

МЕТОД ПОВЫШЕНИЯ КАЧЕСТВА ЗВУЧАНИЯ СИНТЕЗИРОВАННОГО РЕЧЕВОГО СИГНАЛА В ЦИФРОВОМ ВОКАДЕРЕ С ПРЕДСКАЗАНИЕМ

С.Ф.Лихачев, М.В.Назаров, Ю.Н.Прохоров

Московский электротехнический институт связи

Р е ф е р а т

Предложен новый метод оценивания параметров речевых сигналов на основе метода обновляющего процесса (ОП). По сравнению с традиционным методом наименьших квадратов (МНК), разработанный метод позволяет повысить качество звучания синтезированного речевого сигнала в цифровом вокадере, работающем при действии шумов умеренной интенсивности.

1. Введение

Предметом настоящего доклада является исследование и разработка методов повышения эффективности рекуррентных алгоритмов оценивания параметров, обеспечивающих повышение качества цифровой передачи речи при низких скоростях 2400...4800 бит/с и наличии входных шумов умеренной интенсивности.

В настоящее время наиболее перспективными в этом направлении являются системы передачи с предсказанием [1,2,3].

Известно, что в вокадерах с линейным предсказанием не учитываются особенности слухового восприятия в частотной области.

В работах [1,2,4,5] и других доказано,

что область низких частот наиболее важна для слуха. В тоже время методы линейного предсказания, построенные в рамках метода наименьших квадратов, обеспечивают малую ошибку в описании высокочастотной области спектра, к которой ухо менее чувствительно. Исходя из этого, отличие спектров исходного и синтезированного речевых сигналов оказывается довольно значительным. Низкое качество восстановленного сигнала в низкоскоростных системах с линейным предсказанием обусловлено тем, что погрешность предсказания содержит информацию о спектре, которую не может извлечь анализатор системы передачи, построенный на методе наименьших квадратов. Поэтому большой научный и технический интерес представляет разработка метода такого изменения спектрального состава восстановленного речевого сигнала, при котором учитываются особенности слухового восприятия.

2. Метод оценивания параметров предсказания.

Введем авторегрессионную модель сигнала

$$x_t = \vec{a}^T \vec{\varphi}(x_{t-1}) + \xi_t, \quad (1)$$

где $\vec{a}^T = (a^{(1)}, \dots, a^{(p)})$ - вектор параметров авторегрессии; $\vec{\varphi}(x_{t-1})$ - функция регрессии; ξ_t - порождающий процесс; p - размерность модели.

Если сигнал и модель стохастически

эквивалентны, то обновляющий процесс v_t оказывается последовательностью некоррелированных случайных величин, обладающих теми же характеристиками, что и процесс ξ_t . Поэтому синтез сигнала на приемной стороне системы передачи можно представить как прохождение обновляющего процесса v_t через линейное звено с передаточной функцией $K(\omega)$. Отклонение формантных максимумов в спектре речевого сигнала заметно на слух, если оно превышает +1дБ. Таким образом, из физических соображений следует, что показатель качества, отражающий спектральные свойства синтезированного аналогового речевого сигнала, должен представлять собой меру отклонения спектров исходного ($G_{xx}(\omega)$) и синтезированного ($\hat{G}_{xx}(\omega)$) речевых сигналов. Такую меру можно представить в виде:

$$\rho = \int | [G_{xx}(\omega) - K(\omega) \cdot G_{vv}(\omega)] \cdot M(\omega) | d\omega, \quad (2)$$

где $M(\omega)$ - функция веса.

Функция потерь (2) не позволяет получить аналитически простые алгоритмы оценивания параметров речевого сигнала.

Пусть $M=0$, если $G_{xx}(\omega) - K(\omega) \cdot G_{vv}(\omega) < 0$. Тогда можно записать

$$\rho = \int | [G_{xx}(\omega) - K(\omega) \cdot G_{vv}(\omega)] \cdot M(\omega) \cdot d\omega |. \quad (3)$$

Используя преобразование Винера-Хинчина меру ρ (3) теперь можно представить в виде:

$$\rho = \int | [B_{xx}(\tau) - q(\tau) B_{vv}(\tau)] \cdot M(\tau) | d\tau, \quad (4)$$

где $M(\tau)$ и $M(\omega)$ связаны преобразованием Фурье; $q(\tau) = \frac{K(\omega)}{M(\omega)}$ и $B(\tau)$ - преобразование Фурье от $K(\omega) \cdot M(\omega)$.

