ные результаты, как только импульс основного тона появится внутри интервала анализа. Поэтому в большинстве практических систем, в которых нельзя организовать анализ синхронно с основным тоном, используются интервалы того же порядка, что и в корреляционном методе. В последующих параграфах будут представлены результаты экспериментальной оценки влияния длины и положения интервала анализа на погрешность предсказания для ковариационного и автокорреляционного методов[1]. Однако сначала коротко рассмотрим свойства погрешности предсказания и нормированной погрешности, получаемой на ее основе.

8.5. Погрешность предсказания

В результате анализа сигнала с помощью линейного предсказания возникает погрешность предсказания, определяемая как (8.97):

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) = Gu(n). \qquad (8.97)$$

Если речевой сигнал действительно порождается моделью линейного предсказания порядка p с переменными во времени параметрами, то $e(n)$ должна представлять собой хорошее приближение для сигнала на выходе источника возбуждения. Основываясь на этом обстоятельстве, можно ожидать, что для вокализо-

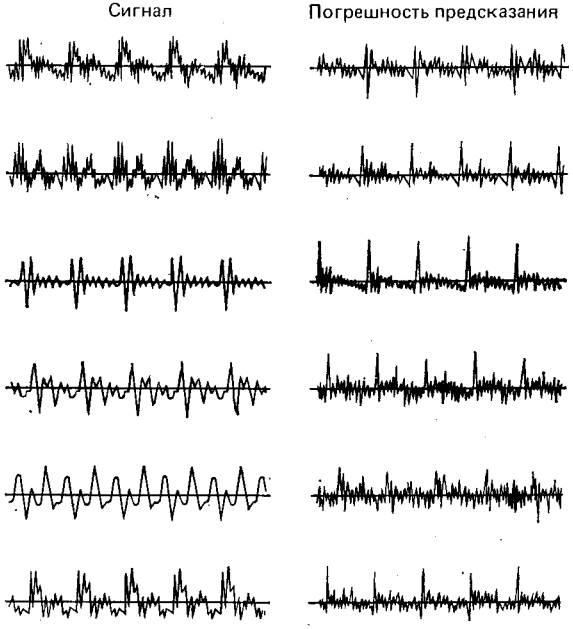


Рис. 8.5. Примеры сигналов и погрешностей предсказания для гласных (*i*, *e*, *a*, *o*, *u*, *y*) [14]

[1] Исследование Рабинера и др. [16] показали, что выбор параметров для лестничного и ковариационного методов совпадают. Поэтому в данном параграфе различия между ними не делается.

ванных сегментов речевого сигнала погрешность предсказания должна быть большой в начале каждого периода основного тона. Таким образом, период основного тона можно определить, оценивая координаты достаточно больших отсчетов и определяя период как разность координат во времени двух соседних отсчетов погрешности, превысивших соответствующий порог. С другой стороны, период основного тона можно определить на основе корреляционного анализа погрешности предсказания путем обнаружения максимального пика в подходящем диапазоне задержек. Другим объяснением того, что погрешность предсказания удобна для оценивания периода основного тона, является тот факт, что спектральная плотность погрешности практически равномерна во всей полосе частот, что обусловлено устранением из нее формант.

Чтобы показать особенности сигнала погрешности предсказания, на рис. 8.5 представлен ряд фрагментов гласных звуков соответствующих фрагментов погрешности (Страубе [14]). Для всех этих фрагментов гласных звуков в сигнале погрешности явно видны импульсы на интервалах, соответствующих периоду основного тона. На рис. 8.6—8.9 представлен ряд других экспериментов с погрешностью предсказания. На всех рисунках в части *а*) показан сегмент обрабатываемого речевого сигнала, в части *б*) — погрешность предсказания, в части

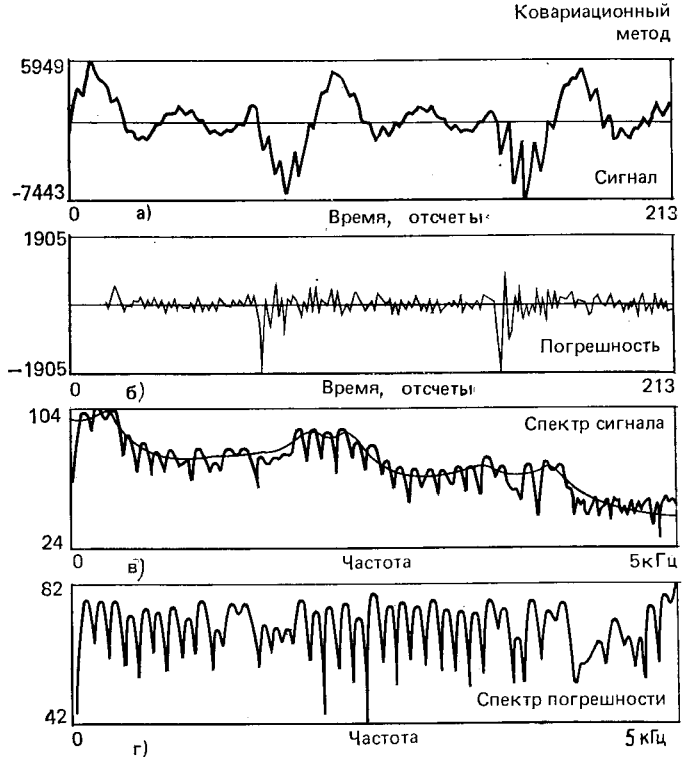


Рис. 8.6. Типичные сигналы и спектры ковариационного метода линейного предсказания для мужского голоса [16]

в) — логарифм дискретного преобразования Фурье сигнала из *а*) (вычисленный с использованием БПФ) совместно с логарифмом $H(e^{i\omega T})$ в качестве огибающей, в части *д*) представлен логарифм спектральной плотности погрешности предсказания (рассчитанный на основе БПФ). Рисунки 8.6 и 8.7 содержат результаты обработки гласного звука *i* (как в слове we), произнесенного муж-

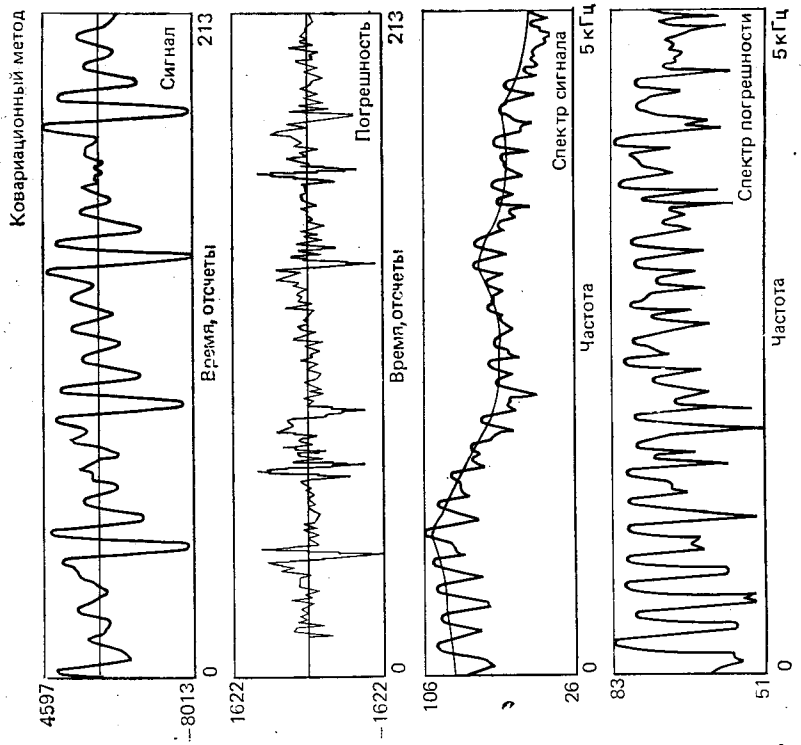


Рис. 8.8. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для женского голоса [16]

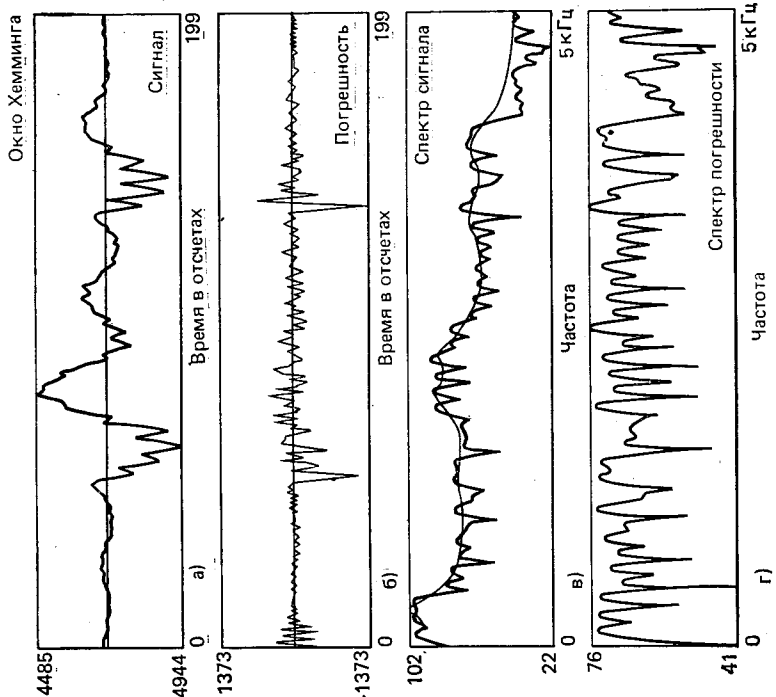


Рис. 8.7. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для мужского голоса [16]

ским голосом, с использованием ковариационного и автокорреляционного методов соответственно (с окном Хемминга). Продолжительность интервала анализа составляла 20 мс. Легко видеть, что в погрешности предсказания имеются пики точно в начале каждого периода основного тона, а спектральная плотность достаточно равномерна, хотя в ней просматривается линейчатая структура, возникающая за счет основного тона. Отметим слишком большую погрешность предсказания в начале сегмента на рис. 8.7 при использовании автокорреляционного

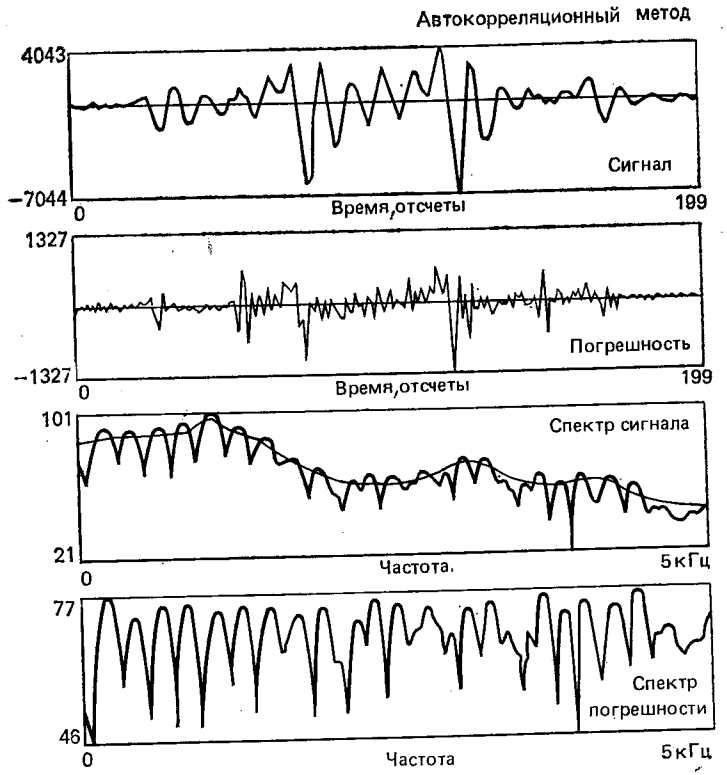


Рис. 8.9. Типичные сигналы и спектры автокорреляционного метода линейного предсказания для женского голоса [16]

метода. Это, конечно, обусловлено попыткой предсказать отсчеты сигнала по нулевым значениям вне интервала $0 \leqslant m \leqslant 199$. Скорость убывания окна Хемминга в данном случае оказалась недостаточной для эффективного устранения этой ошибки.

На рис. 8.8 и 8.9 показаны аналогичные результаты, полученные на гласном *a* (как в слове father), произнесенном женским голосом. У этого диктора в интервал анализа попадает около пяти периодов основного тона. Таким образом, на рис. 8.8 в сигнале погрешности наблюдается большое количество острых пиков в начале каждого периода основного тона при ковариационном методе анализа. Однако использование окна Хемминга в автокорреляционном методе привело к тому, что пики в погрешности предсказания уменьшились вследствие убывания функции окна к концу интервала.

Поведение сигналов погрешности предсказания, представленного на предыдущих рисунках, позволяет рассчитывать, что они являются как раз теми сигналами, по которым наиболее просто оценить период основного тона. В [5] пока-

зано, что для звуков, не имеющих явно выраженной гармонической структуры, например плавных, как r, l, или носовых, как m, n, пики в погрешности предсказания выражены не столь отчетливо. Кроме того, на переходах между вокализованными и невокализованными звуками выбросы за счет основного тона в погрешности часто просто не проявляются.

