Пример 2. Неравномерное разнесение фильтров используется в тех случаях, когда хотят воспользоваться ухудшением частотного разрешения слуха с ростом частоты. Предположим, что требуется перекрыть тот же самый диапазон от 200 до 3200 Гц, применяя четырехоктавные фильтры, т. е. каждый следующий фильтр имеет полосу вдвое шире предыдущей. Отсюда следует, что диапазон частот от 200 до 3200 Гц следует разбить на четыре полосы шириной в 200, 400, 800 и 1600 Гц с центральными частотами 300, 600, 1200 и 2400 Гц соответственно. Допустив опять, что требуемое затухание составляет 60 дБ, видим, что наименьшая частота среза равна 100 Гц, так что наименьшая переходная область равна 200 Гц. Гребенка фильтров, соответствующая этим параметрам, иллюстрируется рис. 6.31. Дополняющее соотношение между спадами и нараста-

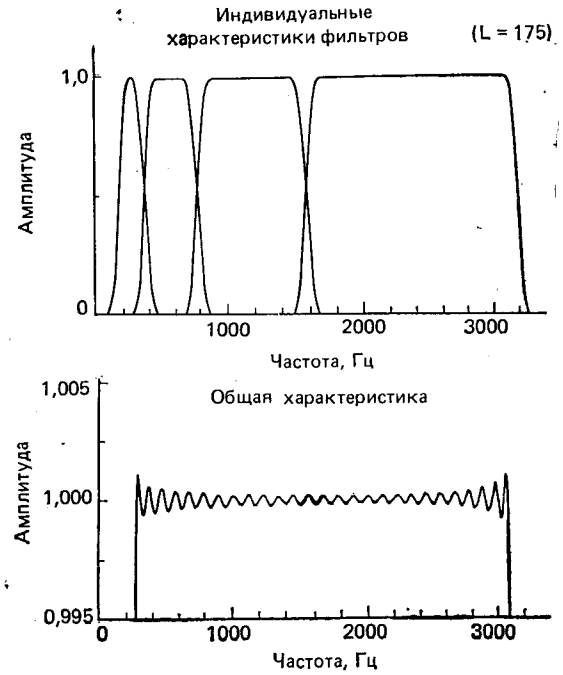

Рис. 6.31. Индивидуальная и общая характеристики гребенки из четырех полосовых фильтров с разными полосами пропускания при $L=175$ [2]

ниями переходов смежных каналов видно в верхней части рисунка. Понятно, что, поскольку L и α одинаковы для каждого из фильтров нижних частот — прототипов, форма кривых в переходной области не зависит от ширины полосы. В нижней части рис. 6.31 приведена общая характеристика, причем отклонение от единицы опять не превосходит 0,001. Как и в примере 1, фаза линейна и соответствует задержке, равной 87 отсчетам. Сравнение рис. 6.30 и 6.31 подтверждает, что в обоих случаях получается одинаковая общая характеристика.

Пример 3. Положим все параметры такими, как и в примере 2, за исключением того, что потребуем более узких переходных областей. Это означает, что необходимо большее значение L. В действительности (6.116а) показывает, что L и Δf, грубо говоря, обратно пропорциональны. На рис. 6.32 показаны характеристики соответствующей гребенки фильтров с параметрами, взятыми из примера 2 с $L=301$ и $\Delta f=0{,}012082$; ширина области перехода равна 116 Гц. В

верхней части рисунка видны более крутые переходы, а из нижней части ясно, что общая характеристика осталась плоской. Задержка в этом случае равна 150 отсчетам.

