Выражение для среднеквадратической ошибки можно записать в виде

$$\langle E^2(n)\rangle = \sum_{n=1}^{\infty} [s(n) - \sum_{k=1}^{p} a_k s(n-k)]^2. \qquad (12.47)$$

Чтобы определить коэффициенты прогнозирующего фильтра, продифференцируем правую часть суммы (12.47) по a_j, $j = 1, 2, \ldots, p$, и, приравняв производные нулю, получим систему уравнений

$$\sum_{k=1}^{p} a_k \sum_{n=1}^{\infty} s(n-k)\, s(n-j) = \sum_{n=1}^{\infty} s(n)\, s(n-j), \quad j = 1, 2, \ldots, p. \qquad (12.48)$$

В матричной форме она записывается следующим образом:

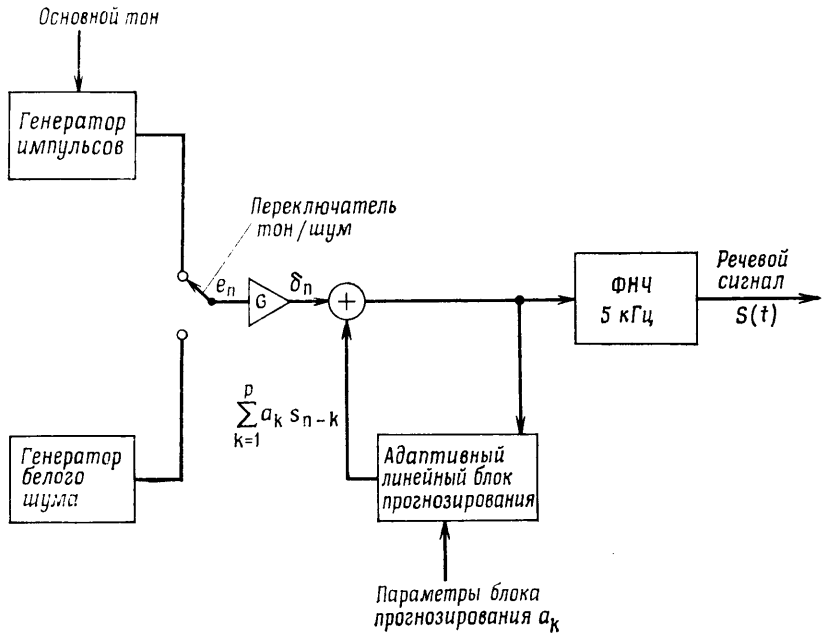
$$\Phi \underline{a} = \psi, \qquad (12.49)$$

где

$$\varphi_{ij} = \sum_{n=1}^{\infty} s(n-i)\, s(n-j), \qquad (12.50)$$

причем

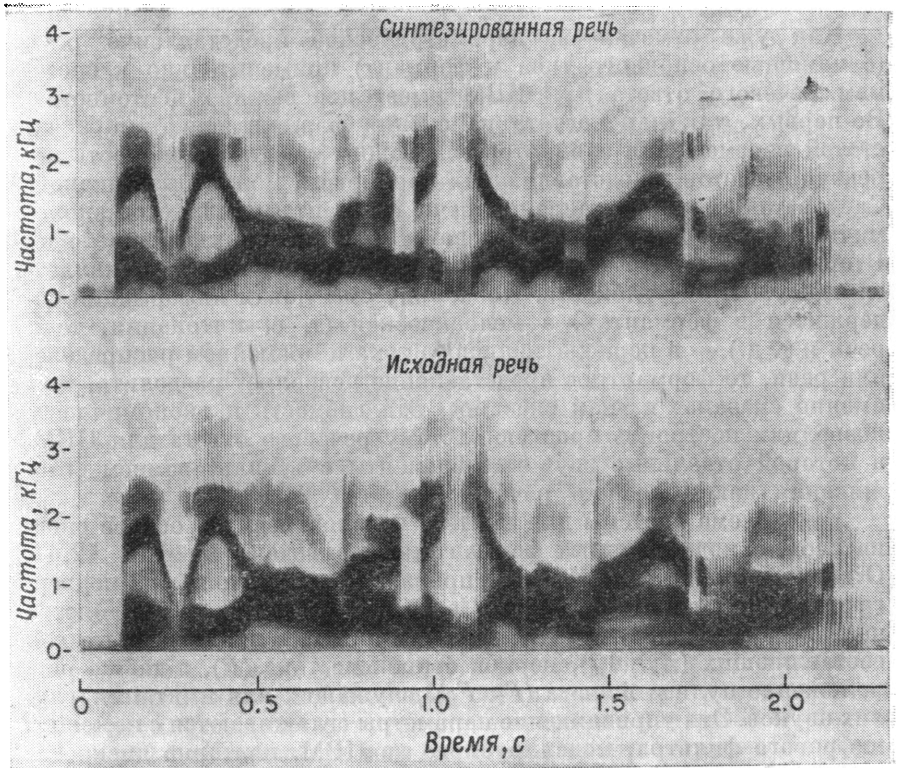
$$\psi_j = \varphi_{0j}. \qquad (12.51)$$



