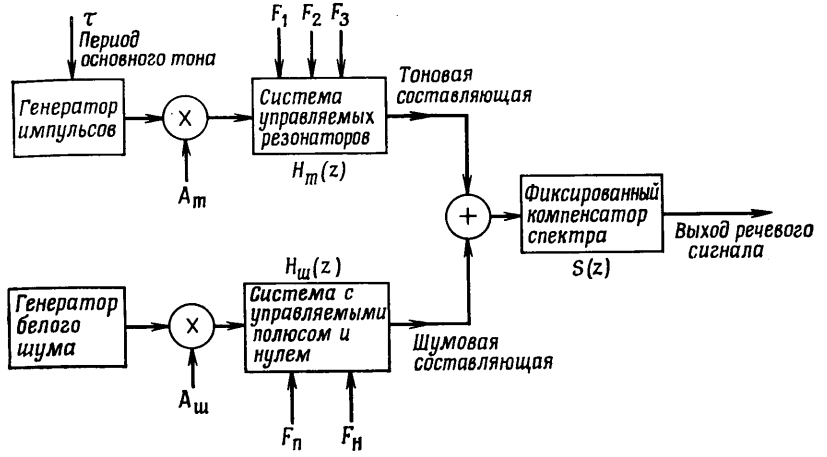
Такая обратная система состоит из нелинейности с экспоненциальной характеристикой (компенсирующей логарифмирование) и блока, выполняющего обратное ДПФ (компенсирующего ДПФ). На выходе системы получается $u(n)$ — оценка импульсной характеристики голосового тракта. Период возбуждения (полученный из анализа кепстра) используется для формирования либо квазипериодической, либо случайной последовательности импульсов $\hat{s}(n)$, заменяющей истинный сигнал возбуждения $s(n)$. Для получения синтезированной речи образуют свертку последовательностей $\hat{s}(n)$ и $\hat{u}(n)$. Сравнение на фиг. 12.35 спектрограмм (по Оппенгейму) исходного высказывания и синтезированного, полученного при гомоморфной обработке, показывает весьма близкое их сходство.

12.15. Формантный синтез

В исследованиях речи одной из наиболее важных задач является синтез речевого сигнала на основе некоторых параметров сигнала возбуждения. Синтез речи применяется в нескольких видах систем речевого общения ЭВМ с человеком, и знакомство с ними существенно проясняет основные механизмы образования и восприятия речи. Одним из основных наборов упомянутых выше параметров является набор значений частот основных формант, заданных в функции времени. Ниже будет показано, каким образом такое представление речи обеспечивает значительную гибкость и эффективность в разнообразных применениях искусственной речи. В настоящем разделе рассмотрены некоторые задачи обработки сигналов, связанные с синтезом речи по данным о ее формантах. Предполагается, что для получения этих данных из реализаций естественной речи используется система анализа, подобная, например, рассмотренной в разд. 12.14.

Блок-схема универсального формантного синтезатора (фиг. 12.36), аналогичного применяемому в нескольких системах речевого общения ЭВМ с человеком, содержит два источника возбуждения: генератор импульсов с внешней синхронизацией (источник звонких звуков), вырабатывающий единичные импульсы с частотой основного тона (т. е. через каждые P отсчетов), и генератор псевдослучайных чисел с равномерным распределением (источник глухих звуков), играющий роль генератора белого шума.

В синтезаторе имеются две основные ветви обработки сигналов. Верхняя состоит из амплитудного модулятора ($A_т$) и цифрового фильтра с переменными параметрами, образованного цепочкой из L перестраиваемых резонаторов (полюсов). Передаточная

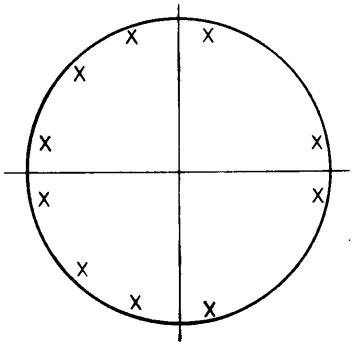


Фиг. 12.36. Упрощенная блок-схема формантного синтезатора.

