих параметров фильтра

$$f^m[n] = f^{m-1}[n] + k_m[n-1] b^{m-1}[n-1];$$
$$b^m[n] = b^{m-1}[n-1] + k_m[n-1] f^{m-1}[n]$$

и находится очередные отсчет ошибки предсказания.

5. На основе полученных отсчетов сигналов ошибок производится адаптация всех коэффициентов фильтра:

$$k_i[n] = k_i[n-1] - \mu\left(\psi\left(f^p[n]\right) b^{i-1}[n-1] + \right.$$
$$\left. + \psi\left(b^p[n]\right) f^{i-1}[n-p+i]\right),$$
$$i = \overline{1,p}.$$

6. Если значения какого-либо из вновь полученных коэффициентов отражения выходят за пределы допустимых значений, то параметр μ уменьшается и процедура градиентного поиска начинается заново.

Для проверки разработанного алгоритма использовались модельные сигналы, сформированные с использование рекурсивных фильтров 2-го порядка, которые возбуждались короткими прямоугольными импульсами результаты исследования приведены в табл. 1 и на рис. 4. Приведенные результаты получены после двух проходов адаптации на 3-х периодах сигнала. Из табл. 1 видно, что оценки коэффициентов отражения сходятся к истинным значениям. Из рис. 4, е также видно, что сигнал остатка предсказания, найденный на основе предложенного метода, значительно ближе к истинному сигналу возбуждения, чем остаток предсказания для метода наименьших квадратов.

Таблица 1

Динамика изменений коэффициентов отражения в процессе адаптации

Коэффициенты	Истинное значение	Метод Берга	Метод робастного градиентного поиска					
			Итерация 1			Итерация 2		
			Период 1	Период 2	Период 3	Период 1	Период 2	Период 3
k1	–0,8	–0,9088	–0,8731	–0,8170	–0,8030	–0,8030	–0,8029	–0,8028
k2	0,8	0,9125	0,8826	0,8458	0,8257	0,8237	0,8220	0,8204