Фиг. 12.45. Синтезатор речи с линейным прогнозированием (по Аталу и и Ханауэру).

Таким образом, Ф является автокорреляционной матрицей, а ψ — вектором автокорреляции. Поскольку матрица Ф симметричная и положительно определенная, то для решения системы (12.48) можно применить известные эффективные методы. Поэтому анализ при линейном прогнозировании достаточно прост.

Для высококачественного представления естественного речевого сигнала используют систему синтеза, схема которой изображена на фиг. 12.45. Рассмотрим ее отличия от схемы формантного синтезатора, описанного в предыдущем разделе. Наиболее важное из них состоит в использовании единственного рекурсивного фильтра p-го порядка вместо цепочки фильтров второго порядка. При стационарном речевом сигнале, например при продолжительном звучании гласной, обе схемы полностью эквивалентны. В случае нестационарного сигнала (т. е. в большей части речи) они не эквивалентны. Для формантного синтезатора важно, что-



Фиг. 12.46. Сравнение спектрограмм естественного и синтезированного высказываний (по Аталу и Ханауэру).

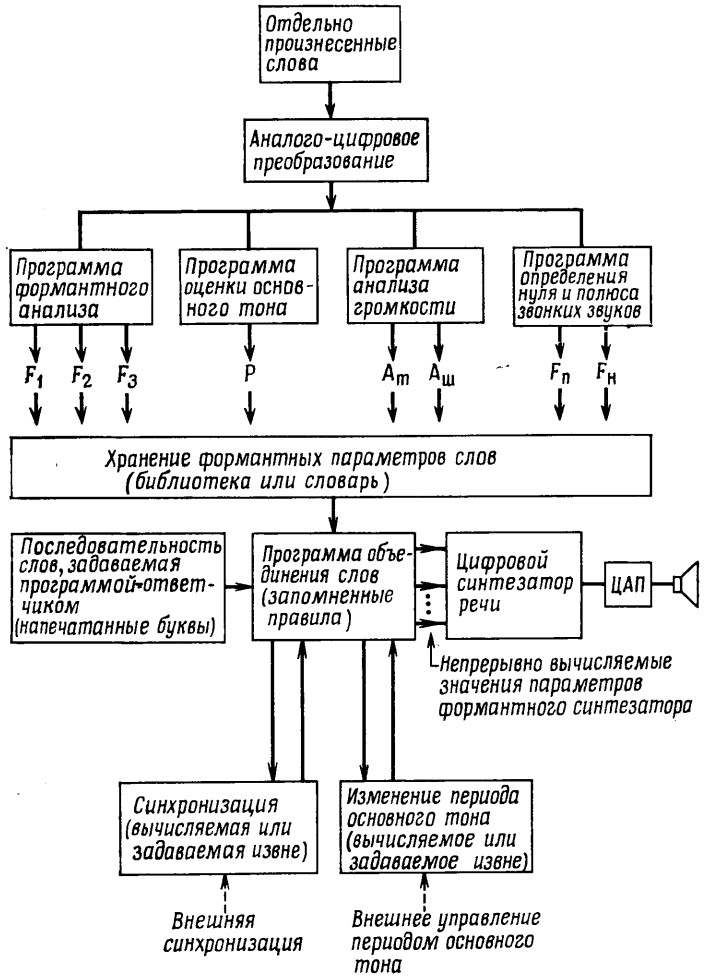
бы каждый из резонаторов соответствовал своей форманте, так как иначе синтезатор сбивается. Этого не требуется при линейном прогнозировании, так как все форманты синтезируются одним рекурсивным фильтром. Другое важное отличие состоит в том, что амплитуда импульсов основного тона, как и амплитуда белого шума, подстраивается с помощью усилителя G таким образом, чтобы получить нужное среднеквадратическое значение отсчетов синтезированной речи. Обычно такой подстройки при синтезе речи не производят.