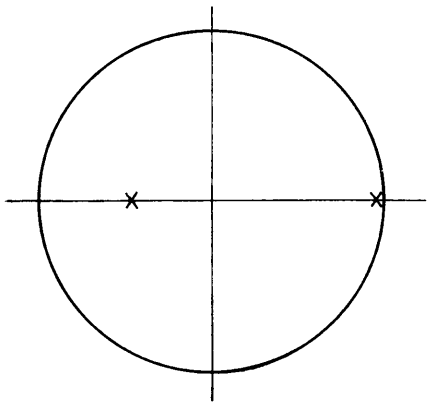
функция этого фильтра (в стационарном режиме) равна

$$H_т(z) = \prod_{k=1}^{L} \left[\frac{1 - \exp(-\alpha_k T)\, 2\cos(b_k T) + \exp(-2\alpha_k T)}{1 - \exp(-\alpha_k T)\, 2\cos(b_k T)\, z^{-1} + \exp(-2\alpha_k T) z^{-2}} \right], \quad (12.39)$$

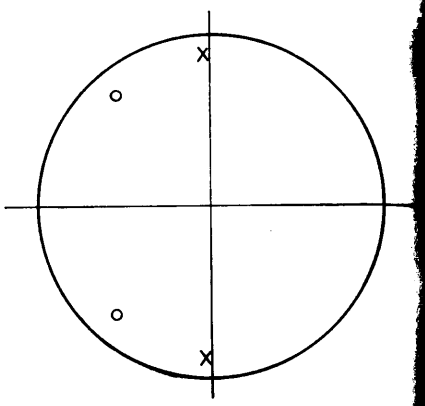
где α_k и b_k — ширина полосы и центральная частота k-го резонатора в радианах, а T — период дискретизации. Типичная схема расположения полюсов в z-плоскости для гласной ($L = 5$) изображена на фиг. 12.37. Хотя управлять можно и шириной полосы, и центральными частотами всех полюсов, обычно подстраивают только три нижние центральные частоты. Поэтому блок перестраиваемых резонаторов (фиг. 12.36) имеет три управляющих входа (F_1, F_2, F_3). Эта управляемая резонансная система позволяет



Фиг. 12.37. Расположение полюсов для типичной гласной



Фиг. 12.38. Расположение полюсов, описывающих функцию возбуждения.



Фиг. 12.39. Расположение нулей и полюсов для типичного шумового звука.

учесть влияние временно́го изменения формы голосового тракта на спектр речевого сигнала.

Следует также учесть форму импульсов возбуждения и характеристики излучения звука изо рта (или носа) в воздух. Для этого предназначена неперестраиваемая схема компенсации с передаточной функцией вида

$$S(z) = \frac{[1-\exp(-\alpha T)]\,[1+\exp(-bT)]}{[1-\exp(-\alpha T)\,z^{-1}]\,[1+\exp(-bT)\,z^{-1}]}. \qquad (12.40)$$

Схема реализует два полюса, расположенных на действительной оси (один в правой, а другой в левой половине z-плоскости), и аппроксимирует выбранную передаточную функцию. Положение полюсов в z-плоскости показано на фиг. 12.38.