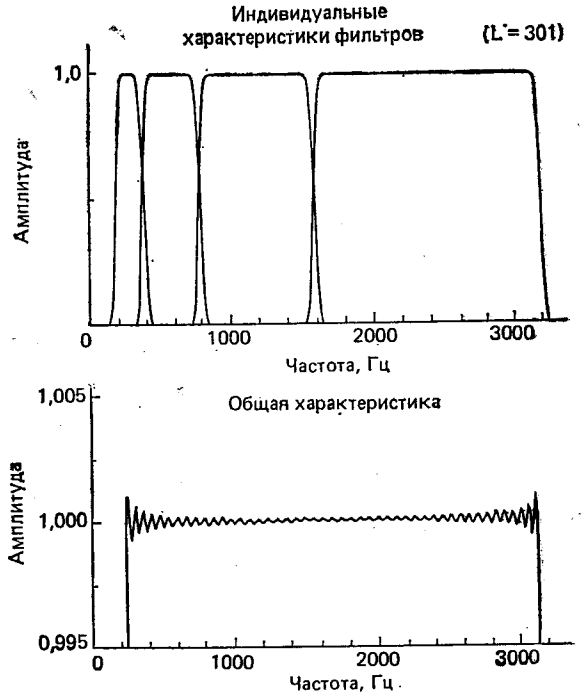

Рис. 6.32. Индивидуальная и общая характеристики гребенки из четырех полосовых фильтров с разными полосами пропускания при $L=301$ [2]

6.3. Реализация метода суммирования выходов гребенки фильтров с помощью БПФ

В предыдущем разделе показано, что можно спроектировать гребенку физически реализуемых фильтров, выход которой совпадает со входом с точностью до задержки и масштабного множителя. В частности, для этой цели особенно хорошо подходят фильтры с конечной импульсной характеристикой. Поскольку кратковременные анализ и синтез Фурье эквивалентны такой гребенке фильтров, то при проектировании систем анализа — синтеза можно эффективно использовать анализирующие окна конечной длительности. Одним из основных недостатков систем на КИХ-фильтрах является большое число вычислений при их реализации. В частном случае кратковременного анализа Фурье имеется удачное обстоятельство — существует ряд методов, позволяющих снизить объем вычислений по сравнению с прямой реализацией.

6.3.1. Методы анализа

Рассмотрим систему кратковременного анализа—синтеза Фурье с равноразнесенными частотами анализа $\omega_k=2\pi k/N$, $0\leqslant k\leqslant N-1$. В 6.1.3 было показано, что нет необходимости вычислять $X_n(e^{i\omega_k})$ с частотой дискретизации на входе, поскольку каж-

дая из последовательностей представляет собой по существу выходную последовательность фильтра нижних частот с цифровой частотой среза π/N. Следовательно, входную последовательность можно вычислять всего 1 раз через каждые N последовательных отсчетов на входе. В таких случаях особенно подходят КИХ-системы, так как в них можно ограничиться вычислением только нужных выходных отсчетов, не вычисляя промежуточных $N-1$ отсчетов. В БИХ-системах приходится вычислять на выходе все значения из-за присущей им рекурсивной природы реализации.

Дополнительное увеличение эффективности вычислений можно получить за счет применения метода быстрого преобразования Фурье (БПФ) [6]. Чтобы уяснить это, выразим кратковременное преобразование Фурье в виде

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=\sum_{m=-\infty}^{\infty}x(m)w(n-m)e^{-i\frac{2\pi}{N}km},\ 0\leqslant k\leqslant N-1. \quad (6.130)$$

Видно, что если бы пределами суммирования были 0 и $N-1$, то отношение (6.130) приняло бы форму дискретного преобразования Фурье (ДПФ). В случае конечной длительности $w(m)$ выражение (6.130) можно привести к виду ДПФ и, следовательно, можно для вычисления $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $0\leqslant k\leqslant N-1$ воспользоваться алгоритмом БПФ. При замене переменных суммирования (6.130) превращается в

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{m=-\infty}^{\infty}x_n(m)e^{-i\frac{2\pi}{N}km}, \quad (6.131)$$