Короче говоря, хотя сигнал погрешности предсказания $e(n)$ кажется очень подходящим для построения на его основе детектора основного тона, однако имеется ряд специфических трудностей в определении положения импульса для широкого класса гласных звуков и, таким образом, сигнал погрешности предсказания не следует считать чем-то исключительным при решении указанной задачи. В 8.10.1 рассмотрен один метод определения основного тона по погрешности предсказания.

8.5.1. Другие выражения для нормированного среднего квадрата погрешности предсказания

Нормированная погрешность предсказания для автокорреляционного метода определяется как

$$V_n = \frac{\sum_{m=0}^{N+p-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)}, \qquad (8.98а)$$

где $e_n(m)$ — погрешность, соответствующая речевому сегменту $s_n(m)$ для момента n. Для ковариационного метода соответствующее определение имеет вид

$$V_n = \frac{\sum_{m=0}^{N-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)}. \qquad (8.98б)$$

Полагая, что $\alpha_0 = -1$, погрешность предсказания можно записать в виде

$$e_n(m) = -\sum_{k=0}^{p} \alpha_k s_n(m-k). \qquad (8.99)$$

Подставляя (8.99) в (8.98) и используя (8.13), получим, что

$$V_n = \sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i \frac{\varphi_n(i,j)}{\varphi_n(0,0)} \alpha_j, \qquad (8.100а)$$

и подставляя уравнение (8.14) в (8.100), получим

$$V_n = -\sum_{i=0}^{p} \alpha_i \frac{\varphi_n(0,i)}{\varphi_n(0,0)}. \qquad (8.100б)$$

Еще одно выражение для V_n получено в алгоритме Дарбина, т. е.

$$V_n = \prod_{i=1}^{p} (1-k_i^2). \qquad (8.101)$$

Не все полученные выше выражения эквивалентны между собой. Они требуют дополнительного анализа с учетом используемого метода. Например, (8.101), основанное на алгоритме Дарбина, справедливо лишь для автокорреляционного и лестничного методов. Аналогично, поскольку лестничный метод не использует вычисления автокорреляционной функции в явном виде, выражения (8.100) в данном случае непосредственно неприменимы. В табл. 8.2 объединены все перечисленные выше выражения для нормированного среднего квадрата погрешности и показана область применения каждого из выражения. (Для упрощения таблицы индексы n и p опущены.)

8.5.2. Экспериментальное определение погрешности предсказания

Для получения рекомендаций по выбору параметров p и N при практическом использовании алгоритма линейного предсказания Чандра и Лин [15] провели серию исследований. Они измеряли нормированную среднюю квадратическую погрешность предсказания для предсказателя порядка p при различных параметрах алгоритма для следующих случаев: ковариационный и автокор-

Таблица 8.2

Выражения для нормированной погрешности

Выражение	Ковариационный метод	Автокорреляционный метод	Лестничный метод
$V = \frac{\sum_m e^2(m)}{\sum_m s^2(m)}$	Справедливо	Справедливо *	Справедливо
$V = \sum_i \sum_j \alpha_i \frac{\varphi(i,j)}{\varphi(0,0)} \alpha_j$	Справедливо	Справедливо **	Несправедливо
$V = \sum_i \alpha_i \frac{\varphi(i,j)}{\varphi(0,0)}$	Справедливо	Справедливо **	Несправедливо
$V = \prod_i (1-k_i^2)$	Несправедливо	Справедливо	Справедливо

* Это выражение вычисляется по взвешенному сигналу при верхнем пределе $N-1+p$.
** В этом случае $\varphi(i,j) = R(i-j)$.

реляционный методы; синтетические гласные и натуральная речь; синхронный с основным тоном и асинхронный анализ. Нормированная погрешность определялась в соответствии с табл. 8.2. На рис. 8.10—8.15 показаны результаты, полученные Чандрой и Лином [15].

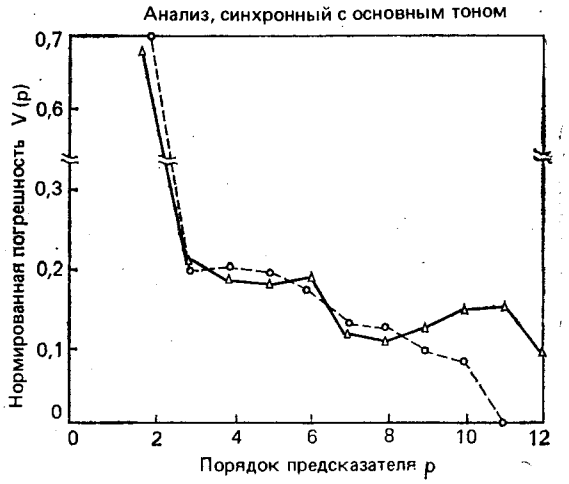


Рис. 8.10. Зависимость погрешности предсказания от порядка предсказателя для вокализованного сегмента синтетического гласного звука при анализе, синхронизированном сигналом основного тона [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

На рис. 8.10 показана нормированная дисперсия V при различном порядке модели для сегмента синтетического звука $|i|$ (в слове «heed») при периоде основного тона, равном 83 отсчетам. Интервал анализа составлял 60 отсчетов и начинался в начале периода основного тона, т. е. результаты соответствуют синхронному анализу. Для ковариационного метода погрешность

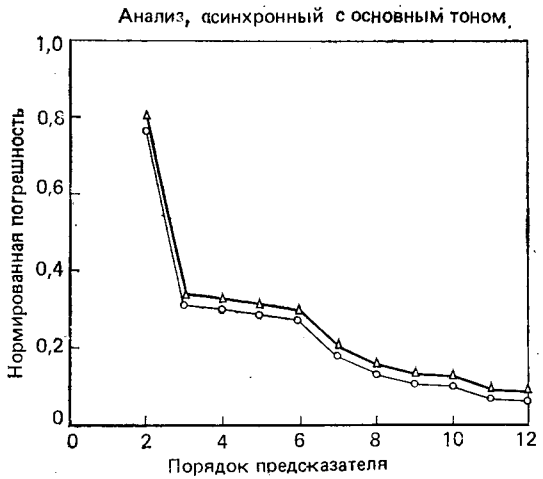


Рис. 8.11. Зависимость погрешности предсказания $V(p)$ от порядка предсказателя p при анализе, синхронизированном сигналом основного тона [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

предсказания монотонно убывает при увеличении порядка модели от 0 до 11, т. е. до порядка модели, используемой при синтезе данного звука. Для автокорреляционного метода погрешность остается на уровне 0,1 при больших p [7]. Это объясняется тем об-

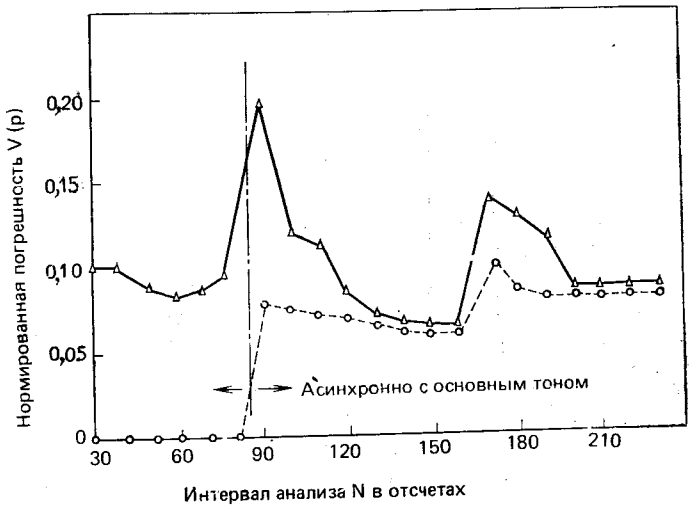


Рис. 8.12. Погрешность предсказания как функция длины интервала анализа N для вокализованного сегмента синтетического речевого сигнала [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

стоятельством, что для случая автокорреляционного анализа при малой протяженности интервала анализа погрешность предсказания в начале сегмента составляет значительную часть общего

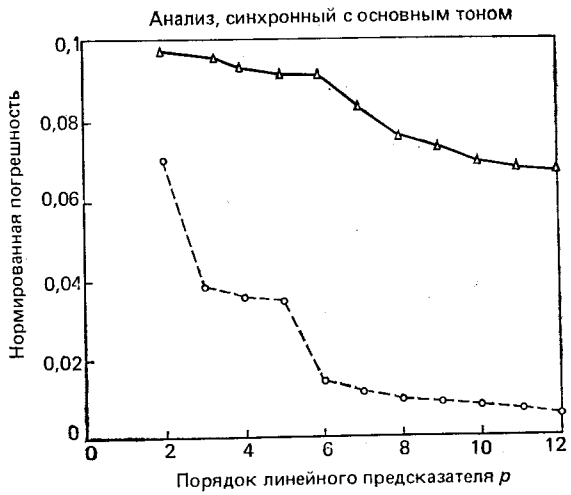


Рис. 8.13. Погрешность предсказания $V(p)$ как функция порядка предсказателя p для вокализованного сегмента натурального речевого сигнала при анализе, синхронизированном основным тоном [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

среднего квадрата погрешности. Этого, конечно, не происходит при ковариационном методе, где для предсказания используются отсчеты вне интервала, на котором ведется предсказание.

На рис. 8.11 показаны результаты анализа с асинхронной обработкой того же сегмента, что и на рис. 8.10. В данном случае

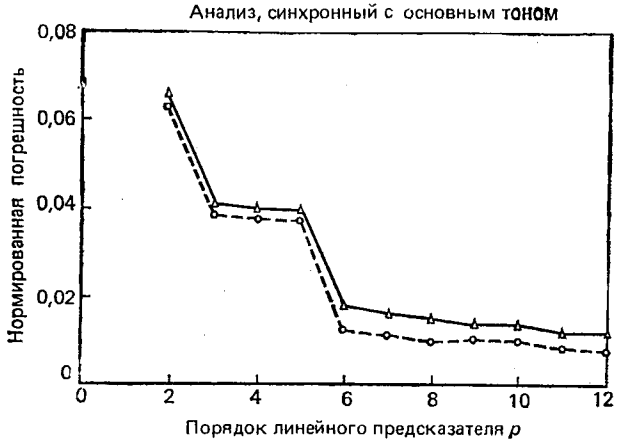


Рис. 8.14. Погрешность предсказания $V(p)$ как функция порядка предсказателя для вокализованного сегмента гласного звука при асинхронном анализе [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

длительность окна составляла 120 отсчетов. При этом ковариационный и автокорреляционный методы дают примерно одинаковые результаты при различных значениях p. Более того, значе-

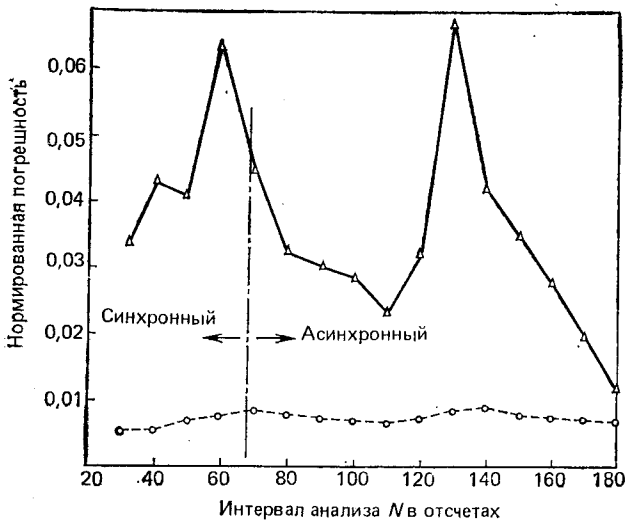


Рис. 8.15. Погрешность предсказания $V(p)$ как функция интервала анализа для вокализованного сегмента реального речевого сигнала [15]:
△ — автокорреляционный метод; ○ — ковариационный метод

ния V монотонно убывают приблизительно до 0,1 при $p=11$. Таким образом, в данном случае при асинхронной обработке, по крайней мере, синтетического звука оба метода приводят к сходным результатам.

На рис. 8.12 показана зависимость V от N для предсказателя при $p=12$ на сегменте синтетической речи. Как и предполагалось, для значений N, значительно меньших периода основного тона (83 отсчета), ковариационный метод приводит к значительно меньшим V, чем автокорреляционный. Величина V резко возрастает в области гармоник основного тона и содержит большие скачки из-за высокой погрешности предсказания в случае использования импульсов для возбуждения системы. Но при достаточно больших значениях N (два и более периодов основного тона) оба подхода приводят к сравнимым значениям V. На рис. 8.13—8.15 приведены аналогичные результаты для вокализованного сегмента реального речевого сигнала. Из рис. 8.13 видно, что при синхронном анализе нормированная погрешность для ковариационного метода значительно меньше, чем для автокорреляционного, а при асинхронном анализе (см. рис. 8.14) результаты сравнимы. Наконец, рис. 8.15 показывает изменение V в зависимости от N при анализе сигнала с $p=12$. В области периода основного тона значение V для автокорреляционного метода заметно изменяется, в то время как при ковариационном методе анализа изменения незначительны. При больших N кривые V для обоих методов приближаются друг к другу.