Нижняя ветвь схемы синтезатора (фиг. 12.36) состоит из модулятора $A_{\text{ш}}$, регулирующего дисперсию шума, и второго цифрового фильтра с переменными параметрами, образованного последовательно соединенными блоками с нулем и полюсом. Передаточная функция фильтра равна

$$H_{\text{ш}}(z) = \frac{H_1(1)\,H_2(z)}{H_1(z)\,H_2(1)}, \qquad (12.41)$$

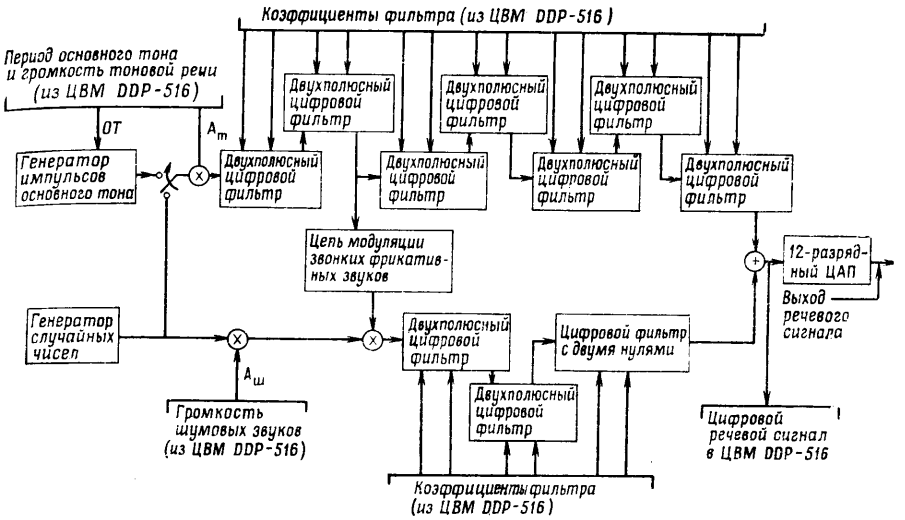
где

$$H_1(z) = 1 - 2e^{-aT}\cos(bT)\,z^{-1} + e^{-2aT}z^{-2}$$

и

$$H_2(z) = 1 - 2e^{-cT}\cos(dT)\,z^{-1} + e^{-2cT}z^{-2}.$$

Здесь a, b, c и d — значения ширины полосы и центральных частот блоков с перестраиваемыми полюсом и нулем, измеренные в радианах. Ширину полос обычно не изменяют, а регулируют



Фиг. 12.40. Блок-схема аппаратурной части синтезатора.

только центральные частоты, поэтому фильтр (фиг. 12.36) имеет два управляющих входа $F_{\text{п}}$ и $F_{\text{н}}$. Типичное для глухого звука расположение нулей и полюсов показано на фиг. 12.39. Выходное колебание проходит через фильтр компенсации спектра и создает на выходе всей системы глухой звук.

Следует отметить, что передаточные функции (12.39)—(12.41) всех фильтров синтезатора на нулевой частоте равны единице независимо от значений ширины полосы и центральной частоты управляемых блоков. Это необходимо для того, чтобы коэффициент передачи голосового тракта на нулевой частоте равнялся единице, что достигается за счет использования отдельно откалиброванных резонаторов.

Рассмотренная схема синтезатора не позволяет получить некоторые звуки, желательные в многоцелевом синтезаторе. Например, в нем нет средств для получения носовых согласных звуков n и m, звонких фрикативных звуков z (как в слове zoo), zh (azure), v (very) и th (there). При синтезе носовых согласных последовательно с перестраиваемым резонатором (фиг. 12.36) следует включить цепь с управляемыми нулем и полюсом. Для качественного синтеза звонких фрикативных звуков необходимо ввести цепь, модулирующую выход генератора шума сигналом из канала тоновых (звонких) звуков. Кроме того, для расширения возможностей синтезатора следует ввести цепи, позволяющие для имитации шепота возбуждать канал тоновых звуков шумовым сигналом.