где

$$x_n(m)=x(n+m)w(-m),\ -\infty<m<\infty. \quad (6.132)$$

Иначе говоря, последовательность $x_n(m)$ получается при переносе начала последовательности $x(m)w(n-m)$ в отсчет n, что концентрирует внимание на членах последовательности в окрестности момента времени, для которого необходимо вычислить $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Далее подстановками $m=Nr+q$, $-\infty<r<\infty$ и $0\leqslant q\leqslant N-1$ можно представить (6.131) двойной суммой:

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{r=-\infty}^{\infty}\left(\sum_{q=0}^{N-1}x_n(Nr+q)\right)e^{-i\frac{2\pi}{N}k(Nr+q)}. \quad (6.133)$$

Поскольку $e^{-i2\pi kr}=1$, можно поменять порядок суммирования и получить

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=e^{-i\frac{2\pi}{N}kn}\sum_{q=0}^{N-1}\left(\sum_{r=-\infty}^{\infty}x_n(Nr+q)\right)e^{-i\frac{2\pi}{N}kq}. \quad (6.134)$$

Определив конечную последовательность

$$u_n(q)=\sum_{r=-\infty}^{\infty}x_n(Nr+q),\ 0\leqslant q\leqslant N-1, \quad (6.135)$$

видим, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой просто произведение $e^{-i\frac{2\pi}{N}kn}$ на N-точечное ДПФ последовательности $u_n(q)$. Или, по-другому, $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой N-точечное ДПФ последовательности $u_n(q)$ после циклического сдвига на n по модулю N. Иначе говоря,

$$X_n\left(e^{i\frac{2\pi}{N}k}\right)=\sum_{m=0}^{N-1}u_n((m-n))_N e^{-i\frac{2\pi}{N}km}, \quad (6.136)$$

где обозначение $((m-n))_N$ указывает на то, что целое в двойных скобках следует рассматривать по модулю N. Мы сумели, таким образом, привести $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ к виду N-точечного ДПФ последовательности конечной длины, полученной по взвешенной окном входной последовательности. Итак, процедура для вычисления $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $0\leqslant k\leqslant N-1$ состоит в следующем:

1. Сформировать последовательность $x_n(m)$, такую, как в (6.132), умножая $x(m+n)$ на обращенную во времени последовательность окна $w(-m)$. На рис. 6.33 показаны $x(m+n)$ и три специальных случая $w(-m)$.

2. Разбить результирующую последовательность на сегменты по N отсчетов и сложить эти сегменты вместе в соответствии с (6.135), что даст последовательность конечной длины $u_n(q)$, $0\leqslant q\leqslant N-1$.

3. Циклически сдвинуть $u_n(q)$ на n по модулю N, что даст $u_n((m-n))_N$, $0\leqslant m\leqslant N-1$.

4. Вычислить N-точечное ДПФ от $u_n((m-n))_N$, что даст $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ $0\leqslant k\leqslant N-1$. Эту процедуру придется повторить для каждого n, при котором необходимо значение $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Ясно, однако, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ можно наращивать любым требуемым способом. Можно, например, вычислять $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при $n=0, \pm R, \pm 2R, \ldots$, т. е. с интервалом R отсчетов входного сигнала. Это оправдано тем, что $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ представляет собой выход фильтра нижних частот с номинальной частотой среза π/N радиан. Следо-

вательно, «отсчетов» $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ будет достаточно для восстановления входного сигнала, если только $R \leqslant N$.