Спектрограммы на фиг. 12.46 показывают, насколько хорошо действует система с линейным прогнозированием. Вверху приведена спектрограмма синтезированной речи, а внизу — естественной. Различить их весьма трудно.

12.20. Система речевого ответа для вычислительной машины

Как уже отмечалось, параметрическое представление речи (с помощью основного тона и формант) применительно к системам речевого ответа для ЦВМ имеет два важных достоинства. Во-первых, так как частота изменения формант сравнима с частотой перемещения элементов голосового тракта, то скорость передачи отсчетов, представляющих параметры, может быть низкой. Следовательно, представление речи формантами является экономичным способом хранения речевой информации в цифровом виде. Вторым преимуществом формантного представления речи является присущая ему гибкость. Поскольку смысловая информация содержится в формантах, а мелодическая (т. е. интонация, темп речи и т. д.) — в периоде основного тона и временно́м распределении речи, то формантное представление позволяет разделить, «что именно сказано» и «как сказано». Эта гибкость и экономичность позволили построить простую систему речевого ответа для ЦВМ, в которой отдельные звуковые элементы с использованием сглаживания дают связную речь.

Блок-схема системы для синтеза связной речи на основе списка слов, закодированных формантами, приведена на фиг. 12.47. Отдельные слова (или фразы), произнесенные человеком, подвергаются формантному анализу. Через каждые 10 мс определяются частоты трех формант (F_1, F_2, F_3), амплитуды звонкой и глухой составляющих ($A_т$, $A_ш$), период основного тона (P), а также расположение нуля и полюса ($F_т$, $F_п$), служащих для имитации глухих звуков. Эти управляющие параметры сглаживаются с помощью цифрового фильтра, моделируемого на ЦВМ, повторно дискретизуются с частотой, определяемой теоремой о дискретизации сигналов с ограниченным спектром (обычно 33 1/3 Гц), квантуются и заносятся в память, образуя справочную библиотеку. Типичная



Фиг. 12.47. Блок-схема системы речевого ответа.

скорость заполнения памяти для хранения управляющих параметров составляет 700 бит/с, если значения периода основного тона сохраняются. Однако чаще всего значения периода не запоминаются, а вычисляются с помощью специальной программы составления речи. Тогда скорость заполнения памяти равна 533 1/3 бит/с. В табл. 12.1 показано, из чего складывается эта скорость. Данные, приведенные в таблице, были получены путем экспериментального исследования влияния сглаживания и квантования на восприятие синтезированной речи.

Таблица 12.1

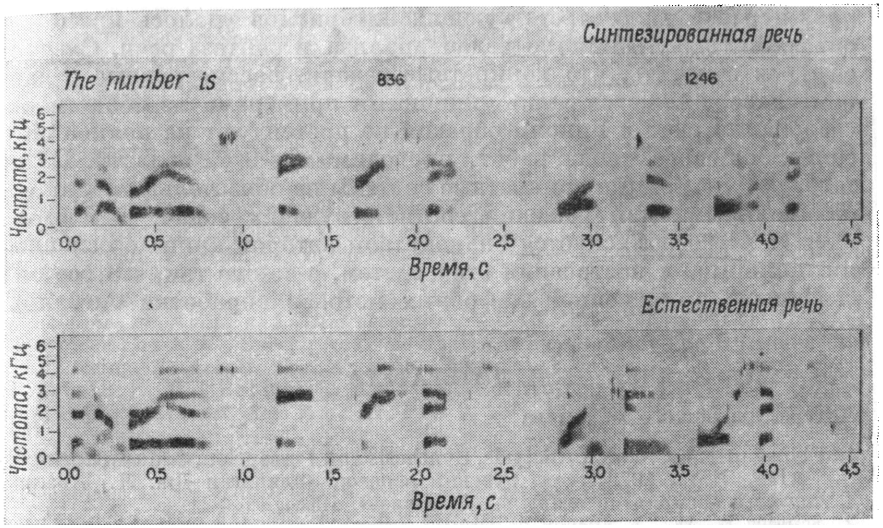
Кодирование формантных параметров

Параметр	Число бит на параметр	Частота дискретизации параметров, Гц	Скорость заполнения памяти, бит/с
F_1 или $F_п$	3	33 1/3	100
F_2 или $F_н$	4	33 1/3	133 1/3
F_3	3	33 1/3	100
P	5	33 1/3	166 2/3
$A_т$ или $A_ш$	3	33 1/3	100
V/U	1	100	100
		Всего	700
		Для периода основного тона	166 2/3
Скорость заполнения памяти при синтезе с вычислением периода основного тона			533 1/3

Как указано в табл. 12.1, каждые 10 мс определяется, какой произносится звук: звонкий или глухой и результат V/U представляется одноразрядным двоичным числом. Поэтому каждый из заносимых в память наборов параметров может быть отнесен либо только к звонкому, либо к глухому звуку. Следует отметить, что частота поступления наборов управляющих параметров (33 1/3 Гц) втрое ниже частоты следования отсчетов сигнала V/U.