Заменяя интеграл интегральной суммой и подставляя в нее вместо $B_{xx}(\tau)$ ее оценку, получим показатель качества в дискретном времени:

$$\gamma_N(\vec{a}) = \left| \sum_{\tau=0}^N w(\tau) \sum_{t=1}^N [x_t - \vec{a}^T \vec{\varphi}(x_{t-1})] [x_{t+\tau} - \vec{a}^T \vec{\varphi}(x_{t+\tau-1})] \right| \quad (5)$$

где $w(\tau)$ - весовая последовательность.

Задача оценивания параметров может быть сформулирована следующим образом: по наблюдаемой последовательности x_t или $z_t = x_t + v_t$, $t=1,2,\dots,N$ и априори заданной модели сигнала (1) определить наилучшую \vec{m}_N из условия:

$$\vec{m}_N = \operatorname{argmin}_{\vec{a}} \gamma_N(\vec{a}), \quad (6)$$

где v_t - шумовая последовательность; \vec{m}_N - оценка вектора \vec{a}

Метод отыскания оценок параметров модели авторегрессии минимизацией целевой функции $\gamma_N(\vec{a})$ является развитием метода обновляющего процесса [1].

Из (5) можно получить оптимальную в смысле (6) оценку

$$\vec{m}_N = \left[\sum_{\tau=0}^N w(\tau) \sum_{t=1}^N \vec{\varphi}(x_{t-1}) \cdot \vec{\varphi}^T(x_{t+\tau-1}) \right]^{-1} \sum_{\tau=0}^N w(\tau) \sum_{t=1}^N x_{t+\tau} \vec{\varphi}(x_t, \lambda \tau)$$

Используя лемму об обращении матриц получим рекуррентные выражения для оценок:

$$\vec{m}_t = \vec{m}_{t-1} + \gamma_{t-1}^{-1} \vec{\varphi}(x_{t-1}) \sum_{\tau=0}^N w(\tau) [x_{t+\tau} - \vec{m}_{t-1}^T \vec{\varphi}(x_{t+\tau-1})];$$

$$\gamma_t = \gamma_{t-1} - \gamma_{t-1} \vec{\varphi}(x_{t-1}) \left[1 + \sum_{\tau=0}^N w(\tau) \cdot \vec{\varphi}(x_{t+\tau-1}) \vec{\varphi}^T(x_{t-1}) \right]^{-1} \gamma_{t-1} \vec{\varphi}(x_{t-1})$$

с начальными условиями $\vec{m}_0 = E \vec{a}$; $\gamma_0 = \epsilon^{-1} I$, $\epsilon \rightarrow 0$, $\forall t \in \mathbb{N}$, γ_t ($p \times p$)

При стохастической эквивалентности сигнала и модели оценка (7), (8) совпадает при $N \rightarrow \infty$ с асимптотической оценкой МНК, но в отличие от нее в линейном случае оказывается несмещенной при стационарных шумах с равномерным спектром, так как, например, при $\rho = 1$, $w(\tau) = \delta_{\tau \tau_0}$, имеем из (7):

$$\lim_{N \rightarrow \infty} m_N = \frac{B_{xx}(\tau_0+1) + B_{xx}(\tau_0+1)}{B_{xx}(\tau_0) + B_{vv}(\tau_0)} = \frac{B_{xx}(\tau_0+1)}{B_{xx}(\tau_0)}. \quad (9)$$

Свойство асимптотической несмещенности сохраняется для линейной модели авторегрессии при любом τ_0 , но может нарушаться в нелинейном случае.