Этот метод даст значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ для всех k. В общем случае можно ограничиться вычислениями, самое большее для половины каналов. Это связано с сопряженной симметрией $X_n(e^{i\omega})$. Часто каналы на очень низких и очень высоких частотах не реализуют. Возникает, следовательно, вопрос: будет ли метод ДПФ эффективнее непосредственной реализации. Чтобы сравнить их, предположим, что нам требуются значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ только при $1 \leqslant k \leqslant M$. Допустим, кроме того, что длительность окна равна L. Тогда для вычисления полного набора значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ потребуется $4LM$ умножений и примерно $2LM$ сложений, если используется метод, иллюстрированный рис. 6.12. Допустив, что применен обычный комплексный алгоритм БПФ (с N, равным целой степени 2)[1], можно показать, что метод БПФ потребует примерно $L+2N\log_2 N$ умножений и $L+2N\log_2 N$ сложений, чтобы получить

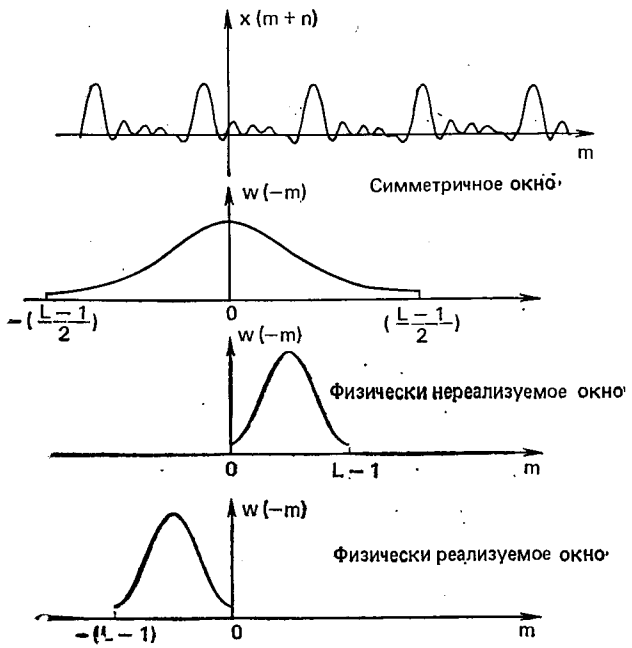

Рис. 6.33. Графики последовательностей $x(m+n)$ и $w(-m)$

[1] Здесь не используется то, что $u_n((m-n))_N$ — действительно. Можно было бы воспользоваться этим и уменьшить число вычислений еще в 2 раза.

все N значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Если за меру сравнения принять число умножений действительных величин, то легко показать, что при $L=N$ метод БПФ потребует меньшего числа вычислений, если только не выполняется неравенство

$$M \leqslant \log_2 N/2. \tag{6.137}$$

Пусть, например, $N=128=2^7$. Видим, что БПФ эффективнее прямого метода, если только не выполняется неравенство $M \leqslant 3{,}5$, т. е. кроме случаев, когда число каналов меньше четырех. Следовательно, во всех приложениях, где необходимо тонкое разрешение по частоте, почти наверняка метод БПФ будет самым эффективным (заметим, что при $L>N$ сравнение еще больше в пользу БПФ).

6.3.2. Методы синтеза

Предыдущее обсуждение методов анализа показало, что, используя алгоритм быстрого преобразования Фурье, можно определить все N равноразнесенные значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ при меньшем объеме вычислений, чем требуется для вычисления M каналов с непосредственной реализацией. Реорганизовав вычисления, необходимые в ходе синтеза, можно получить аналогичный выигрыш, восстанавливая $x(n)$ по значениям $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ через каждые R отсчетов $x(n)$, где $R \leqslant N$ [7].

Из (6.83) при $\omega_k = 2\pi k/N$ получим для выхода системы синтеза

$$y(n) = \sum_{k=0}^{N-1} Y_n(k)\, e^{i\frac{2\pi}{N}kn}, \tag{6.138}$$

где

$$Y_n(k) = P_k X_n\left(e^{i\frac{2\pi}{N}k}\right), \quad 0 \leqslant k \leqslant N-1. \tag{6.139}$$