8.5.3. Зависимость нормированной погрешности предсказания от положения интервала анализа

В 8.5.2 рассмотрены некоторые свойства нормированной погрешности предсказания, а именно зависимость дисперсии погрешности от протяженности временно́го окна N и порядка модели. Остался еще один источник изменения V — изменение дисперсии при изменении положения интервала анализа. Для иллюстрации этого эффекта на рис. 8.16 представлены результаты анализа сегмента длительностью 40 мс гласного звука $|i|$, произнесенного мужским голосом в случае, когда интервал анализа последовательно перемещался каждый раз на один отсчет. На рис. 8.16*а* представлена энергия сигнала (частота дискретизации 10 кГц), на рис. 8.16*б* — нормированная средняя квадратическая погрешность V (частота дискретизации по-прежнему 10 кГц) при использовании ковариационного метода для модели с 14 полюсами ($p=14$) и интервалом анализа 20 мс ($N=200$). На рис. 8.16*в* показана нормированная средняя квадратическая погрешность при использовании окна Хемминга, на рис. 8.16*г* — средняя квадратическая погрешность при использовании автокорреляционного метода в случае прямоугольного окна. Средний период основного тона для этого диктора равен 84 отсчетам (т. е. 8,4 мс), т. е. интервал анализа составляет приблизительно 2,5 периода основного тона, или 20 мс.

Для ковариационного метода имеются значительные изменения погрешности предсказания в зависимости от положения временно́го окна (т. е. погрешность не является гладкой функцией

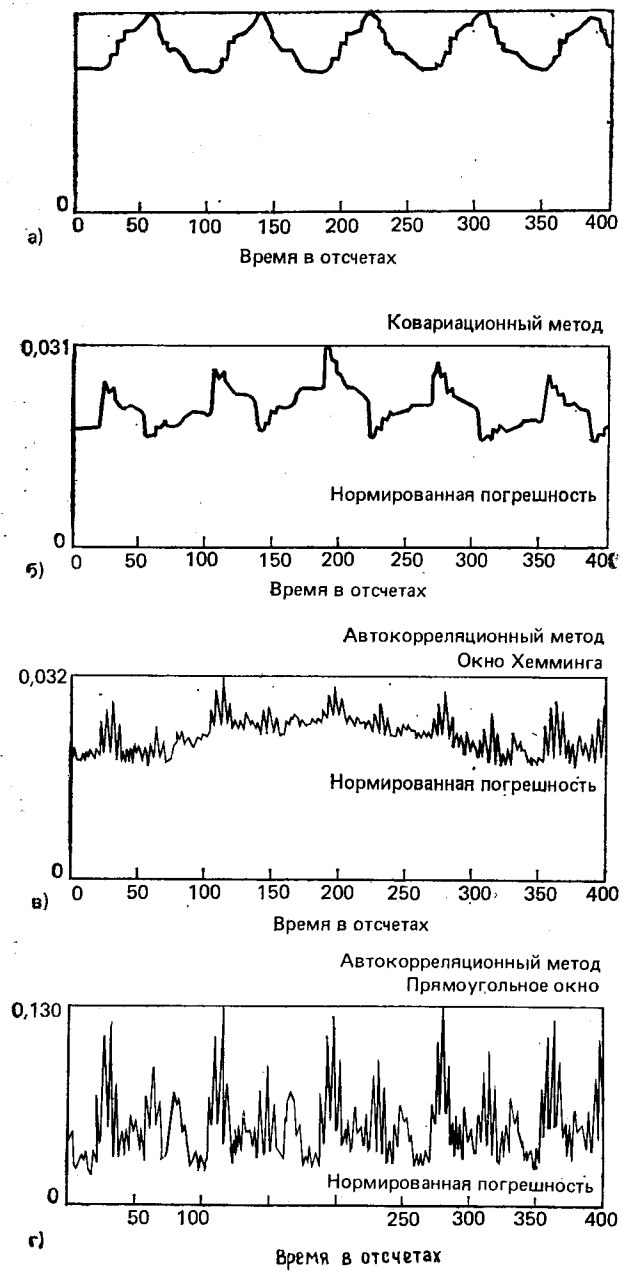


Рис. 8.16. Последовательность погрешности предсказания для 200 отсчетов речевого сигнала и трех систем линейного предсказания [16]

времени). Этот эффект обусловлен наличием значительных пиков погрешности предсказания в начале каждого периода основного тона. Когда в интервал анализа попадают три пика погрешности, нормированная ошибка оказывается больше, чем в случае двух пиков погрешности предсказания. Этим и объясняется наличие резких разрывов в погрешности при попадании в интервал анализа очередного пика в ошибке предсказания. Каждый скачок нормированной погрешности следует непосредственно за участком сглаженной нормированной погрешности предсказания. Точное поведение нормированной погрешности между скачками зависит от особенностей сигнала и метода анализа.

На рис. 8.16*в* и *г* показан различный до некоторой степени характер поведения погрешности при использовании автокорреляционного метода анализа для окна Хемминга и прямоугольного окна. Как видно из рисунков, в данном случае нормированный средний квадрат погрешности предсказания содержит в основном высокочастотные компоненты и слабо зависит от наличия импульсов основного тона в интервале анализа. Высокочастотные компоненты объясняются наличием в каждом интервале анализа первых p отсчетов, которые линейно непредсказуемы. Эти флуктуации при использовании окна Хемминга значительно меньше, чем в случае прямоугольного окна, поскольку окно Хемминга уменьшается к концу интервала анализа. Другая компонента высокочастотных флуктуаций определяется взаимным расположением импульсов основного тона и начала интервала анализа, как это выше излагалось для ковариационного метода. Однако в автокорреляционном методе этот эффект сказывается слабее, чем в ковариационном, особенно при использовании окна Хемминга, поскольку новые импульсы основного тона, попадающие в интервал анализа, ослабляются вследствие применения окна.

Изменения, подобные показанным на рис. 8.16, типичны для большинства гласных звуков [16]. Флуктуации в погрешности из-за различного положения окна могут быть ослаблены применением фильтрации и спектрального предыскажения перед обработкой сигнала с помощью линейного предсказания [16].

8.6. Анализ линейного предсказания в частотной области

До сих пор методы линейного предсказания рассматривались на основе разностных уравнений и корреляционных функций, т. е. с позиций представления сигнала во временной области. Однако предполагалось, что коэффициенты линейного предсказания являются коэффициентами знаменателя передаточной функции, описывающей действие речевого тракта, формы сигнала возбуждения и излучения. Таким образом, располагая совокупностью параметров предсказания, можно определить частотную характеристику модели речеобразования путем простой подстановки в $H(z)$ значения $z=e^{i\omega}$, т. е.

$$H(e^{i\omega}) = \frac{G}{1-\sum_{k=1}^{p}\alpha_k e^{-i\omega_k}} = \frac{G}{A(e^{i\omega})}. \qquad (8.102)$$

Если изобразить $H(e^{i\omega})$ как функцию частоты[1], то можно ожидать, что на формантных частотах будут видны максимумы, как и при рассмотрении спектральных представлений в предыдущей главе. Таким образом, линейное предсказание можно рассматривать как метод кратковременной оценки спектра. Действительно, подобные методы широко применяются не только при обработке сигналов речи [12]. В данном параграфе метод линейного предсказания по минимуму среднего квадрата ошибки трактуется на основе частотного представления и проводится сравнение данного метода с другими методами представления речевого сигнала в частотной области.

8.6.1. Спектральная трактовка среднего квадрата погрешности предсказания

Рассмотрим параметры линейного предсказания, полученные с помощью автокорреляционного метода. В этом случае погрешность представления во временно́й области будет иметь вид

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m), \tag{8.103а}$$

или в частотной области на основе теоремы Парсеваля

$$E_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S_n(e^{i\omega})|^2 |A(e^{i\omega})|^2 d\omega, \tag{8.103б}$$

где $S_n(e^{i\omega})$ — преобразование Фурье для сегмента сигнала $s_n(m)$, а

$$A(e^{i\omega}) = 1 - \sum_{k=1}^{p} \alpha_k e^{-i\omega k}. \tag{8.104}$$

Вспоминая, что

$$H(e^{i\omega}) = G/A(e^{i\omega}), \tag{8.105}$$

и используя (8.103), можно получить

$$E_n = \frac{G^2}{2\pi}\int_{-\pi}^{\pi} \frac{|S_n(e^{i\omega})|^2}{|H(e^{i\omega})|^2} d\omega. \tag{8.106}$$

Поскольку подынтегральное выражение в (8.106) положительно, то минимизация E_n эквивалентна минимизации отношения энергетического спектра сигнала к квадрату модуля частотной характеристики линейной системы в модели речеобразования.

В § 8.2 показано, что автокорреляционная функция $R_n(m)$ сегмента речевого сигнала $s_n(m)$ и автокорреляционная функция $\tilde{R}(m)$ импульсной характеристики $h(m)$, соответствующей системной функции $H(z)$, совпадают для первых $(p+1)$ значений. Таким образом, при $p \to \infty$ соответствующие автокорреляционные функции совпадают при всех значениях и, следовательно,

$$\lim_{p\to\infty} |H(e^{i\omega})|^2 = |S_n(e^{i\omega})|^2. \tag{8.107}$$

[1] См. задачу 8.2, где рассмотрен метод вычисления $H(e^{i\omega})$ с использованием БПФ.

Это означает, что при достаточно большом p можно аппроксимировать спектр сигнала с любой точностью.

Для иллюстрации возможностей спектрального анализа на основе линейного предсказания на рис. 8.17 [7] показаны кривые $20\log_{10}|H(e^{i\omega})|$ и $20\log_{10}|S_n(e^{i\omega})|$. Спектр сигнала получен по

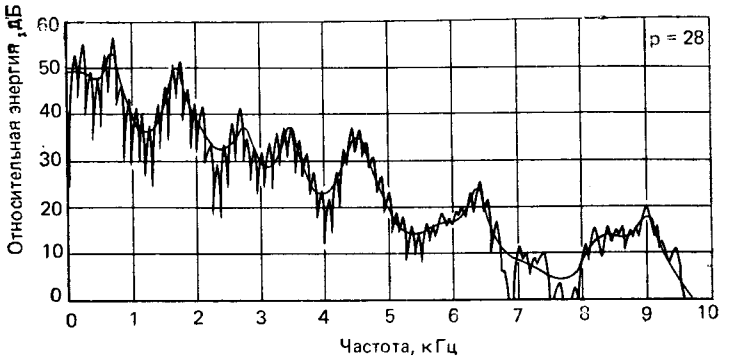


Рис. 8.17. Спектр речевого сигнала и 28-полюсной модели [17]

алгоритму БПФ на сегменте речевого сигнала длительностью 20 мс (при частоте дискретизации 20 кГц) с использованием окна Хемминга (см. гл. 6). Расчет сделан для звука $|ae|$. Спектр системы с линейным предсказанием получен для случая, когда используется предсказатель 28-го порядка ($p=28$), а его коэффициенты рассчитаны по автокорреляционному методу [2]. На рисунке отчетливо видна гармоническая структура спектра сигнала. На том же рисунке проявляется важная особенность анализа спектра с помощью линейного предсказания. Спектральное описание линейного предсказания лучше согласуется со спектром сигнала: более высокая точность обеспечивается в области больших значений спектральной плотности (т. е. вблизи максимумов спектра), более низкая точность — в области малых значений (т. е. в области провалов спектра). Это не является неожиданным, если учесть, что в соответствии с (8.106) в общую погрешность предсказания большой вклад вносят области, где $|S_n(e^{i\omega})| > |H(e^{i\omega})|$, по сравнению с областями, в которых $|S_n(e^{i\omega})| < |H(e^{i\omega})|$. Таким образом, спектральное описание на основе линейного предсказания, отвечая критерию оптимальности, приводит к хорошим результатам в области спектральных максимумов и к значительно худшим — в области минимумов спектра.

Проведенное выше обсуждение позволяет считать, что выбором порядка предсказателя p можно добиться нужной степени сглаживания спектра. Это утверждение иллюстрирует рис. 8.18, на котором показан сегмент речевого сигнала, его преобразование Фурье и спектры, полученные на основе линейного предсказания при различном порядке p. Очевидно, что увеличение p приводит к более детальному описанию спектральной плотности. Поскольку

наша задача сводится к получению лишь спектральных изменений, обусловленных совместным действием источника возбуждения, речевого тракта и излучения, требуется выбирать p таким образом, чтобы сохранить положение формантных максимумов и особенности формы спектральной плотности.

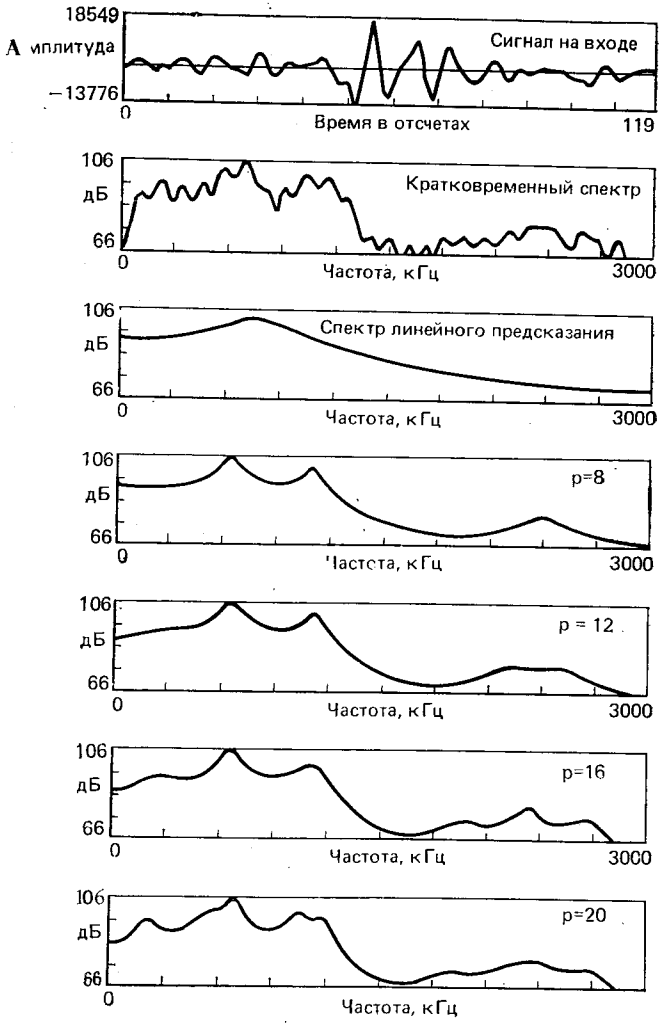


Рис. 8.18. Спектр гласного /a/ при частоте дискретизации 6 кГц для различных значений порядка предсказателя p

Предполагается, что коэффициенты предсказания оцениваются по автокорреляционному методу. Только в этом случае преобразование Фурье кратковременной автокорреляционной функции совпадает с квадратом модуля кратковременного преобразования

Фурье сигнала. Однако это не исключает использования $H(e^{i\omega})$ в качестве оценки спектра, даже если коэффициенты предсказания оцениваются по ковариационному методу.