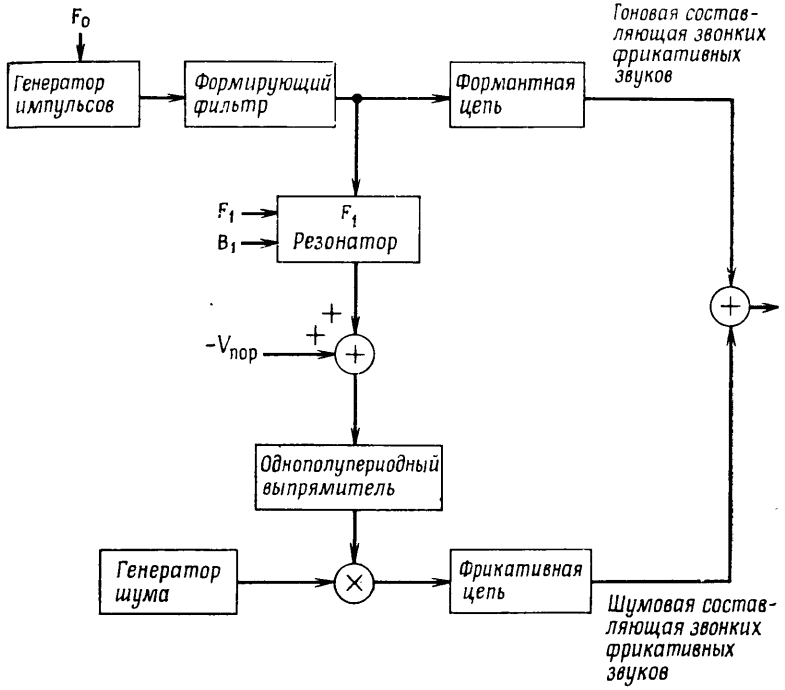
Существует более гибкая схема синтезатора (фиг. 12.40), решающая эти задачи. Она была промоделирована на ЦВМ, а также построена в виде специализированного устройства. Синтезатор получает переменные управляющие параметры (обозначенные как внешние входы в каждый из блоков обработки сигнала) синхронно, т. е. изменяет сразу все параметры в начале каждого периода основного тона. В этот момент энергия, запасенная в каждом из фильтров, минимальна, что уменьшает нежелательные эффекты, вызываемые резкими изменениями управляющих параметров. Управляющие параметры поступают в синтезатор из управляющей ЦВМ типа Honeywell DDP-516.

Рассматриваемый синтезатор в сущности аналогичен рассмотренному выше, хотя и отличается от него в деталях. В частности, верхний канал обработки сигналов содержит шесть цифровых фильтров с двумя полюсами каждый [в формуле (12.39) $L = 6$] и один фильтр с двумя нулями, причем полосы и центральные частоты каждого фильтра перестраиваются. Шестой двухполюсный фильтр и фильтр с двумя нулями введены для образования носовых звуков. При синтезе неносовых звуков они компенсируют друг друга (в цифровых системах легко достичь точной компенсации полюса нулем). Четыре двухполюсных фильтра (или пять при неносовых звуках) формируют изменяющуюся во времени передаточную функцию голосового тракта $H_{\text{т}}(z)$, а последний фильтр с двумя полюсами обеспечивает желаемую функцию компенсации спектра $S(z)$.

Канал глухих звуков состоит из двух двухполюсных фильтров и одного фильтра с двумя нулями. Полосы и центральные частоты каждого из них также устанавливаются извне. Два разнотипных фильтра формируют $H_{\text{ш}}(z)$, а второй двухполюсный фильтр задает $S(z)$ и используется для компенсации спектра. Гибкость синтезатора увеличивается также за счет того, что функции компенсации спектра при синтезе звонких и глухих звуков могут отличаться, так как соответствующие цепи включены в разные каналы синтезатора независимо.