Вспомним, что весовые комплексные коэффициенты P_k позволяют подобрать модуль и фазу каналов. Если имеются значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ только при целых кратных R, то промежуточные значения можно получить с помощью интерполяции. Для этого полезно определить последовательность

$$V_n(k) = \begin{cases} P_k X_n\left(e^{i\frac{2\pi}{N}k}\right), & n = 0, \pm R, \pm 2R, \ldots \\ 0, & \textit{в противном случае.} \end{cases} \tag{6.140}$$

Для каждого значения k имеется последовательность указанного вида. Для каждого k промежуточные значения заполняются теперь за счет обработки последовательности $V_n(k)$ фильтром нижн-

них частот с частотой среза π/N радиан. Если обозначить отклик этого фильтра на единичный импульс через $h(n)$ и допустить, что он симметричен, а полная длина равна $2RQ-1$, то для каждого k из интервала $0 \leqslant k \leqslant N-1$ получим

$$Y_n(k) = \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, V_m(k), \quad -\infty < n < \infty. \tag{6.141}$$

Соотношение (6.141) вместе с (6.138) описывает операции, необходимые при вычислении выхода синтеза в случае, если имеется кратковременное преобразование Фурье, заданное через интервалы в R отсчетов. Этот процесс показан на рис. 6.34, на котором

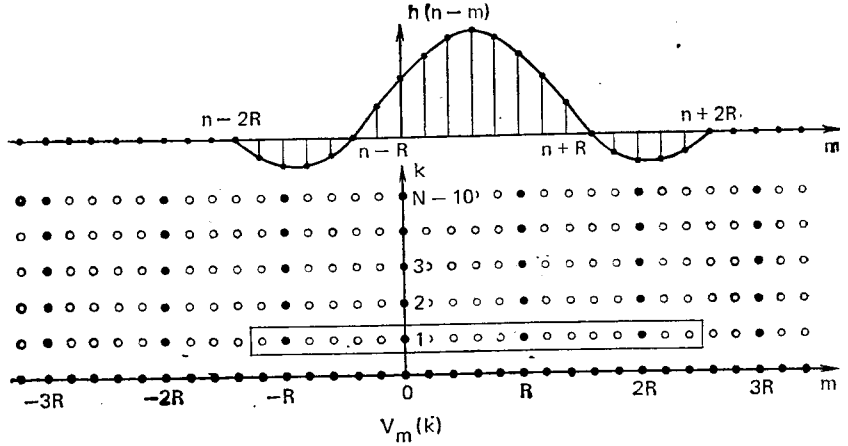

Рис. 6.34. Отсчеты, входящие в вычисление $Y_n(k)$

$V_m(k)$ приведена как функция m и k (напомним, что m — индекс времени, а k — индекс частоты). Точками отмечены координаты, в которых $V_m(k)$ отлична от нуля, т. е. точки, в которых известно $X_m\left(e^{i\frac{2\pi}{N}k}\right)$. Кружками отмечены точки, в которых $V_m(k)$ равна нулю и в которых требуется интерполировать значения $Y_n(k)$. Импульсная характеристика интерполирующего фильтра (для $Q=2$) показана в момент n. Сигнал каждого из каналов интерполируется свертыванием с импульсной характеристикой интерполирующего фильтра. В качестве примера отсчеты, связанные с вычислением $Y_3(1)$, помещены в рамку. В общем случае рамка, указывающая на отсчеты, связанные с вычислением $Y_n(k)$, будет скользить вдоль k-й строки (рис. 6.34), причем центр рамки совпадает с координатой n. Отметим, что каждое из интерполируемых значений зависит от $2Q$ известных значений $X_n\left(e^{i\frac{2\pi}{N}k}\right)$. Предположив, что в процессе синтеза доступны M каналов, легко показать, что потребуется $2(Q+1)M$ умножений и $2QM$ сложений для вычисления каждого из значений выходной последовательности.