8.6.2. Сравнение кратковременного спектрального анализа с оценкой спектра на основе линейного предсказания

В качестве примера на рис. 8.19 показаны четыре логарифмических спектра сегмента синтетического гласного |a| [10]. Первые две кривые получены с использованием методов кратковременного анализа спектра, рассмотренных в гл. 6. В первом случае использовался взвешенный сегмент сигнала длительностью 512 отсчетов (51,2 мс), который преобразовывался (с использованием 512-точечного БПФ) для осуществления относительно узкополосного спектрального анализа, результат которого показан в верхней части рис. 8.19. В данном спектре хорошо просматриваются отдельные гармоники сигнала возбуждения, что объясняется большой протяженностью интервала анализа.

На втором рисунке интервал анализа уменьшен до 128 отсчетов (12,8 мс), что привело к широкополосному спектральному анализу сигнала. Здесь уже отдельные гармоники неразличимы, просматривается огибающая спектра в целом. Хотя в данном случае формантные частоты в спектре видны хорошо, но их однозначная идентификация или оценка затруднительна.

Третий спектр получен на основе гомоморфного сглаживания, рассмотренного в гл. 7. Несглаженный спектр был рассчитан по 300 отсчетам (30 мс) с использованием БПФ. Сглаженный спектр, показанный на рисунке, был получен линейным сглаживанием логарифма спектра. В данном случае отдельные форманты легко разделяются и могут быть измерены с использованием простых методов

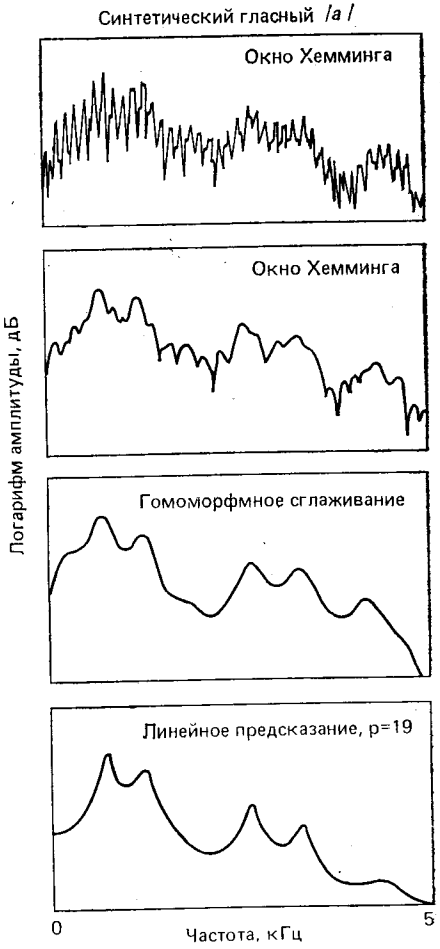


Рис. 8.19. Спектр синтетического звука /a/

поиска экстремумов. Однако оценка полосы форманты в данном случае очень сложна из-за использования сглаживания, применяемого для получения окончательного спектра.

Спектр, показанный в нижней части рисунка, получен в результате анализа на основе линейного предсказания с использованием модели при $p=12$ для сегмента длительностью 128 отсчетов (12,8 мс). Сравнивая этот спектр с другими, можно отметить, что параметрическое описание позволяет четко выявить формантную структуру без дополнительных побочных экстремумов и флуктуаций. Это объясняется тем, что модель линейного предсказания хорошо описывает речевой тракт в случае гласных звуков при правильном выборе порядка p. Поскольку порядок модели можно определить по полосе сигнала, метод линейного предсказания приводит к хорошей оценке спектральных свойств источника возбуждения, речевого тракта и излучателя.

На рис. 8.20 показано сравнение спектров сегмента натурального гласного звука, полученных гомоморфным сглаживанием и

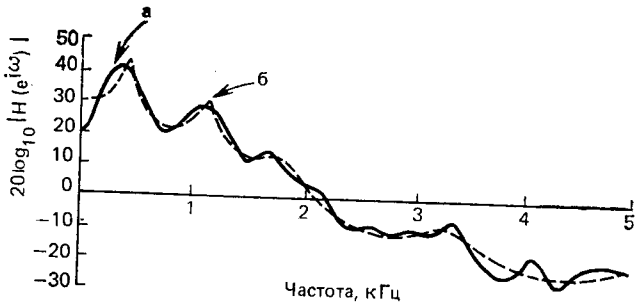


Рис. 8.20. Сравнение спектра речи, полученного с помощью кепстрального сглаживания (*а*) и линейного предсказания (*б*)

с использованием линейного предсказания. Хотя формантные частоты в обоих случаях практически совпадают, спектр линейного предсказания имеет меньше побочных пиков. Это объясняется тем, что при анализе на основе линейного предсказания и порядке модели $p=12$ в спектре может возникнуть не более шести пиков. При гомоморфном анализе такое ограничение отсутствует. Как отмечалось выше, спектральные пики при анализе на основе линейного предсказания более острые, чем при гомоморфном анализе, что обусловлено применением сглаживания спектра в последнем случае.

8.6.3. Селективное линейное предсказание

Изложенные выше идеи можно применять не ко всей полосе спектра сигнала, а только к ее части. Такой метод назван Макхоулом методом селективного линейного предсказания [8]. Появление этого метода объясняется тем, что в ряде случаев необхо-

димо использовать лишь часть спектральной плотности сигнала. Например, для адекватного описания фрикативов в системах распознавания речи используется частота дискретизации, равная 20 кГц. Для вокализованных сегментов при этом требуется диапазон от 0 до 4 кГц, а для невокализованных сегментов более важен диапазон от 4 до 8 кГц. Используя селективное представление, спектр в диапазоне от 0 до 4 кГц можно сформировать на основе предсказателя порядка p_1, а спектр в диапазоне от 4 до 8 кГц — на основе другого предсказателя — порядка p_2.

Селективное линейное предсказание осуществляется следующим способом. Чтобы сформировать сигнал лишь в диапазоне от $f=f_A$ до $f=f_B$, требуется лишь линейно отобразить эту область так, что $f=f_A$ отображается в $f'=0$, а $f=f_B$ в $f'=\omega'/2\pi=0{,}5$ (т. е. в половину частоты дискретизации). Параметры предсказания являются решением системы уравнений предсказания, коэффициенты корреляции в которой получены по формуле

$$R'(i)=\frac{1}{2\pi}\int_{-\pi}^{\pi}|S_n(e^{i\omega'})|^2 e^{i\omega' i}\,d\omega'. \qquad (8.108)$$

На рис. 8.21 представлены результаты селективного линейного предсказания [8]. Исходный сигнал был тем же, что и на рис. 8.17. В диапазоне от 0 до 5 кГц использовалась модель с $p_1=14$,

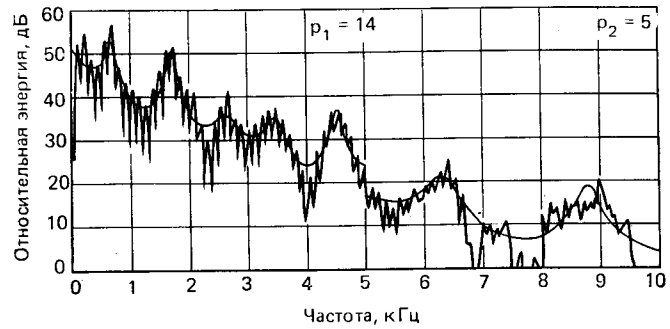


Рис. 8.21. Применение селективного линейного предсказания к спектру сигнала, изображенному на рис. 8.17, с использованием 14-полюсной модели в диапазоне 0—5 кГц и 5-полюсной модели в диапазоне 5—10 кГц [2]

а диапазон от 5 до 10 кГц описывался независимо моделью с $p_2=5$. На частоте 5 кГц имеется разрыв в спектральной плотности, что объясняется отсутствием условий согласования обоих спектров.

8.6.4. Сравнение методов линейного предсказания с методами анализа через синтез

Как излагалось в гл. 6, мера погрешности, обычно используемая в методе анализа через синтез, представляет собой логарифм

отношения спектральной плотности мощности сигнала к квадрату модуля спектральной плотности мощности модели:

$$E' = \int_{-\pi}^{\pi} \left\{ \log \left[\frac{|S_n(e^{i\omega})|^2}{|H(e^{i\omega})|^2} \right] \right\}^2 d\omega. \qquad (8.109)$$

Таким образом, минимизация E' в методе анализа через синтез эквивалентна минимизации среднего квадратического отклонения между логарифмами спектров.

Сравнение мер погрешности моделирования на основе методов линейного предсказания и анализа через синтез приводит к следующим выводам:

1. Оба метода связаны с соотношением спектров сигнала и модели.

2. Оба метода одинаково учитывают различные частотные диапазоны.

3. Оба метода пригодны для минимизации погрешности в некотором выбранном диапазоне.

4. Критерий качества в линейном предсказании более чувствителен к тем участкам спектра, на которых $|S_n(e^{i\omega})|^2 > |H(e^{i\omega})|^2$, в то время как в методе анализа через синтез критерий качества одинаково чувствителен во всем диапазоне.

Из сказанного следует, что при анализе несглаженного спектра (см. рис. 8.17) критерий метода линейного предсказания приводит к лучшим результатам, чем критерий метода анализа через синтез [7]. Более того, объем вычислений, необходимых при использовании линейного предсказания, значительно меньше. Если же анализируется сигнал с гладким спектром (например, с помощью набора фильтров), то как анализ на основе линейного предсказания, так и метод анализа через синтез приводят к хорошему совпадению спектров. На практике для спектральных плотностей такого типа почти всегда применяется метод анализа через синтез.

8.7. Применение анализа на основе линейного предсказания к моделям речевого тракта в виде труб без потерь

В гл. 3 рассматривалась модель речеобразования, включавшая в себя последовательное соединение N акустических труб без потерь (рис. 8.22). Коэффициенты отражения r_k на рис. 8.22 связаны с площадями поперечного сечения соотношением

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}. \qquad (8.110)$$

В 3.3.4 получена передаточная функция такой системы в предположении, что коэффициент отражения от источника возбуждения $r_G = 1$, т. е. сопротивление источника предполагается бесконечно

большим. Передаточная функция системы, представленной на рис. 8.22, как показано в 3.3.4, имеет вид

$$V(z) = \frac{\prod_{k=1}^{N} (1 + r_k)\, z^{-N/2}}{D(z)}, \qquad (8.111)$$

где $D(z)$ удовлетворяет соотношениям:

$$D_0(z) = 1; \qquad (8.112а)$$

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}); \qquad (8.112б)$$

$$D(z) = D_N(z). \qquad (8.112в)$$

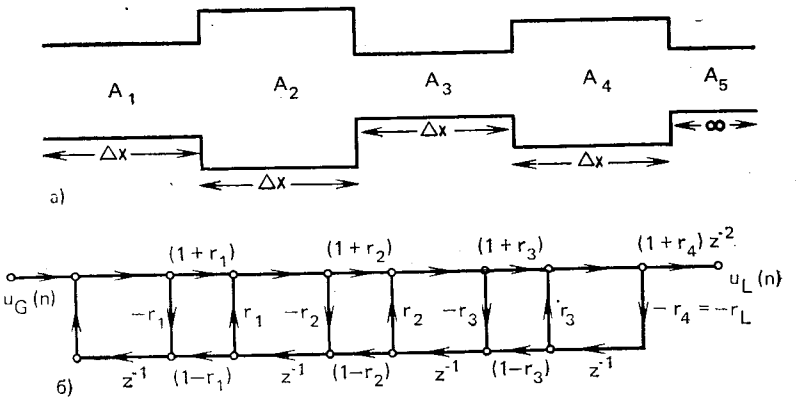


Рис. 8.22. Модель трубы без потерь, нагруженной на трубу бесконечной длины (*а*) и граф прохождения сигнала при бесконечном сопротивлении источника (*б*)

Все это весьма напоминает обсуждение лестничного метода в 8.8.3. Действительно, там было показано, что полином

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}, \qquad (8.113)$$

полученный при анализе на основе линейного предсказания, можно получить с использованием рекурсивной процедуры:

$$A^{(0)}(z) = 1; \qquad (8.114а)$$

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}); \qquad (8.114б)$$

$$A(z) = A^{(p)}(z), \qquad (8.114в)$$

где параметры $\{k_i\}$ названы коэффициентами частной корреляции. Сравнивая уравнения (8.112) и (8.114), замечаем, что передаточная функция

$$H(z) = G/A(z), \qquad (8.115)$$

полученная на основе линейного предсказания, имеет тот же вид, что и передаточная функция акустической трубы без потерь, имеющей p секций. Если

$$r_i = -k_i, \tag{8.116}$$

то очевидно, что

$$D(z) = A(z). \tag{8.117}$$

Используя (8.110) и (8.116), легко показать, что эквивалентные площади поперечных сечений модели в виде неоднородной трубы связаны с коэффициентами частных корреляций соотношением

$$A_{i+1} = \left[\frac{1-k_i}{1+k_i}\right] A_i. \tag{8.118}$$

Заметим, что частные корреляции определяют соотношение площадей соседних секций. Таким образом, площади поперечного сечения модели в виде неоднородной трубы не определяются абсолютно точно, а все модели с подходящими условиями нормализации дают одинаковую передаточную функцию.