Сложность технической реализации алгоритма (8) обусловлена необходимостью вычисления матрицы γ_t . Для упро-

щения вычислений был предложен приближенный алгоритм оценивания:

$$\vec{m}_t = \vec{m}_{t-1} + \vec{P}_t^{-1} \sum_{\tau=0}^t \omega(\tau) \vec{\varphi}(x_{t-1}) [x_{t-1} - \vec{\varphi}(x_{t-1}) \vec{m}_{t-1}] \quad (10)$$

где $\vec{P}_t = \sigma_0^2 / t$ - матрица коэффициентов. Можно показать, что $\lim_{t \rightarrow \infty} \vec{m}_t = \vec{a}$.

3. Экспериментальное исследование.

Экспериментальная проверка алгоритма оценивания параметров проводилась для линейной и нелинейной моделей предсказания. В случае линейной модели показано, что разработанные алгоритмы по сравнению с алгоритмами МНК обеспечивают меньшее смещение допредельных оценок параметров в шумах с равномерным спектром. На рис.1 показано смещение оценки \vec{m}_t параметра \vec{a} в шумах (при истинном значении $\vec{a} = -0,8$) для линейной модели авторегрессии первого порядка, при различных отношениях сигнал-шум.

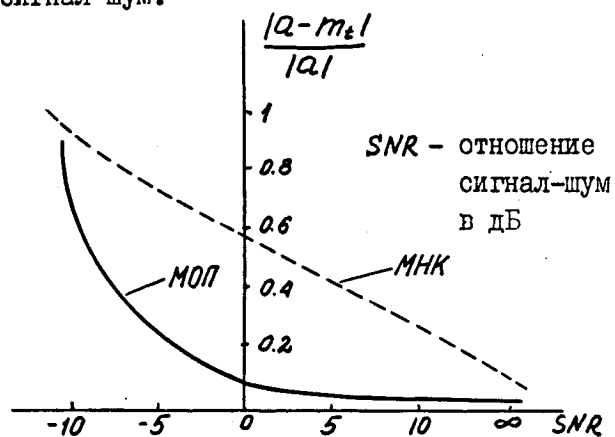


Рис.1 Смещение оценок параметров.

Из рисунка видно, что при $SNR = 5 \dots 10$ дБ, смещение оценок в 2...3 раза меньше, чем при МНК.

В качестве нелинейной модели рассмотрена модель предсказания, в которой регрессии $\vec{\varphi}(x_{t-1})$ представлена в виде ряда по функциям Уолша:

$$\vec{\varphi}(x_{t-1}) = \sum_{k=1}^m \sum_{i=0}^{N-1} a_k^{(i)} \cdot wal^{(i)}(x_{t-1})$$

где $\{a_k^{(i)}\}, k=1, \dots, m; i=0, \dots, N-1$ - параметры нелинейного предсказания; $wal^{(i)}(x)$ - функция Уолша i -го порядка. Такое представление позволяет учесть негауссовское распределение вероятностей сигнала.

Алгоритмы оценивания параметров в случае нелинейной модели обеспечивают "обеление" погрешности предсказания. В частности, в ней подавляются импульсы основного тона. На рис.2 представлена корреляционная функция погрешности предсказания речевого сигнала получения по МНК МОР. Отрезком показан 95%-ый доверительный интервал; T_{0T} - период основного тона.

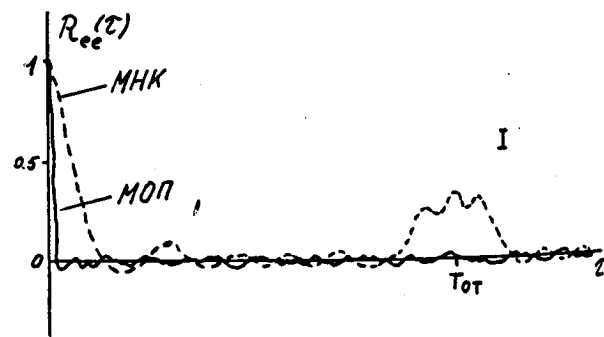


Рис.2 Корреляционные функции ошибки.

4. Разработка вокодера.

На основе алгоритма (8) на ЭВМ проведено моделирование цифрового вокодера с улучшенным качеством звучания восстановленного речевого сигнала в акустических шумах умеренной интенсивности $SNR = (+5 \dots +10)$ дБ. Скорость передачи 2400 бит/с.