Чтобы понять, как можно вычислить выходной сигнал более эффективно, подставим (6.141) в (6.138), что даст

$$y(n) = \sum_{k=0}^{N-1} \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, Y_m(k)\, e^{i\frac{2\pi}{N}kn}. \tag{6.142}$$

Изменив порядок суммирования, получим

$$y(n) = \sum_{m=n-RQ+1}^{n+RQ-1} h(n-m)\, v_m(n), \tag{6.143}$$

где

$$v_m(r) = \sum_{k=0}^{N-1} V_m(k)\, e^{i\frac{2\pi}{N}kr}. \tag{6.144}$$

Воспользовавшись (6.140), видим, что

$$v_m(r) = \begin{cases} \sum_{k=0}^{N-1} P_k X_m\left(e^{i\frac{2\pi}{N}k}\right) e^{i\frac{2\pi}{N}kr}, & m = 0, \pm R, \pm 2R, \ \ldots \\ 0, & \textit{в противном случае.} \end{cases} \tag{6.145}$$

Следовательно, вместо того чтобы интерполировать кратковременное преобразование Фурье, а затем вычислять (6.138), можно вычислять $v_m(r)$ во все те моменты, в которых известна $X_n\left(e^{i\frac{2\pi}{N}k}\right)$, т. е. $m=0, \pm R, \ldots$, а затем интерполировать $v_m(r)$, как в (6.143).

Можно видеть, что $v_m(r)$ имеет вид обратного дискретного преобразования Фурье и, следовательно, $v_m(r)$ периодична по r с периодом N. Поэтому в (6.143) переменную n в $v_m(n)$ надо интерпретировать по модулю N. Такой процесс интерполяции приведен на рис. 6.35. Жирными точками в двумерной сети представлены точки, в которых $v_m(r)$ отлична от нуля. Оставшиеся точки на рис. 6.35 можно интерполировать так, как это было описано при интерполяции $Y_n(k)$. Однако нет необходимости интерполировать все эти точки, поскольку требуется получить только значения одномерной последовательности $y(n)$. Из (6.143) и периодического характера $v_m(r)$ видно, что $y(n)$ равна значениям интерполируемой последовательности вдоль «пилы» (см. рис. 6.35).

Реализуя таким способом систему синтеза, можно вычислить N-точечную последовательность $v_m(r)$, $0 \leqslant r \leqslant N-1$ для каждого значения m, в котором известна $X_n\left(e^{i\frac{2\pi}{N}k}\right)$, используя при этом алгоритм быстрого преобразования Фурье для выполнения вычислений ДПФ в (6.145). Чтобы исключить канал, достаточно просто приравнять значение соответствующего ему сигнала нулю до вычисления обратного дискретного преобразования Фурье. Аналогичным образом, если требуется реализовать линейный фазовый

сдвиг выбором $P = e^{-i\frac{2\pi}{N}kn_0}$, то, как нетрудно показать, результат сведется к простому циклическому сдвигу последовательности $v_m(r)$. Можно, следовательно, избежать умножения на $e^{-i\frac{2\pi}{N}kn_0}$, выполнив обратное дискретное преобразование Фурье непосредственно над $X_m\left(e^{i\frac{2\pi}{N}k}\right)$ и сдвинув затем циклически результат на n_0 отсчетов. Коль скоро получены последовательности $v_m(r)$, выход можно вычислить, интерполируя $v_m(r)$, как в (6.143). Для каждого значения $y(n)$ необходимы $2Q$ значения $v_m(r)$. Отсчеты, необходимые для двух различных значений n, показаны заключенными в рамки на рис. 6.35. Для R последовательных значений $y(n)$ значения $v_m(r)$ получаются из одних и тех же $2Q$ столбцов. Поэтому выход удобно вычислять блоками по R отсчетов.