«Функция площади», полученная с использованием (8.118), не является соответствующей функцией для речевого тракта человека. Однако Вакита [17] показал, что при использовании предыскажений, устраняющих влияние источника возбуждения и излучения, функции площади, описывающие речевой тракт, часто бывают весьма сходными с конфигурацией голосового тракта, используемого человеком при речеобразовании.

8.8. Соотношения между различными параметрами речи

Хотя коэффициенты предсказания α_k, $1 \leqslant k \leqslant p$, часто считаются основными параметрами при анализе речи на основе линейного предсказания, обычно сразу же возникает задача преобразования этих параметров в некоторые другие для получения иных представлений речевого сигнала. Эти представления часто оказываются более удобными при применении линейного предсказания. В этом разделе рассматриваются методы получения других полезных описаний сигнала на основе непосредственного использования параметров линейного предсказания [1, 2].

8.8.1. Корни полинома передаточной функции предсказателя

Вместо коэффициентов линейного предсказания можно использовать корни полинома

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \sum_{k=1}^{p} (1 - z_k z^{-1}). \tag{8.119}$$

Множество корней $\{z_i,\ i=1,\ 2,\ \ldots,\ p\}$ представляет собой эквивалентное представление $A(z)$. При необходимости пересчитать корни на z-плоскости в корни на s-плоскости воспользуемся подстановкой

$$z_i = e^{s_i T}, \tag{8.120}$$

где $s_i = \sigma_i + i\,\Omega_i$ — корень на s-плоскости, соответствующий корню z_i на z-плоскости. Если $z_i = z_{ir} + i\,z_{ii}$, то

$$\Omega_i = (1/T)\,\tan^{-1}(z_{ii}/z_{ir}) \tag{8.121}$$

и

$$\sigma_i = (1/2T)\,\log(z_{ir}^2 + z_{ii}^2). \tag{8.122}$$

Соотношения (8.121) и (8.122) полезны в случаях применения линейного предсказания в формантном анализе.

8.8.2. Кепстр

Другим представлением сигнала является кепстр импульсной характеристики всей системы линейного предсказания. Если система с линейным предсказанием имеет передаточную функцию $H(z)$ с импульсной реакцией $h(n)$ и комплексным кепстром $\hat{h}(n)$, то можно показать, что кепстр $\hat{h}(n)$ получается с помощью рекурсивных соотношений

$$\hat{h}(n) = \alpha_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \hat{h}(k)\,\alpha_{n-k},\ n \geqslant 1, \tag{8.123}$$

где

$$H(z) = \sum_{n=0}^{\infty} h(n)\,z^{-n} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}. \tag{8.124}$$

8.8.3. Импульсная характеристика полюсной системы

Импульсная характеристика $h(n)$ полюсной системы с передаточной функцией (8.124) может быть определена на основе рекурсивного уравнения

$$h(n) = \sum_{k=1}^{p} \alpha_k\,h(n-k) + G\,\delta(n), \tag{8.125}$$

где $h(n)$ предполагается (по определению) равной 0 для $n<0$ и G — амплитуда возбуждения.

8.8.4. Автокорреляционная функция импульсной характеристики

Как отмечалось в § 8.2, автокорреляционная функция импульсной характеристики определяется выражением (см. задачу 8.1)

$$\widetilde{R}(i)=\sum_{n=0}^{\infty} h(n)\,h(n-i)=\widetilde{R}(-i). \tag{8.126}$$

и удовлетворяет соотношениям

$$\widetilde{R}(i)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(|i-k|). \tag{8.127}$$

и

$$\widetilde{R}(0)=\sum_{k=1}^{p}\alpha_k\widetilde{R}(k)+G^2. \tag{8.128}$$

Уравнения (8.127) и (8.128) можно использовать для определения $\widetilde{R}(i)$ по коэффициентам предсказания и наоборот.

8.8.5. Коэффициенты автокорреляции полиномиальной передаточной функции предсказателя

Полином передаточной функции предсказателя (обратного фильтра) имеет вид

$$A(z)=1-\sum_{k=1}^{p}\alpha_k z^{-k}, \tag{8.129}$$

а импульсная характеристика равна

$$a(n)=\delta(n)-\sum_{k=1}^{p}\alpha_k\delta(n-k).$$

Автокорреляционная функция импульсной характеристики обратного фильтра определяется соотношением

$$R_a(i)=\sum_{k=0}^{p-i}a(k)\,a(k+i),\ 0\leqslant i\leqslant p. \tag{8.130}$$

8.8.6. Коэффициенты частной корреляции

Для автокорреляционного метода коэффициенты предсказания можно получить по коэффициентам частной корреляции, используя рекурсивные соотношения:

$$a_i^{(i)}=k_i; \tag{8.131а}$$

$$a_j^{(i)}=a_j^{(i-1)}-k_i\,a_{i-j}^{(i-1)},\ 1\leqslant j\leqslant i-1, \tag{8.131б}$$

решая (8.131а) и (8.131б) для $i=1, 2, \ldots, p$ и устанавливая последний набор коэффициентов равным

$$\alpha_j=a_j^{(p)},\ 1\leqslant j\leqslant p. \tag{8.131в}$$

Аналогично частные корреляции можно рассчитать по коэффициенту линейного предсказания, используя возвратную рекурсию в виде:

$$k_i=a_i^{(i)}; \tag{8.132а}$$

$$a_j^{(i-1)}=\frac{a_j^{(i)}+a_i^{(i)}\,a_{i-j}^{(i)}}{1-k_i^2},\ 1\leqslant j\leqslant i-1, \tag{8.132б}$$

где i изменяется от p к $p-1$ и т. д. до 1 и финальное решение есть

$$a_j^{(p)}=\alpha_j,\ 1\leqslant j\leqslant p. \tag{8.132в}$$

8.8.7. Логарифм отношения площадей

Важной совокупностью эквивалентных параметров, которые можно получить по коэффициентам частных корреляций, является совокупность логарифмов отношений площадей поперечного сечения, определяемых как

$$g_i=\log\left[\frac{A_{i+1}}{A_i}\right]=\log\left[\frac{1-k_i}{1+k_i}\right],\ 1\leqslant i\leqslant p. \tag{8.133}$$

Параметры g_i эквивалентны логарифму отношения площадей поперечных сечений соседних секций в модели неоднородной акустической трубы без потерь, которая имеет ту же передаточную функцию, что и модель линейного предсказания (см. § 8.7). Параметры g_i, как установлено в [2] и другими [1], наиболее удобны для квантования вследствие незначительной спектральной чувствительности величин g'_i.

Параметры k_i можно непосредственно получить по параметрам g_i с помощью обратного преобразования:

$$k_i=\frac{1-e^{g_i}}{1+e^{g_i}},\ 1\leqslant i\leqslant p. \tag{8.134}$$

8.9. Синтез речевого сигнала по параметрам линейного предсказания

Речевой сигнал может быть синтезирован по параметрам линейного предсказания различными способами. Простейший способ состоит в использовании для синтеза системы, описываемой теми же параметрами, которые применялись при анализе. На рис. 8.23 изображена структурная схема такого синтезатора. Для синтеза речевого сигнала в данном случае используются такие меняющиеся во времени параметры, как период основного тона, признак тон—шум, коэффициент усиления или минимальное среднее квадратическое значение и p коэффициентов линейного предсказания. Импульсный генератор в данном случае работает как источник возбуждения на вокализованных звуках, формирующий импульсы

в начале каждого периода основного тона. Генератор шума представляет собой источник возбуждения на невокализованных сегментах, формирующий некоррелированный равномерно распределенный случайный процесс с единичной дисперсией и нулевым

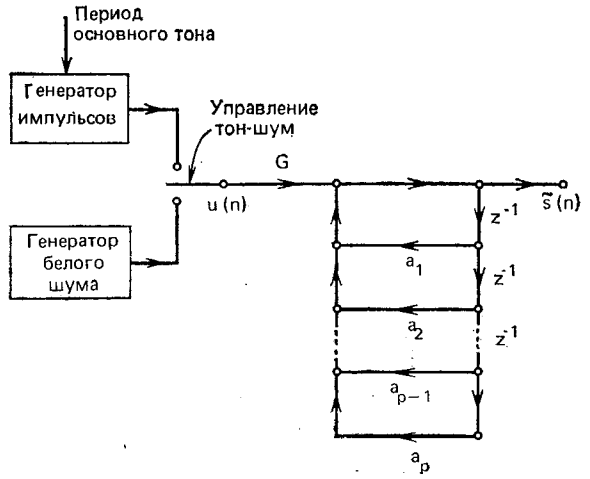


Рис. 8.23. Структурная схема синтезатора на основе линейного предсказания

средним. Выбор между двумя источниками обеспечивается с помощью признака тон—шум. Коэффициент усиления G определяет полную амплитуду возбуждения. Отсчет синтезированной речи определяется соотношением

$$\tilde{s}(n)=\sum_{k=1}^{p}\alpha_k\,\tilde{s}(n-k)+G\,u(n). \qquad (8.135)$$

Схема устройства, реализующего (8.135), представлена на рис. 8.23. Эта схема представляет собой простой и непосредственный способ реализации синтезатора речи по параметрам предсказания. Для воспроизведения каждого отсчета требуется p умножений и p сложений.

В модели синтеза, представленной на рис. 8.23, параметры синтезатора должны изменяться во времени. Хотя обычно оценка параметров производится периодически на интервалах вокализованной речи, управляющие параметры синтезатора изменяются в начале каждого периода основного тона. Для невокализованной речи они изменяются 1 раз на интервале (т. е. через каждые 10 мс для скорости 100 отсчетов на интервал анализа). Установлено, что подстройка управляющих параметров в начале каждого периода основного тона (называемая синтезом, синхронным с основным тоном) является более эффективной по сравнению с подстройкой 1 раз на интервале анализа (которая называется асинхронной). Это, в свою очередь, требует интерполяции параметров

для того, чтобы получить их значения в начале любого периода основного тона.

Установлено, что параметры усиления и основного тона следует интерполировать геометрически (т. е. линейно в логарифмическом масштабе) [3], однако вследствие требования устойчивости параметры линейного предсказания непосредственно интерполировать нельзя. Это обусловлено тем, что интерполяция между двумя устойчивыми множествами параметров может привести к неустойчивым параметрам. Одним из путей преодоления этой трудности, в соответствии с результатами Атала, является интерполяция первых p отсчетов автокорреляционной функции импульсной реакции фильтра (см. рис. 8.21). Используя соотношения, полученные в § 8.4, коэффициенты предсказания можно получить по первым p отсчетам автокорреляции импульсной характеристики и наоборот. Более того, интерполяция автокорреляционных коэффициентов всегда приводит к получению устойчивого фильтра[1].

Синтезатор, изображенный на рис. 8.23, используется в ряде приложений при моделировании систем с линейным предсказанием. Его основным достоинством является простота технической реализации. Существенный недостаток заключается в том, что синтезатор представляет собой прямую форму рекурсивного цифрового фильтра, что требует высокой точности при вычислении коэффициентов, ибо прямая форма программирования весьма чувствительна к изменениям коэффициентов. Другой, более удобный способ синтеза речевого сигнала может быть основан на использовании коэффициентов отражения или коэффициентов частных корреляций в рамках модели неоднородной трубы без потерь. Другими словами, схема фильтра, изображенного на рис. 8.23, может быть заменена схемой рис. 8.22. Преимущество этого подхода заключается в том, что в данном случае синтез проводится на основе ограниченных коэффициентов отражения $r_i=-k_i$ ($|k_i|<1$), которые можно интерполировать непосредственно, без нарушения устойчивости фильтра. Такая структура также менее чувствительна к погрешностям квантования, возникающим при цифровой реализации, чем фильтр в прямой форме.

Из рис. 8.22*б* очевидно также, что для реализации предсказателя порядка p в данном случае требуется $4p+2$ умножений и $2(p-1)$ сложений на отсчет по сравнению с p сложениями и умножениями в фильтре с прямой формой. В 3.3.3 показано, что секции с четырьмя умножениями можно заменить секциями с двумя умножениями за счет увеличения числа сложений. Осуществляя преобразования, показанные на рис. 3.41, граф рис. 8.22 можно изобразить в виде рис. 8.24. Рисунок 8.24*а* приводит к фильтру с $2p-1$ умножениями и $4p-1$ сложениями, а рис. 8.24*б*

[1] Аналогично можно интерполировать частные корреляции и логарифмы площади; устойчивость сохраняется при условии устойчивости исходных совокупностей параметров.