Слова и фразы, представленные формантами, легко приспособить для использования в программе синтеза речи. Слова можно удлинить или укоротить; форманты легко изменить; можно ввести закон изменения основного тона, отличающийся от исходного. Таким образом, характеристики речевого тракта представлены в форме, достаточно гибкой для согласования с временной синхронизацией и высотой основного тона, задаваемыми программой составления речи.

В нижней части фиг. 12.47 показано, каким образом система составляет синтезированное сообщение, сочетая слова и фразы из справочной библиотеки. Во-первых, программа-ответчик для составления каждого конкретного ответа запрашивает последовательность слов. С помощью программы составления речи получаются (с использованием вспомогательной программы) данные о распределении времени в ответной фразе в виде значений продолжительности каждого слова, а затем последовательно извлекаются параметры слов. Слова корректируются так, чтобы их длина соответствовала выбранной длине слов. После этого осуществляются сглаживание и интерполяция значений формантных параметров, если конец некоторого слова и начало следующего содержат звонкие звуки. Для этого используется алгоритм интерполяции, имитирующий переход формант от слова к слову в естественной речи. Наконец, для всего ответа получается закон изменения частоты основного тона. Все вычисленные параметры передаются в специализированный цифровой синтезатор речи. Непрерывный синтезированный речевой сигнал получается с помощью цифро-аналогового преобразователя.



Фиг. 12.48. Спектрограммы телефонного номера, произнесенного человеком и синтезатором речи.

Описанная система речевого ответа была использована для голосового воспроизведения телефонных номеров и соединений по устному запросу с помощью вычислительной машины. Из сравнения спектрограмм типичного телефонного номера, произнесенного человеком и машиной (фиг. 12.48), видно, что моменты произнесения звуков и значения формант довольно хорошо согласуются.

12.21. Заключение

Материалы настоящей главы позволяют проследить, насколько тесно цифровая обработка сигналов и исследования речевых сигналов связаны друг с другом. Фактически только после появления вычислительных машин, предназначенных для обработки сигна-

лов, и развития соответствующих алгоритмов удалось практически решить большинство задач анализа и синтеза речи. Следует еще раз отметить, что конкретные задачи, рассмотренные в данной главе, являются лишь типичными примерами из области исследования речи и никоим образом не претендуют на полное описание большого числа работ, проводимых в этой области. Более того, рассмотренные конкретные системы не обязательно являются оптимальными для решения соответствующих задач. Их выбор прежде всего объясняется знакомством авторов книги с основными принципами построения этих систем, а также тем, что все они связаны с применением цифровых методов обработки сигналов.

ЛИТЕРАТУРА

Литература общего характера

1. Flanagan J. L., Speech Analysis, Synthesis and Perception, 2nd ed., Springer-Verlag, N.Y., 1972; есть русский перевод: Фланаган Дж. Л., Анализ, синтез и восприятие речи, изд-во «Связь», 1968.
2. Schafer R. W., A Survey of Digital Speech Processing Techniques, *IEEE Trans. on Audio and Electroacoustics*, **AU-20**, No. 4, 28—35 (March 1972).
3. Flanagan J. L., Coker C. H., Rabiner L. R., Schafer R. W., Umeda N., Synthetic Voices for Computers, *IEEE Spectrum*, **7**, No. 10, 22—45 (1970).

Кратковременный спектральный анализ

1. Flanagan J. L., Golden R. M., Phase Vocoder, *Bell Syst. Tech. J.*, **45**, 1493—1509 (1966).
2. Schafer R. W., Rabiner L. R., Design of Digital Filter Banks for Speech Analysis, *Bell Syst. Tech. J.*, **50**, No. 10, 3097—3115 (Dec. 1971).
3. Schafer R. W., Rabiner L. R., Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis, *IEEE Trans. on Audio and Electroacoustics*, **AU-21**, 165—174 (June 1973).