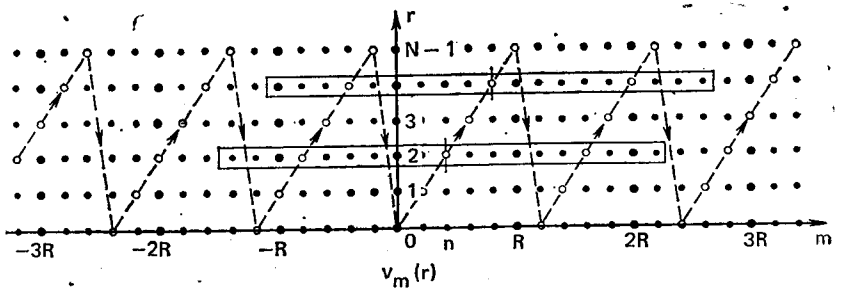

Рис. 6.35. Процесс интерполяции $v_m(r)$ по [7]

Объем вычислений, требуемых при реализации кратковременного синтеза Фурье указанным выше способом, можно снова оценить, предположив, что N — степень двойки и что применен алгоритм БПФ для вычислений обратных преобразований в (6.145). При таком допущении синтез требует $(2QR+2N\log_2 N)$ умножений действительных величин и $(2QR-1+2N\log_2 N)$ сложений действительных величин для вычисления группы из R последовательных значений выхода $y(n)$. Прямой метод синтеза требует $2(Q+1)MR$ умножений и $2QMR$ сложений для вычисления R последовательных отсчетов выхода. Рассмотрев случай, когда прямой метод требует меньшего, чем метод БПФ, числа умножений, найдем

$$M < \frac{Q + (N/R)\log_2 N}{Q+1}. \qquad (6.146)$$

Для типичных значений $N=128$, $Q=2$ (интерполяция через четыре отсчета, как на рис. 6.35) и при $R=N$ (наименьшая возможная частота дискретизации $X_n\left(e^{i\frac{2\pi}{N}k}\right)$) видим, что прямой метод эффективнее только в том случае, если $M<3$. Следовательно, для большинства приложений БПФ дает значительное увеличение эффективности вычислений операций синтеза.

6.4. Спектрографическое отображение

Идеи кратковременного преобразования Фурье появились задолго до появления методов цифровой обработки сигналов. Исследователи речи полагались на спектральные методы анализа начиная с 30-х годов нашего века. Одним из устройств, использующих кратковременное представление Фурье, был звуковой спектрограф — устройство, ставшее важным инструментом в почти каждой фазе исследования речи. В этом приборе короткие (2 с) отрезки речи многократно модулируют сигнал генератора переменной частоты. Модулированный сигнал поступает на полосовой фильтр. Средняя энергия на выходе полосового фильтра для заданных частоты и времени представляет собой грубое приближение кратковременного преобразования Фурье. Эта энергия регистрируется остроумной электромеханической системой на электрочувствительной бумаге. Результат, называемый спектрограммой, представляет собой двумерное представление зависящего от времени спектра, при этом по вертикали откладывается частота, а по горизонтали — время. Амплитуда спектра представляется затемнением отметок на бумаге. Если полосовые фильтры имеют большую ширину полосы (300 Гц), то на спектрограмме получается хорошее разрешение по времени и плохое по частоте. При узкой полосе (45 Гц) спектрограмма имеет хорошее разрешение по частоте и плохое по времени.

На рис. 6.36*а* приведена широкополосная спектрограмма фразы «Every salt breeze comes from the sea». Этот пример иллюстрирует ряд характерных свойств широкополосного кратковременного спектра. Во-первых, видно, что в фиксированный момент времени спектр меняется с частотой, как это показано на рис. 6.3, 6.5, т. е. спектр состоит из ряда широких пиков, соответствующих частотам формант. На спектрограмме четко отражены изменения во времени частот формант. Кроме того, широкополосная спектрограмма имеет линейчатый характер в областях вокализованной речи. Она возникает из-за того, что импульсная характеристика (т. е. анализирующее окно спектра) имеет длительность, примерно совпадающую с периодом основного тона. Поэтому энергия на выходе фильтра максимальна, когда пик импульсной характеристики совпадает с максимумом каждого периода основного тона. В другие моменты энергия на выходе значительно меньше. Для невокализованной речи, которая непериодична, линейчатость исчезает и спектр оказывается более «рваным».