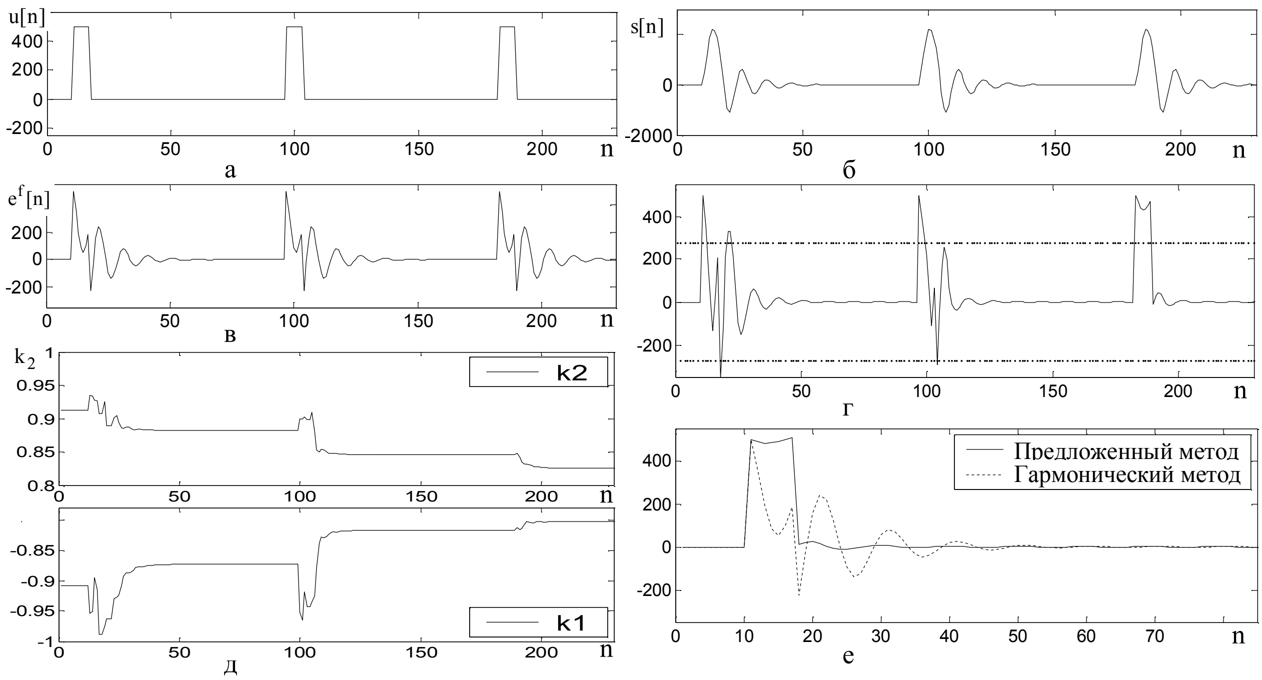

Рис. 4. Сигналы, полученные в ходе исследований для модели 2-го порядка:
а – сигнал возбуждения; б – модельный сигнал 2-го порядка; в – оценка сигнала возбуждения, полученная методом Берга; г – сигнал ошибки лестничного фильтра в процессе адаптации; д – траектории адаптации коэффициентов фильтра на первой итерации; е – ошибки линейного предсказания, полученные гармоническим и предложенным методом.

Заключение

В настоящее время повышение качества речи, передаваемой с помощью вокодеров, является одной из важнейших проблем в области телекоммуникаций. Однако вопросам адаптации вокодеров к особенностям формирования вокализованных речевых сигналов в настоящее время уделяется недостаточное внимание.

В работе показано, что критерий минимума суммы квадратов ошибок предсказания не позволяет наилучшим образом оценить коэффициенты формирующего фильтра, а также определить форму импульсов возбуждения. Предложен модифицированный метод оценивания параметров модели линейного предсказания, учитывающий импульсный характер возбуждения вокализованной речи. Разработан алгоритм оценивания коэффициентов отражения на основе метода адаптивного градиентного поиска и использованием не квадратичных функций потерь. Исследования показали, что предложенный метод перспективен для анализа сигналов с импульсным возбуждением. Дальнейшим направлением исследований является распространение разработанного метода для обработки речевых сигналов, в частности для кодирования речевых сигналов.

Список литературы

1. Хогг Р.В. Введение в помехоустойчивое оценивание // Устойчивые статистические методы оценки данных: Под ред. Н.Г. Волкова. – М.: Машиностроение, 1984. – С. 12-26.

2. Мартин Р.Д. Устойчивый авторегрессионный анализ временных рядов // Устойчивые статистические методы оценки данных: Под ред. Н.Г. Волкова. – М.: Машиностроение, 1984. – С. 121-145.

3. Коротаев Г. А. Анализ и синтез речевого сигнала методом линейного предсказания // Зарубежная радиоэлектроника. – 1990. – № 3. – С 31-51.

4. Уидроу Б., Стринз С. Адаптивная обработка сигналов: Пер. с англ. – М.: Радио и связь, 1989. – 440 с.

5. Марпл. – мл. С.Л. Цифровой спектральный анализ и его приложения: Пер с англ. – М.: Мир, 1990. – 584 с.

6. Маркел Дж.Д., Грэй А.Х. Линейное предсказание речи: Пер с англ. / Под ред. Ю.Н. Прохорова и В.С. Звездина. – М.: Связь, 1980. – 308 с.

7. Сергиенко А.Б. Цифровая обработка сигналов: 2-е изд. – С.-Пб.: Питер, 2006. – 751 с.

Поступила в редколлегию 18.06.2007

Рецензент: д-р техн. наук, проф. М.В. Безрук, Харьковский национальный университет радиоэлектроники, Харьков.