Полосные вокодеры

1. Schroeder M. R., Vocoders: Analysis and Synthesis of Speech, *Proc. IEEE*, **54**, 720—734 (1966); есть русский перевод: Шредер, Вокодеры: анализ и синтез речи, *ТИИЭР*, т. 54, № 5, стр. 5—29 (1966).
2. Gold B., Rader C. M., Systems for Compressing the Bandwidth of Speech, *IEEE Trans. on Audio and Electroacoustics*, **AU-15**, No. 3, 131—135 (Sept. 1967).
3. Golden R., Vocoder Filter Design: Practical Considerations, *J. Acoust. Soc. Am.*, **43**, 803—810 (April 1968).
4. Gold B., Rader C. M., The Channel Vocoder, *IEEE Trans. on Audio and Electroacoustics*, **AU-15**, No. 4, 148—160 (Dec. 1967).

Выделение основного тона

1. Gold B., Computer Program for Pitch Extraction, *J. Acoust. Soc. Am.*, **34**, 916—921 (1962).
2. Gold B., Description of a Computer Program for Pitch Detection, Proc. Int. Cong. Acoustics, 4th, Copenhagen, Paper G34, 1962.
3. Gold B., Note on Buzz — Hiss Detection, *J. Acoust. Soc. Am.*, **36**, 1659—1661 (1964).

4. Gold B., Rabiner L. R., Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain, *J. Acoust. Soc. Am.*, **46**, No. 2, 442—449 (Aug. 1969).
5. Noll A. M., Cepstral Pitch Determination, *J. Acoust. Soc. Am.*, **41**, 293—309 (1967).

Гомоморфная обработка речи

1. Oppenheim A. V., Schafer R. W., Stockham T. G., Nonlinear Filtering of Multiplied and Convolved Signals, *Proc. IEEE*, **56**, 1264—1291 (1968); есть русский перевод: Оппенхейм, Шефер, Стокхэм мл., Нелинейная фильтрация сигналов, представленных в виде произведения и свертки, *ТИИЭР*, т. 56, № 8, стр. 5—46 (1968).
2. Oppenheim A. V., Schafer R. W., Homomorphic Analysis of Speech, *IEEE Trans. on Audio and Electroacoustics*, **AU-16**, 221—226 (1968).
3. Oppenheim A. V., Speech Analysis-Synthesis System Based on Homomorphic Filtering, *J. Acoust. Soc. Am.*, **45**, 459—462 (1969).
4. Schafer R. W., Rabiner L. R., System for Automatic Analysis of Voiced Speech, *J. Acoust. Soc. Am.*, **47**, Part 2, 634—648 (1970).

Формантные синтезаторы

1. Rabiner L. R., Digital-Formant Synthesizer for Speech Synthesis Studies, *J. Acoust. Soc. Am.*, **43**, 822—828 (1968).
2. Gold B., Rabiner L. R., Analysis of Digital and Analog Formant Synthesizers, *IEEE Trans. on Audio and Electroacoustics*, **AU-16**, 81—94 (March 1968).
3. Rabiner L. R., Jackson L. B., Schafer R. W., Coker C. H., Digital Hardware for Speech Synthesis, *IEEE Trans. on Communication Tech.*, **COM-19**, 1016—1020 (1971).

Линейное прогнозирование речи

1. Atal B. S., Hanauer S. L., Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, *J. Acoust. Soc. Am.*, **50**, 637—655 (1971).
2. Itakura F., Saito S., An Analysis — Synthesis Telephony System Based on Maximum Likelihood Method, *Electronics and Communication in Japan*, **53A**, 36—43 (1970).
3. Makhoul J. I., Wolf J. J., Linear Prediction and the Spectral Analysis of Speech, Bott, Beranek, and Newman Report 2304, Aug. 1972.
4. Markel J. D., Gray A. H., Jr., Wakita H., Linear Prediction of Speech-Theory and Practice, Speech Communication Research Lab. Monograph No. 10, Sept. 1973.

Системы речевого ответа для ЦВМ

1. Rabiner L. R., Schafer R. W., Flanagan J. L., Computer Synthesis of Speech by Concatenation of Formant-Coded Words, Bell Syst. Tech. J., **50**, No. 5, 1541—1558 (May—June 1971).
2. Flanagan J. L., Rabiner L. R., Schafer R. W., Denman J., Wiring Telephone Apparatus from Computer-Generated Speech, Bell Syst. Tech. J., **51**, No. 2, 391—397 (Feb. 1972).