В качестве исходного материала был использован речевой сигнал с полосой частот до 4,7 кГц при частотах дискретизации 8; 16 кГц соответственно и числе уровней квантования 2^{12} .

Блок-схема передающей части (анализатора) вокодера представлена на рис.3.

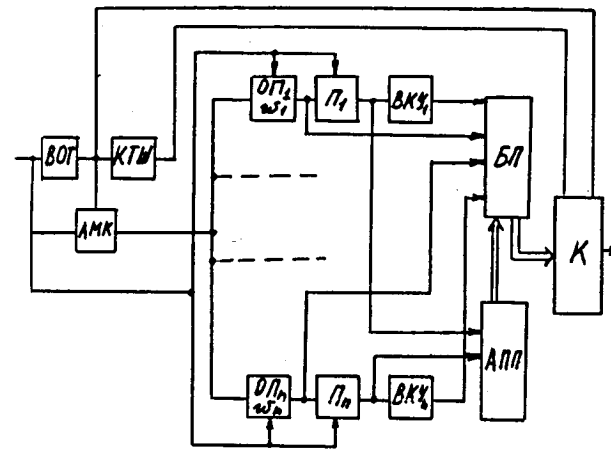


Рис.3 Блок-схема анализатора вокодера.

Передатчик состоит из:

блоков предварительной обработки (выделитель основного тона (БОТ), классификатор тон-шум (КТШ), анализатор максимума корреляции (АМК), осуществляющих оценку периода основного тона, признака вокализованности, а также поиска точки взвешивания оценки корреляционной функции текущего сегмента сигнала;

пяти ветвей анализа (блок оценивания параметров ОП), блок предсказания (П), блок вычисления коэффициентов усиления (ВКУ) с различными значениями функции веса $\omega(\tau)$;

блока памяти, в который записываются реализации погрешности предсказания и коэффициенты усиления;

блока анализа погрешности предсказания (АПП), который осуществляет выбор номера ветви по минимальному расстоянию между оценкой функции корреляции погрешности предсказания для данной ветви и функцией корреляции порождающего процесса;

блока квантования (К).

Экспериментальные исследования вокодера показали следующее:

1) При $SNR = 5 \dots 10$ дБ разборчивость слов равна 97%.

2) Улучшение качества синтезирован-

ного сигнала по сравнению с МНК достигается за счет уточнения спектров сигнала на переходных участках речи, которые плохо воспроизводятся в традиционных вокодерах. Близость спектров или корреляционных функций исходного и синтезированного сигналов в разработанном вокодере улучшается на 15...20%.

В таблице I приведены количественные соотношения для квадратичных отклонений функций исходного и синтезированного сигналов при различных SNR . Ошибка ϵ_R^2 равна нормированному квадрату нормы разности корреляционных функций исходного и синтезированного сигналов на периоде основного тона.

Таблица I.

SNR (дБ)	∞	12	8
ϵ_R^2 (МНК)	$0,41 \pm 0,04$	$0,45 \pm 0,03$	$0,53 \pm 0,01$
ϵ_R^2 (МОР)	$0,307 \pm 0,03$	$0,37 \pm 0,04$	$0,48 \pm 0,01$

3) Методом парных сравнений установлено, что число действительных суждений, высказанных аудитором в пользу разработанного вокодера составило 80...85%.

Л и т е р а т у р а

1. Прохоров Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов. - М.: Радио и связь, 1984. - 240с.
2. Лихачев С.Ф. Нелинейное предсказание речевых сигналов. - Материалы Всесоюз. семинара APCO-12. - Киев, 1982, с.112-114.
3. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. - М.: Радио и связь, 1985. - 176с.
4. Jain V.K. Speech signal analysis by error-weighted LPC-GLOBESOM'82: IEEE Global Telecommun. conf. Miami Beach, Fla, 29 Nov-2 Dec, 1982, Conf.Rec.Vol 3,

New York, 1982, pp 1321-1324.

5. Un C.K., Lee J.R. On spectral flattening techniques in residual - excited linear prediction vocoding - ICASSP' 82, Proc. IEEE; Int. Conf. Acoust. Speech and Signal Proc, Paris, May 3-5, 1982, NY, pp 216-219.