12.16. Цепь возбуждения звонких фрикативных звуков

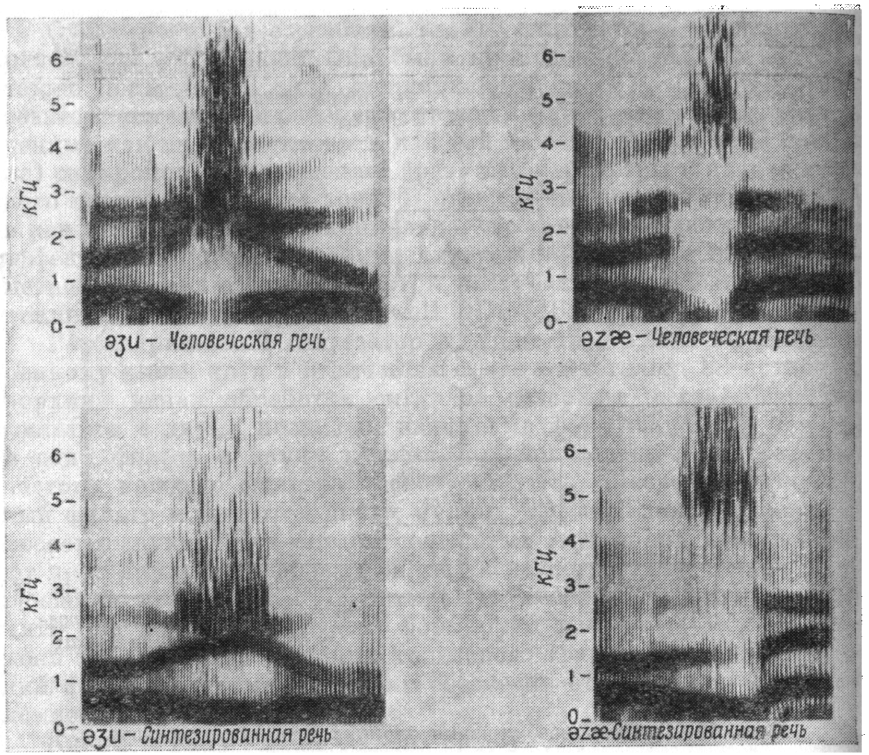
Цепь возбуждения звонких фрикативных звуков соединяет некоторую точку канала тоновых звуков с каналом шумовых (глухих) звуков. Она используется для моделирования образования глухой составляющей звонких фрикативных звуков. Основные цепи, необходимые для синтеза звонкого фрикативного звука, изображены на фиг. 12.41. Глухой сигнал возбуждения формируется следующим образом. Импульсы основного тона возбуждают резонатор, настроенный на частоту первой форманты звонкой составляющей фрикативного звука. Характеристика



Фиг. 12.41. Цепь возбуждения звонких фрикативных звуков.

резонатора в первом приближении аппроксимирует спектр объемной скорости воздушной струи на участке от голосовой щели до точки сужения голосового тракта. Из выходного сигнала резонатора вычитается пороговый уровень ($V_{\text{пор}}$), и результат выпрямляется однополупериодным выпрямителем. Эти операции моделируют известное физическое явление, состоящее в том, что турбулентность не возникает, пока объемная скорость струи не превысит порогового значения.

Выпрямленное колебание модулирует шум, поступающий с генератора шума. При этом образуется сигнал возбуждения глухой составляющей фрикативного звука, синхронный с основным тоном. Полученный сигнал проходит через фрикативную цепь (в нижнем канале синтезатора), что дает глухую составляющую звука. Звонкая составляющая образуется при возбуждении формантной цепи обычным образом. Спектрограммы синтезированного и естественного звонких фрикативных звуков $|zh|$ и $|z|$ приведены на фиг. 12.42. Тщательный анализ спектрограмм позволяет заметить как в естественной, так и в синтезированной речи эффекты модуляции, синхронной с основным тоном.



Фиг. 12.42. Спектрограмма синтезированного и естественного звонких фрикативных звуков.

12.17. Генератор случайных чисел

Для образования псевдослучайных чисел, служащих источником возбуждения глухих звуков, можно применить любой из множества существующих алгоритмов. При построении специализированной системы для генерации случайных чисел используется 16-разрядная последовательность максимальной длины на сдвиговом регистре. В данном алгоритме очередной случайный двоичный разряд образуется сложением по модулю 2 всех 16 предыдущих разрядов, после чего все разряды сдвигаются на один. При этом разряд, образованный 16 тактов назад, теряется, а новый вводится в сдвиговый регистр.