изображает фильтр с p умножениями и $3p-2$ сложениями. При использовании модели в виде неоднородной трубы без потерь для целей синтеза выбор той или иной формы зависит от ряда факторов, так что невозможно утверждать однозначно, какая форма является более эффективной.

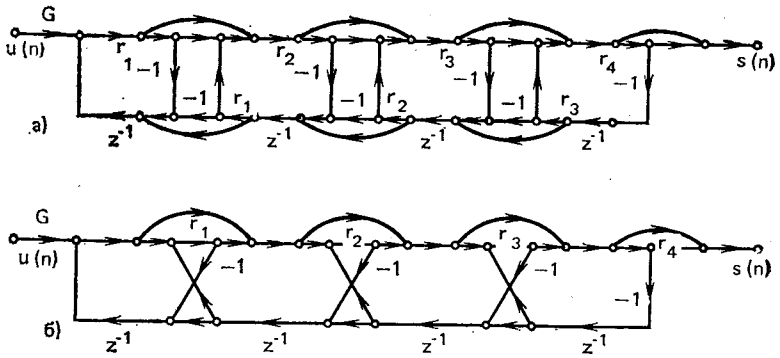


Рис. 8.24. Эквивалентные модели акустической трубы без потерь:
а) с двумя операциями умножения; *б*) с одной операцией умножения

8.10. Применение параметров линейного предсказания

Как следует из результатов предыдущего параграфа, теория линейного предсказания достаточно хорошо разработана. Разработаны методы оценки всех основных параметров речевого сигнала. На основе такого анализа проведены обширные исследования вокодеров, что привело к пониманию свойств различных представлений линейного предсказания применительно к квантованию. Эти методы, наконец, получили распространение при решении задач верификации и идентификации дикторов, распознавания и классификации речи, устранения реверберации и т. д. В § 8.10 будут представлены методы оценивания параметров речевого сигнала на основе линейного предсказания.

8.10.1. Оценивание основного тона на основе коэффициентов линейного предсказания

Выше уже обсуждался вопрос о том, как на основе использования сигнала погрешности можно, по крайней мере теоретически, построить оценку основного тона. Хотя этот метод, вообще говоря, позволяет оценить период основного тона достаточно точно, в [19] предложен несколько иной алгоритм, называемый SIFT-методом (метод обратной фильтрации). Сходный подход предложен в [20].

На рис. 8.25 представлена структурная схема SIFT-алгоритма. Входной сигнал $s(n)$ поступает на вход фильтра нижних частот с частотой среза около 900 Гц и затем обычная частота дискретизации 10 кГц снижается до 2 кГц путем прореживания (т. е. каждые четыре из пяти отсчетов выбрасываются). Прореженный

выход $x(n)$ затем анализируется с использованием автокорреляционного метода. Обратного фильтра четвертого порядка оказывается вполне достаточно для этих целей, поскольку в диапазоне до 1 кГц имеется не более двух формант. В результате анализа на выходе обратного фильтра получается сигнал с почти равномерным спектром[1]. Цель линейного предсказания, таким образом, заключается в выравнивании спектра подобно тому, как это делалось при клиппировании (см. гл. 4). Затем вычисляются кратковременная автокорреляционная функция погрешности предсказания и положение максимума в ней в подходящем интервале задержек выбирается в качестве оценки периода основного тона. Для получения дополнительной точности при оценке основного тона применяется интерполяция автокорреляционной функции в области максимального значения. Сегмент речи классифицируется как невокализованный, если максимальное значение автокорреляционной функции (нормированной соответствующим образом) оказывается ниже некоторого выбранного порога.

На рис. 8.26 [19] показаны колебания, полученные в различных точках анализатора. На рис. 8.26*а* изображен отрезок анализируемого входного сигнала, на рис. 8.26*б* — спектр входного сигнала и спектр сигнала на выходе обратного фильтра. В данном примере имеется лишь одна форманта на частоте 250 Гц. На рис. 8.26*в* показан спектр, а на рис. 8.26*г* — временна́я диаграмма сигнала на выходе обратного фильтра. Наконец, на рис. 8.26*д* представлена нормированная автокорреляционная функция входного сигнала, на которой хорошо виден период основного тона длительностью 8 мс.

Линейное предсказание в алгоритме SIFT используется для выравнивания спектра с целью облегчения оценивания основного тона. Метод дает весьма точные оценки периода основного тона до тех пор, пока спектр сигнала выравнивается достаточно хорошо. Однако для голосов с малым периодом основного тона (например, детских) этот метод выравнивания спектра приводит к плохим результатам

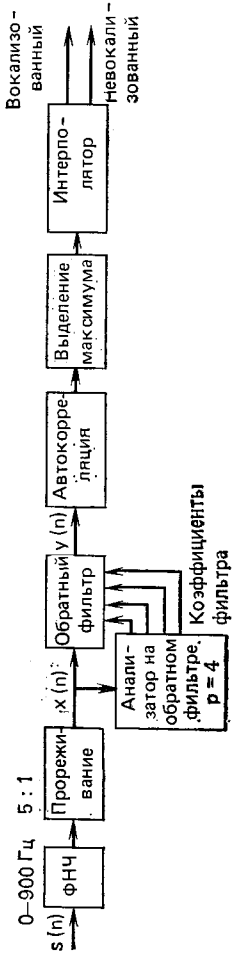


Рис. 8.25. Структурная схема алгоритма SIFT для выделения основного тона

[1] Сигнал на выходе фильтра — это погрешность предсказания для предсказателя четвертого порядка.

из-за отсутствия высших гармоник основного тона в полосе от 0 до 900 Гц (особенно при использовании сигналов с телефонных ли-

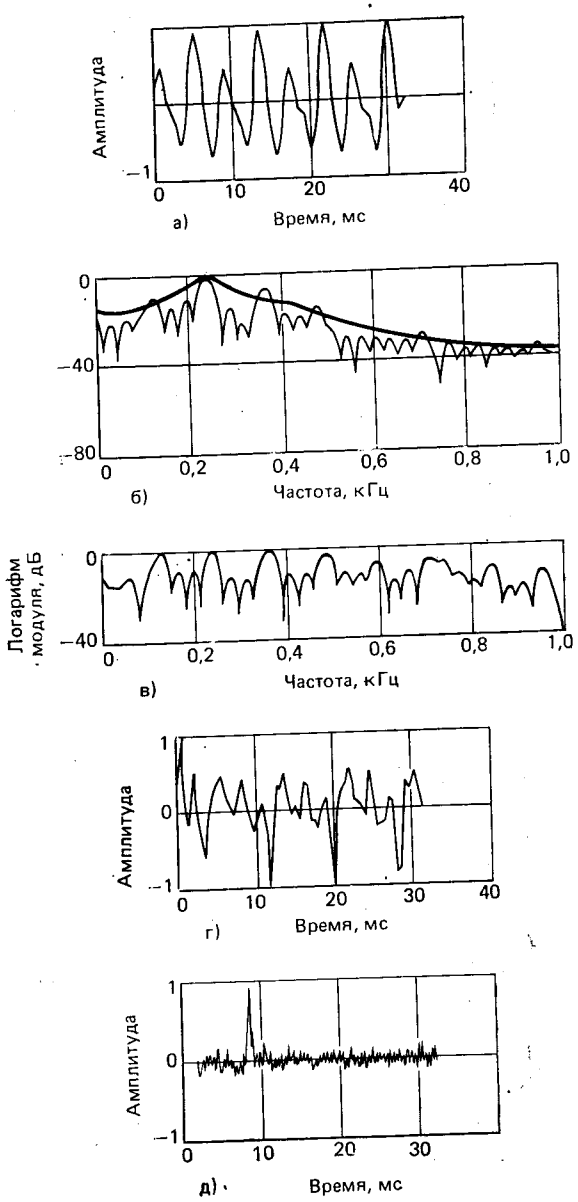


Рис. 8.26. Типичные сигналы алгоритма SIFT [19]

ний). Для таких дикторов и условий передачи лучшие результаты могут дать другие методы выделения основного тона.

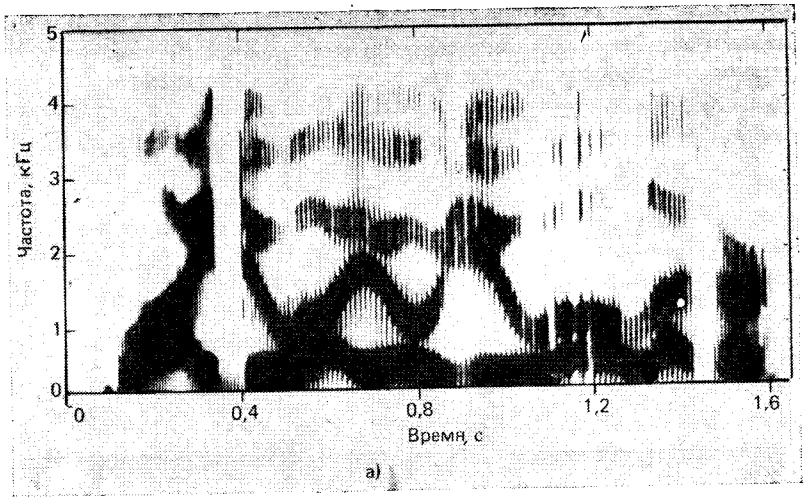
8.10.2. Формантный анализ с использованием коэффициентов линейного предсказания

Анализ речи на основе линейного предсказания при использовании его доли оценивания формантных частот вокализованного сигнала имеет как преимущества, так и недостатки. Форманты можно оценить по коэффициентам предсказания двумя способами. Первый состоит в факторизации полинома предсказания на основе полученных корней и вынесении решения о том, какие из корней описывают форманты, а какие — форму спектра [21, 22]. Другой способ заключается в оценивании спектра и использовании метода выделения максимумов, рассмотренного в гл. 7 [23].

Особое преимущество, присущее методу линейного предсказания в формантном анализе, состоит в том, что как центральные частоты формант, так и их полосы можно оценивать достаточно точно с помощью факторизации полинома предсказателя. Поскольку порядок полинома p выбирается заранее, количество комплексно-сопряженных полюсов составляет $p/2$. Таким образом, упомянутая выше проблема классификации корней полинома с целью определения того, какие из корней описывают форманты, в данном случае оказывается значительно менее сложной, чем при использовании сходных методов, например кепстрального сглаживания. Кроме того, побочные полюсы легко устраняются вследствие того, что полоса соответствующих им формант оказывается во много раз больше, чем можно ожидать для обычного речевого сигнала. На рис. 8.27 показан пример, показывающий, что положение полюсов в самом деле дает хорошее представление о формантных частотах [3].

Недостатком метода линейного предсказания является использование для описания спектра сигнала полюсной модели. Так, хотя для носовых звуков и получается неплохое описание спектра полюсной моделью, совпадение корней полинома и действительных формантных частот неочевидно. Совершенно неясно, чему соответствуют получающиеся корни: нулям и полюсам носовой полости или искомым резонансным частотам. Другая трудность заключается в том, что хотя оценки ширины формант можно определить с использованием полученных корней, однако непонятно, как они соотносятся с истинными формантами. Это объясняется тем, что полученная оценка зависит от расположения и длительности интервала анализа и метода анализа.

С учетом этих достоинств и недостатков предложен ряд методов оценивания формантных частот с использованием линейного предсказания как на основе метода выделения максимумов в спектре, так и на основе факторизации полинома предсказателя. После выбора совокупности формантных параметров устанавливается соответствие между формантными параметрами и номерами формант, как и во всех других методах анализа. Сюда входят и требования непрерывности формант, необходимости предыскажений для исключения взаимного поглощения одной фор-



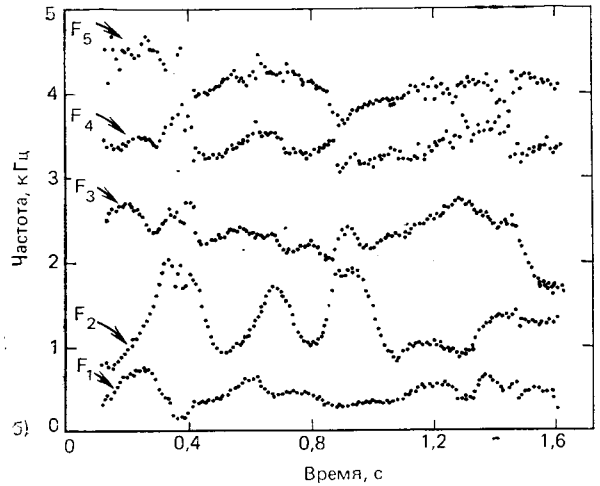


Рис. 8.27. Спектрограмма исходного сигнала (*а*) и центральные частоты (*б*) расположения комплексных полюсов 12-полюсной модели линейного предсказания [3]

манты другой и использования методов обострения пиков в спектре с помощью перемещения старшего параметра линейного предсказания к границе единичной окружности. Обсуждение различных методов содержится в работах Маркела [21, 22], Атала [3], Маххоула и Вольфа [5] и Мак-Кандлесса [23].