Формула образования нового разряда имеет вид

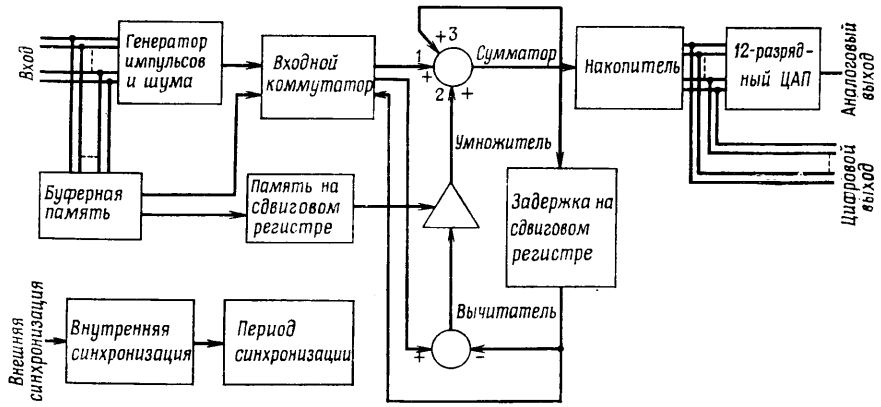
$$X_n = X_{n-1} \oplus X_{n-2} \oplus \ldots \oplus X_{n-15} \oplus X_{n-16}, \quad n = 1, 2, 3, \ldots, \quad (12.42)$$

где X равно 0 или 1, причем единице физически соответствует положительный возбуждающий импульс, а нулю — отрицательный. Таким образом, выходная последовательность является случайной последовательностью положительных и отрицательных импульсов с гладким спектром.

12.18. Цифровая обработка в формантном синтезаторе

Основным принципом, используемым при построении цифровых синтезаторов, является разделение времени, т. е. реализация всех фильтров с двумя полюсами или нулями с помощью одного арифметического устройства. Для получения очередного отсчета на выходе двухполюсного фильтра, например, требуется выполнить два сложения, два вычитания и два умножения. Современные быстродействующие интегральные схемы позволяют выполнить за время между отсчетами (100 мкс при частоте выборок 10 кГц) в 25 раз больше арифметических действий. Поэтому идея разделения времени применительно к синтезаторам оказывается вполне реальной. Если ввести память для коэффициентов фильтров и выходных отсчетов, то при соответствующем управлении одно арифметическое устройство может обслужить весь синтезатор.

В логической блок-схеме цифрового синтезатора (фиг. 12.43) арифметическое устройство состоит из трехвходового сумматора, блока задержки на регистре сдвига (для запоминания промежуточных результатов), схемы вычитания и умножителя. Емкость регистра сдвига равна 480 двоичным разрядам (20 чисел по 24 разряда в каждом). Еще один регистр сдвига на 320 разрядов используется для запоминания коэффициентов цифровых фильтров (20 коэффициентов по 16 разрядов в каждом). Данное арифмети-



Фиг. 12.43. Логическая блок-схема аппаратурной части синтезатора.

ческое устройство позволяет за интервал порядка 3,9 мкс выполнить одновременно сложение, вычитание и умножение. Следовательно, выполнение операций, соответствующих одному фильтру, занимает 7,8 мкс, а всем 10 фильтрам — около 78 мкс. Поэтому синтезатор может работать при частотах отсчетов вплоть до 12,8 кГц.

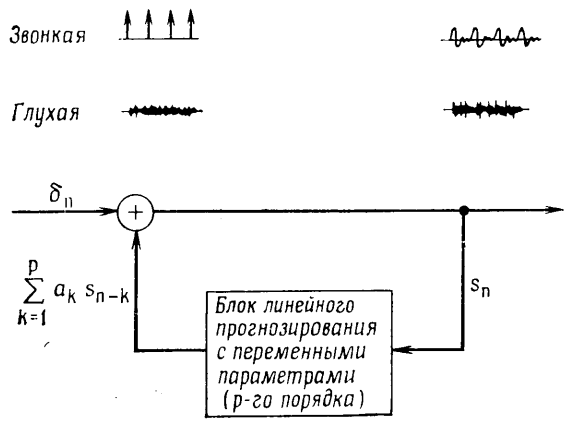
Остальные элементы блок-схемы работают следующим образом. Сигналы управления синтезатором поступают на его вход из вычислительной машины. Вспомогательная память передает значения периода основного тона и коэффициентов усиления в регистр сдвига, в генераторы импульсов и шума, а также на вход устройства многоканального уплотнения. Генераторы импульсов и шума вырабатывают сигнал возбуждения, который через устройство уплотнения поступает в арифметическое устройство. Накопитель складывает звонкую и глухую составляющие и возвращает старшие 16 разрядов в вычислительную машину, преобразуя одновременно 12 старших разрядов в аналоговую форму. Переключения и временна́я последовательность выполнения операций синхронизированы извне. Таким образом, частота отсчетов легко изменяется без каких-либо регулировок в самом синтезаторе.

12.19. Линейное прогнозирование речи

Формантные анализ и синтез основаны на том, что получение речи хорошо моделируется возбуждением цепочки цифровых линейных фильтров второго порядка с переменными параметрами (формантных резонаторов) с помощью квазипериодической последовательности импульсов или шумового сигнала. При этом основная трудность заключается во введении найденных формант в соответствующие блоки второго порядка. При синтезе одних звуков форманты, по-видимому, исчезают. При синтезе других звуков, наоборот, могут возникать дополнительные форманты. При большом количестве перечисленных ошибок синтезированная речь быстро становится неразборчивой или, в лучшем случае, имеет недопустимо низкое качество. В длинных фразах такие ошибки нередки.

Чтобы устранить эти трудности, основную модель образования речи следует несколько изменить (фиг. 12.44). L отдельных систем второго порядка формантной модели объединяют в одну линейную систему p-го порядка (где $p \geqslant 2L$). В ней задаются одновременно передаточная функция голосового тракта, форма возбуждающих импульсов и характеристики излучения звуков. На вход системы поступает или последовательность единичных отсчетов, или квазислучайная последовательность $\delta(n)$. Передаточная функция фильтра имеет вид

$$H(z) = \frac{X(z)}{\delta(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (12.43)$$



Фиг. 12.44. Модель формирования речи с помощью линейного прогнозирования (по Аталу и Ханауэру).

Выделение периода основного тона и обнаружение тон—шум осуществляются, как в любой другой системе, с помощью рассмотренного выше измерителя основного тона или каким-либо другим методом. Коэффициенты прогнозирующего фильтра $\{a_k, k = 1, 2, \ldots, p\}$ определяются методом наименьших квадратов. Разностное уравнение, описывающее систему, имеет вид

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + \delta(n). \qquad (12.44)$$

Для звонких звуков все отсчеты $\delta(n)$, за исключением тех, с которых начинаются периоды основного тона, равны нулю. Поэтому везде, кроме этих ненулевых точек,

$$s(n) = \sum_{k=1}^{p} a_k s(n-k). \qquad (12.45)$$

Итак, в принципе, если модель является верной, отсчеты речи $s(n)$ можно в точности предсказать, используя равенство (12.45). Однако модель описывает речь не полностью, поэтому можно определить ошибку между $s(n)$, истинным значением n-го отсчета, и $\hat{s}(n)$, его значением, предсказанным с помощью равенства (12.45). Пусть $E(n)$ — ошибка, т. е.

$$E(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k). \qquad (12.46)$$

Коэффициенты прогнозирующего фильтра выбирают так, чтобы обеспечить минимум среднеквадратической ошибки предсказания $\langle E^2(n) \rangle$, усредненной по всем n.