8.10.3. Вокодер на основе линейного предсказания

Наиболее важными областями применения линейного предсказания являются низкоскоростная передача речи (вокодеры) и ее хранение (для систем с машинным речевым ответом). На рис. 8.28

представлена структурная схема вокодера, построенного на основе линейного предсказания. Вокодер состоит из передатчика, канала связи и приемника. В передатчике вычисляются коэффициенты линейного предсказания и основной тон, которые затем

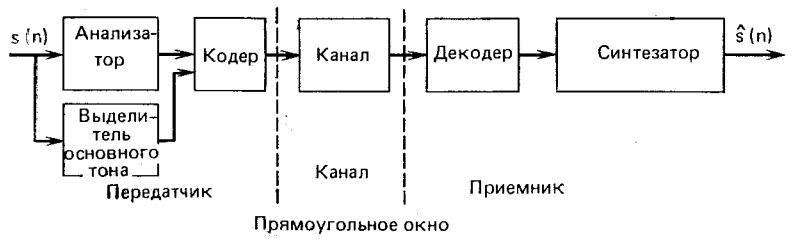


Рис. 8.28. Структурная схема вокодера с линейным предсказанием

кодируются для передачи по каналу связи. В приемнике происходят декодирование параметров и синтезирование выходного речевого сигнала. Выше уже рассматривались вопросы как анализа, так и синтеза сигнала. Для простоты предположим, что канал не вносит ошибок в передаваемое сообщение. Таким образом, в данном разделе рассматриваются различные множества параметров с точки зрения их пригодности для кодирования речи при заданной скорости передачи.

Основными параметрами являются: p коэффициентов линейного предсказания, период основного тона, признак тон—шум и коэффициент усиления. Методы подходящего кодирования периода основного тона, признака тон—шум и коэффициента усиления достаточно хорошо известны. Так, для представления периода основного тона необходимо 6 бит, для признака тон—шум — 1 бит, а для коэффициента усиления — 5 бит при логарифмическом квантовании [3].

Хотя принципиально можно и непосредственно квантовать параметры предсказания, такой подход из-за условий устойчивости требует относительно высокой точности представления (8—10 бит на параметр). Это связано с тем, что малые изменения параметров предсказания приводят к большим изменениям в расположении полюсов. Поэтому непосредственное квантование параметров предсказания не находит широкого применения.

Естественно, возникает вопрос выбора подходящей совокупности параметров, удобных для кодирования и передачи. Среди известных параметров наиболее подходящими являются корни полинома и коэффициенты отражения. Корни полинома можно квантовать таким образом, чтобы обеспечить устойчивость системы: расположение корней внутри единичного круга гарантирует устойчивость. Используя этот подход, Атал показал, что необходимо 5 бит на корень (т. е. 5 бит на полосу и 5 бит на центральную частоту). При этом синтетическая речь практически не отличима от синтетической речи, полученной без квантования параметров.

Используя такой метод кодирования, получаем, что скорость передачи составляет $72F_s$ бит/с, где F_s — количество интервалов анализа в секунду. Обычно значение F_s составляет 100, 67 и 33, что дает скорости 7200, 4800 и 2400 бит/с соответственно.

Другими параметрами, которые легко квантовать и для которых просто проверяется условие устойчивости, являются частные корреляции k_i. В данном случае условие устойчивости легко обеспечить при квантовании параметров. Макхоул и Висвансан [25] показали, что распределения частных корреляций весьма асимметричны, поэтому для правильного распределения фиксированного количества двоичных единиц необходимо предварительное преобразование параметров перед квантованием. Используя меру спектральной чувствительности, Макхоул и Висвансан [25] определили оптимальное преобразование вида

$$g_i = f(k_i) = \log\left[\frac{1-k_i}{1+k_i}\right] = \log\left[\frac{A_{i+1}}{A_i}\right], \quad 1 \leqslant i \leqslant p, \qquad (8.136)$$

где A_i — функции площади поперечного сечения неоднородной акустической трубы без потерь. Таким образом, оптимальными параметрами при кодировании речевого сигнала являются логарифмы отношений площадей поперечного сечения в рамках модели неоднородной трубы без потерь. Легко видеть, что соотношение (8.136) отображает интервал $-1 \leqslant k_i \leqslant 1$ в интервал $-\infty \leqslant g_i \leqslant \infty$. Используя это преобразование, Атал [27] показал, что коэффициенты g_i имеют почти равномерное распределение и малую корреляцию между параметрами и, таким образом, являются весьма удобными для цифрового представления. При использовании этих параметров для кодирования речи требуется 5—6 бит на параметр для получения такого же качества синтезированного сигнала, как и без квантования.

Во всех перечисленных выше случаях предполагалось, что для кодирования параметров используется один из методов ИКМ. В [26] показано, что использование методов кодирования различных параметров при линейном предсказании (см. гл. 5) позволяет дополнительно уменьшить скорость передачи сигнала. Используя АРИКМ при кодировании параметров предсказания, можно получить хорошее восприятие речи при скоростях 1000—2000 бит [26].

8.10.4. Полувокодер с линейным предсказанием [1]

Ранее было показано, что наиболее слабым звеном всех известных вокодеров является необходимость точной оценки источника возбуждения. В гл. 6 рассматривались некоторые вокодерные системы, которые не требовали непосредственной оценки основного тона и признака тон—шум, но описывали возбуждение через фазу или производную фазы сигнала. Другой подход, позволяющий избежать непосредственной оценки параметров возбуждения, состоит в использовании вокодеров, возбуждаемых речевым сигналом. Системы такого типа исследовались Аталом и др. [26] и Винстеном [27].

[1] Имеется в виду вокодер, в котором в качестве сигнала возбуждения применяется преобразованный речевой сигнал. (Прим. ред.)

На рис. 8.29 представлена структурная схема вокодера с речевым возбуждением. В этой системе имеются два отдельных подканала передачи, один из которых используется для передачи

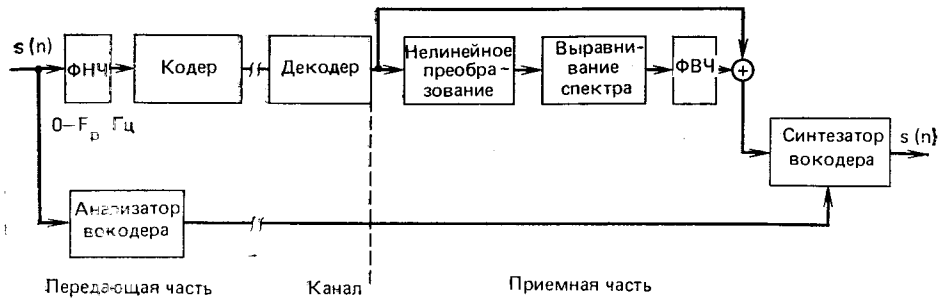


Рис. 8.29. Структурная схема вокодера, возбуждаемого речью (полувокодер)

узкополосного речевого сигнала, а другой — для передачи параметров обычного вокодера (например, коэффициентов предсказания, огибающей спектра и т. д.). Узкополосный сигнал, который при передаче можно закодировать любым из методов гл. 5, используется в синтезаторе для образования сигнала возбуждения путем соответствующих нелинейных преобразований и выравнивания спектра. Причина высокой эффективности данного метода заключается в том, что низкочастотная часть сигнала содержит всю необходимую информацию о возбуждении, т. е. она синхронна с точным периодом основного тона при периодическом возбуждении и шумоподобна в других случаях.

При использовании такого метода можно избежать оценивания основного тона и признака тон—шум. Однако в данном случае по каналу передается дополнительная информация, необходимая для описания низкочастотной части спектра, поэтому вокодеры с возбуждением речевым сигналом требуют более высоких скоростей передачи, чем обычные вокодеры. Так, например, при использовании речевого возбуждения скорость передачи составляет около 3000—4000 бит/с, т. е. на 1000—2000 бит/с больше, чем в обычных вокодерах. Выигрыш, получаемый за счет увеличения скорости передачи, сводится к меньшей зависимости качества передачи от замены диктора или изменений условий передачи. Это объясняется отсутствием устройств выделения основного тона и классификаторов тон—шум. Более детальное обсуждение вокодера, возбуждаемого речевым сигналом, можно найти в [27], [28].

8.11. Заключение

В данной главе рассматривались методы линейного предсказания речи. Основное внимание уделялось подходам, позволяющим в наибольшей мере понять процессы речеобразования. Были рассмотрены некоторые аспекты применения этих методов, а также предпринята попытка выявить там, где это возможно, сходства и различия между основными методами обработки сигналов.

Задачи

8.1. Рассмотрим разностное уравнение

$$h(n) = \sum_{k=1}^{p} \alpha_k h(n-k) + G\delta(n).$$

Автокорреляционная функция $h(n)$ определяется как

$$\tilde{R}(m) = \sum_{n=0}^{\infty} h(n)\, h(n+m).$$

а) Показать, что $\tilde{R}(m) = \tilde{R}(-m)$.
б) Подстановкой разностного уравнения в $\tilde{R}(-m)$ показать, что

$$\tilde{R}(m) = \sum_{k=1}^{p} \alpha_k \tilde{R}(|m-k|), \quad m = 1, 2, \ldots, p.$$

8.2. Системная функция $H(z)$, вычисленная в N равноотстоящих точках единичной окружности, есть

$$H\left(e^{i\frac{2\pi}{N}k}\right) = \frac{G}{1 - \sum_{n=1}^{p} \alpha_n e^{-i\frac{2\pi}{N}kn}}, \quad 0 \leqslant k \leqslant N-1.$$

Описать процедуру использования алгоритма БПФ для вычисления $H\left(e^{i\frac{2\pi}{N}k}\right)$

8.3. Уравнение (8.30) можно использовать для сокращения объема вычислений, необходимых для получения ковариационной матрицы в ковариационном методе.

а) Используя определение $\varphi_n(i, k)$ в ковариационном методе, показать, что $\varphi_n(i+1, k+1) = \varphi_n(i, k) + s_n(-i-1)s_n(-k-1) - s_n(N-1-i)s_n(N-1-k)$, считая, что $\varphi_n(i, 0)$ вычислено для $i = 0, 1, 2, \ldots, p$.

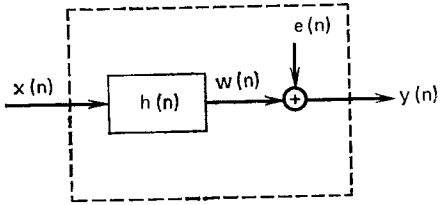


Рис. 3.8.1

б) Показать, что элементы на главной диагонали можно вычислить, основываясь на $\varphi_n(0, 0)$, т. е. получить рекурсивную формулу для $\varphi_n(i, i)$.

в) Показать, что элементы на нижних диагоналях также вычисляются рекурсивно, начиная с $\varphi_n(i, 0)$.

г) Как получить элементы на верхних диагоналях?

8.4. Линейное предсказание можно рассматривать как оптимальный метод оценивания линейной системы, основанный на ряде предположений. На рис. 3.8.1 представлен другой способ оценки параметров линейной системы. Предположим, что наблюдению доступен как $x(n)$, так и $y(n)$ и что $e(n)$ — белый гауссовский шум с нулевым средним и дисперсией σ^2_e, статистически не связанный с $x(n)$. Оценка импульсной характеристики линейной системы должна быть такой, чтобы минимизировать квадрат ошибки $\varepsilon = E[(y(n) - \hat{h}(n) * x(n))^2]$, где $\hat{h}(n)$, $0 \leqslant n \leqslant M-1$ — оценка $h(n)$.

а) Определить систему линейных уравнений относительно $\hat{h}(n)$ через автокорреляционную функцию $x(n)$ и взаимно-корреляционную функцию между $y(n)$ и $x(n)$.

б) Как решить систему уравнений, полученную в п. а)? Как соотнести полученную систему с методом линейного предсказания, рассмотренным в данной главе?

в) Получить выражение для ε-минимального среднего квадрата ошибки.

8.5. При выводе лестничного алгоритма фильтр погрешности i-го порядка определялся как

$$A^{(i)}(z) = 1 - \sum_{k=1}^{i} \alpha_k^{(i)} z^{-k}.$$

Коэффициенты предсказания удовлетворяют соотношениям (8.131). Подставляя выражения $\alpha_j^{(i)}$, $1 \leqslant j \leqslant i$, в выражение для $A^{(i)}(z)$, получить

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}).$$

8.6. Дан отрезок речевого сигнала $s(n)$, который имеет период N_p отсчетов, при этом $s(n)$ можно представить в виде дискретного преобразования Фурье

$$s(n) = \sum_{k=1}^{M} \left(\beta_k e^{i\frac{2\pi}{N_p}kn} + \beta_k^* e^{i\frac{2\pi}{N_p}kn} \right),$$

где M — число имеющихся гармоник основной частоты $(2\pi/N_p)$. С целью спектрального выравнивания сигнала (для выделения основного тона) запишем сигнал $y(n)$ в виде

$$y(n) = \sum_{k=1}^{M} \left(e^{i\frac{2\pi}{N_p}kn} + e^{-i\frac{2\pi}{N_p}kn} \right).$$

Эта задача связана с процедурой выравнивания спектра сигнала с использованием комбинированного метода, основанного на линейном предсказании и гомоморфной обработке.

а) Показать, что выравненный по спектру сигнал можно выразить в виде

$$y(n) = \frac{\sin\left[\frac{\pi}{N_p}(2M+1)n\right]}{\sin\left[\frac{\pi}{N_p}n\right]} - 1.$$

(Отметим, что эта последовательность изображена на рис. 6.20 для $N_p = 15$ и $M = 2$.)

Теперь предположим, что линейное предсказание сделано по сигналу $s(n)$ с использованием окна длиной в несколько периодов основного тона, а значение p в анализе таково, что $p = 2M$. При этом получена системная функция вида

$$H(z) = 1/(1 - \sum_{k=1}^{p} \alpha_k z^{-k}) = 1/A(z).$$

Знаменатель можно представить в виде $A(z) = \prod_{k=1}^{(p)} (1 - z_k z^{-1})$.

б) Как связаны $p = 2M$ полюсов в $A(z)$ с частотами, представленными в сигнале?

Кепстр $\hat{h}(n)$ импульсной характеристики $h(n)$ определяется как последовательность, z-преобразование которой имеет вид $\hat{H}(z)=\log H(z)=-\log A(z)$. Отметим, что $\hat{h}(n)$ можно вычислить по α_k, используя (8.123). Показать, что $\hat{h}(n)$ связано с нулями $A(z)$ соотношением

$$\hat{h}(n)=\sum_{k=1}^{p}\frac{z_k^n}{n},\quad n>0.$$

в) Используя результаты пп. а) и б), доказать, что $y(n)=n\hat{h}(n)$ является сигналом с выравненным спектром, таким, как это необходимо для выделения основного тона.

8.7. «Стандартный» метод вычисления кратковременного спектра для сегмента речевого сигнала показан на рис. 3.8.2*а*. Более сложный метод, требующий больше вычислений для получения $\log|X(e^{i\omega})|$, показан на рис. 3.8.2*б*.

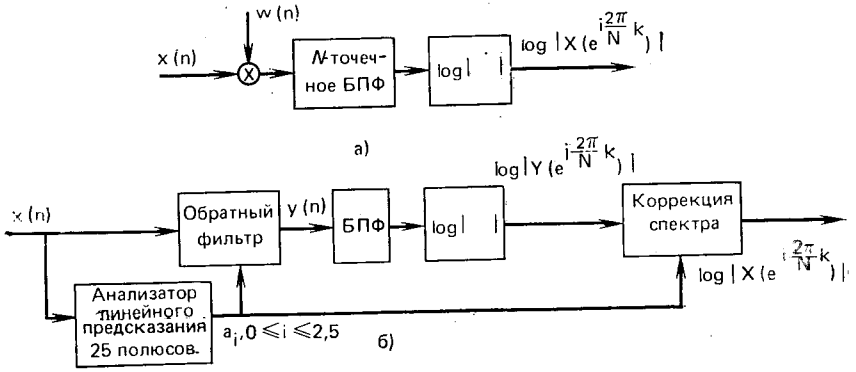


Рис. 3.8.2

а) Рассмотрите новый метод вычисления спектра и объясните функции устройства коррекции спектра.

б) В чем возможные преимущества нового метода? Рассмотрите использование окон, присутствие нулей в спектре и т. д.

8.8. Предлагается метод определения основного тона на основе линейного предсказания с использованием автокорреляционной функции погрешности предсказания $e(n)$. Вспомним, что $e(n)$ можно представить в виде

$$e(n)=\hat{s}(n)-\sum_{i=1}^{p}\alpha_i\hat{s}(n-i),$$

и если обозначить $\alpha_0=-1$, тогда

$$e(n)=-\sum_{i=0}^{p}\alpha_i\hat{s}(n-i),$$

где сигнал взвешивается с окном $\hat{s}(n)=s(n)w(n)$ и отличен от нуля на интервале $0\leqslant n\leqslant N-1$.

а) Показать, что автокорреляционная функция $e(n)$, $R_e(m)$, может быть выражена в виде

$$R_e(m)=\sum_{l=-\infty}^{\infty}R_a(l)\,R_{\hat{s}}(m-l),$$

где $R_a(l)$ — автокорреляционная функция параметров предсказания: $R_{\hat{s}}(l)$ — автокорреляционная функция сигнала $\hat{s}(n)$.

б) Определить число сложений и умножений для вычисления, если частота дискретизации равна 10 кГц, а $R_e(m)$ заключено в интервале от 3 до 15 мс?

8.9. В этой книге рассматривался ряд вокодеров: канальный последовательный формантный, параллельный формантный, гомоморфный, фазовый и вокодер на основе линейного предсказания. Чисто теоретически упорядочите эти вокодеры по качеству сигнала. Объясните подробно полученную очередность. При обсуждении следует рассмотреть вопросы используемой модели, теряемую при анализе информацию, необходимость слежения за основным тоном и т. д.

8.10. Предположим, что два диктора пытаются установить связь, используя вокодеры различного типа, как это показано на рис. 3.8.3. У диктора *1* имеются анализатор вокодера с линейным предсказанием типа рассмотренного в 8.3.10 и синтезатор в прямой форме, который описан в § 8.9. Диктор *2* располагает гомоморфным вокодером, обсуждавшимся в § 7.5.

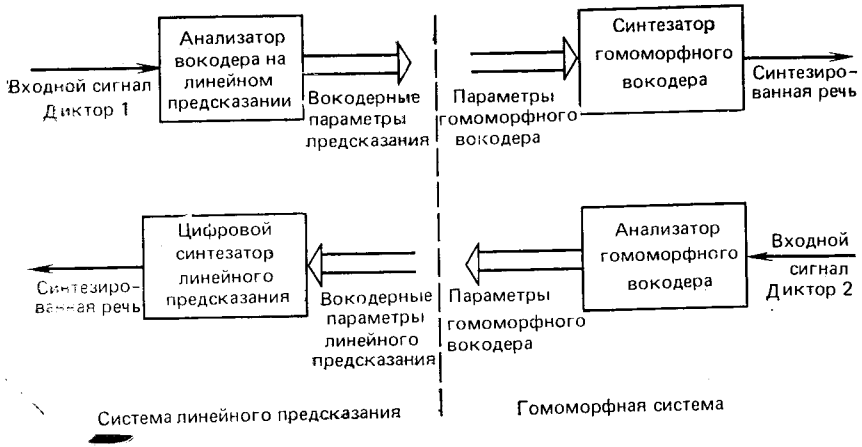


Рис. 3.8.3

а) Чтобы диктор *1* мог связаться с диктором *2*, необходимо преобразовать описание сигнала на основе линейного предсказания в гомоморфное описание для синтеза речи с помощью гомоморфного синтезатора. Придумать метод такого преобразования.

б) Придумать метод преобразования гомоморфного описания в описание на основе линейного предсказания с тем, чтобы диктор *2* мог установить связь с диктором *1*.

8.11. Рассмотрим два взвешенных сегмента речи $x(n)$ и $\hat{x}(n)$, определенные на интервале $0\leqslant n\leqslant N-1$ (вне этого интервала оба сегмента равны нулю). Осуществим анализ на основе линейного предсказания на каждом из сегментов. Таким образом, получим автокорреляционные функции, определяемые как

$$R(k)=\sum_{n=0}^{N-1-k}x(n)\,x(n+k),\quad 0\leqslant k\leqslant p;$$

$$\hat{R}(k)=\sum_{n=0}^{N-1-k}\hat{x}(n)\,\hat{x}(n+k),\quad 0\leqslant k\leqslant p.$$

На основе автокорреляционных функций найдем параметры предсказания $\boldsymbol{\alpha}=(\alpha_0,\alpha_1,\ldots,\alpha_p)$ и $\hat{\boldsymbol{\alpha}}=(\hat{\alpha}_0,\ldots,\hat{\alpha}_1,\hat{\alpha}_p)$, $(\alpha_0=\hat{\alpha}_0=1)$.

а) Показать, что погрешность предсказания

$$E^{(p)} = \sum_{n=0}^{N-1+p} e^2(n) = \sum_{n=0}^{N-1+p} \left[-\sum_{i=0}^{p} \alpha_i x(n-i) \right]^2$$

может быть записана в виде $E^{(p)} = \boldsymbol{\alpha} \mathbf{R}_{\alpha} \boldsymbol{\alpha}^t$, где $\mathbf{R}_{\alpha}$ — матрица $(p+1) \times (p+1)$. Определить $\mathbf{R}_{\alpha}$.

б) Предположим, что сигнал $\hat{x}(n)$ пропущен через обратный фильтр с коэффициентами α, что дает погрешность предсказания $\hat{e}(n)$, определяемую выражением

$$\tilde{e}(n) = -\sum_{i=0}^{p} \alpha_i \hat{x}(n-i).$$

Показать, что средняя квадратическая ошибка $\tilde{E}^{(p)}$, определяемая как $\tilde{E}^{(p)} = \sum_{n=0}^{N-1+p} [\tilde{e}(n)]^2$, может быть записана в виде $\tilde{E}^{(p)} = \mathbf{d} \underset{\alpha}{\mathbf{R}} \boldsymbol{\alpha}^t$, где $\underset{\alpha}{\mathbf{R}}$ — матрица $(p+1) \times (p+1)$. Определить $\underset{\alpha}{\mathbf{R}}$.

в) Если определить отношение $D = \tilde{E}^{(p)}/E^{(p)}$, то что можно сказать о диапазоне значения D?

8.12. Предложена следующая мера различимости между двумя сегментами речевого сигнала с параметрами предсказания $\boldsymbol{\alpha}$ и $\hat{\boldsymbol{\alpha}}$ и корреляционными матрицами $\mathbf{R}_{\alpha}$ и $\underset{\alpha}{\mathbf{R}}$ (см. задачу 8.11):

$$D(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = \frac{\boldsymbol{\alpha} \underset{\alpha}{\mathbf{R}} \boldsymbol{\alpha}^t}{\hat{\boldsymbol{\alpha}} \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t}.$$

а) Показать, что мера различимости $D(\alpha, \hat{\alpha})$ может быть записана в следующей удобной для вычислений форме:

$$D(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = \left[\frac{\left(b(0) \hat{R}(0) + 2 \sum_{i=1}^{p} b(i) \hat{R}(i) \right)}{\hat{\boldsymbol{\alpha}} \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t} \right],$$

где $b(i)$ — автокорреляционная функция вектора $\boldsymbol{\alpha}$ — равна:

$$b(i) = \sum_{j=0}^{p-i} \alpha_j \alpha_{j+i}, \quad 0 \leqslant i \leqslant p.$$

б) Предположим, что величины (вектора, матрицы, скаляры) $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, $\mathbf{R}$, $\hat{\mathbf{R}}$, $(\alpha \underset{\alpha}{\mathbf{R}} \hat{\boldsymbol{\alpha}}^t)$, $\underset{\alpha}{\mathbf{R}}$ и $\mathbf{b}$ вычислены заранее, т. е. известны к моменту расчета меры расстояния. Сравнить объем вычислений, необходимый для определения $D(\alpha, \hat{\alpha})$, используя оба выражения для D, рассмотренные в данной задаче.

9

Цифровая обработка речи в системах речевого общения человека с машиной[1]

9.0. Введение

В предыдущих главах внимание было сконцентрировано на основных теоретических вопросах, необходимых для понимания современных методов цифровой обработки речевых сигналов. Еще не рассматривалась обширная область применения разработанных методов, т. е. способов использования моделей и связанных с ними параметров в системах передачи или автоматического выделения информации из сигнала речи. В данной главе приведены характерные примеры цифровой обработки речи применительно к системам общения между человеком и машиной (ЭВМ) посредством голоса. Существует ряд причин, по которым имеет смысл ограничиться рассмотрением примеров связи между человеком и машиной. Прежде всего, эта область наиболее плодотворна с точки зрения возможностей использования методов цифровой обработки речи и позволяет, таким образом, проиллюстрировать почти все рассмотренные выше методы обработки. Кроме того, эта область является чрезвычайно важной, дающей все новые и новые приложения, область, которая только еще развивается и демонстрирует огромные возможности для широкого применения.

Системы речевого обмена между человеком и машиной можно подразделить на три класса: с речевым ответом, распознавания диктора и распознавания речи.

Системы с речевым ответом предназначаются для выдачи информации пользователю в форме речевого сообщения. Таким образом, системы с речевым ответом — это системы односторонней связи, т. е. от машины к человеку. С другой стороны, системы второго и третьего классов — это системы связи от человека к машине. В системах распознавания диктора задача состоит в верификации диктора (т. е. в решении задачи о принадлежности данного диктора к некоторой группе лиц) или идентификации диктора из некоторого известного множества. Таким образом, класс задач распознавания диктора распадается на два подкласса: верификации и идентификации говорящего. Различия и сходство между этими задачами будут рассмотрены в последующем.

Последний класс задач распознавания речи также можно разделить на подклассы в зависимости от таких факторов, как размер словаря, количество дикторов, условия произнесения слов и т. д. Основная задача распознающей системы сводится либо к точному распознаванию произнесенной на входе фразы (т. е. система фонетической или орфографической печати произнесенного текста), либо к «пониманию» произнесенной фразы (т. е. к правильной реакции на сказанное диктором). Именно задача понимания, а не распознавания наиболее важна для систем с достаточно большим словарем непрерывных речевых сигналов, в то время как задача точного распознавания более важна для систем с ограниченным словарем, малым количеством дикторов, систем распознавания изолированных слов. Различные аспекты построения систем распознавания речи также рассматриваются в данной главе.

В заключительной части главы рассмотрены некоторые системы, типичные для каждой области речевого общения человека с машиной. Более подробно рассматриваются особенности обработки речевого сигнала с целью закрепления результатов предшествующих глав. Однако как для полноты обсуждения, так и для более глубокого понимания здесь излагаются и общие аспекты обработки информации в системе, поскольку часто они оказываются достаточно важными для успешной работы системы в целом.

[1] Имеются в виду цифровые ЭВМ. (Прим. ред.)