На рис. 6.36*б* показана узкополосная спектрограмма той же фразы. В этом случае у фильтра полоса пропускания выбрана так, что в вокализованной области разделяются отдельные гармоники. Хотя по-прежнему видны частоты формант, сечение в фиксированный момент времени напоминает спектр рис. 6.2 и 6.4. В вокализованной области уже нет линейчатости, поскольку импульсная характеристика при узкой полосе захватывает несколько периодов основного тона; теперь, однако, четко видны основная

частота и ее гармоники. Невокализованные области выделяются отсутствием периодичности по частоте.

На широкополосных и узкополосных спектрограммах отображается существенная часть информации о свойствах речи. Когда

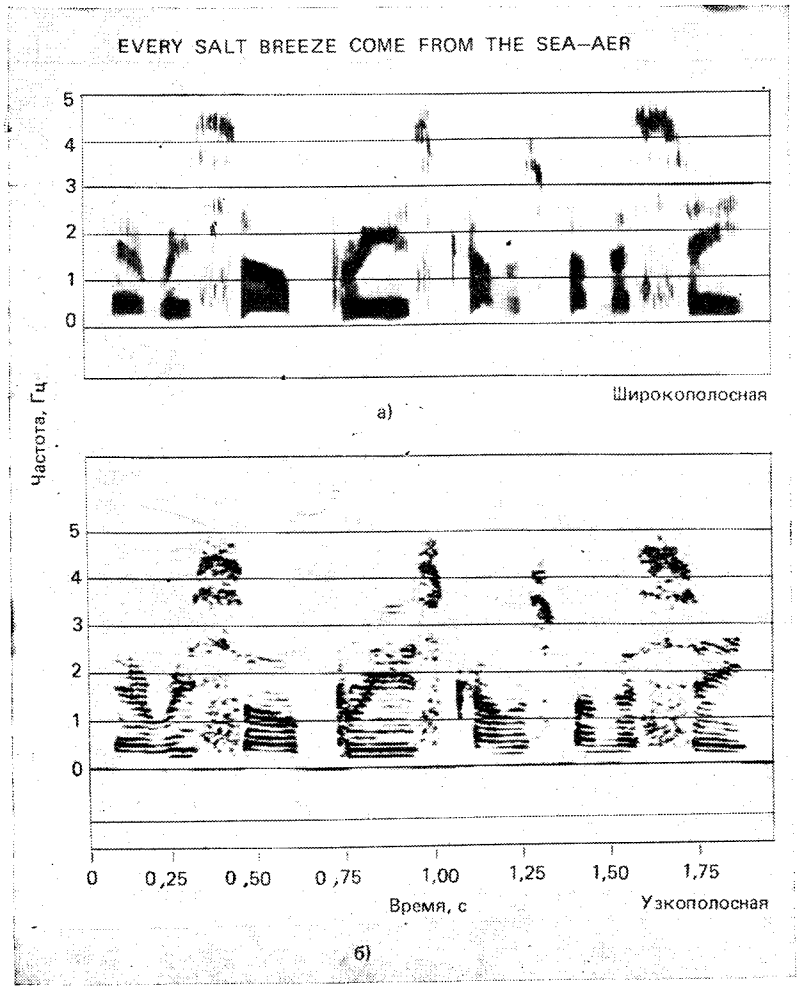

Рис. 6.36. Широкополосная и узкополосная спектрограммы предложения

аппаратура для такого отображения кратковременного представления Фурье появилась впервые, надеялись даже, что получен новый «язык» для общения с глухими. И хотя этим надеждам не суждено было воплотиться в жизнь, последующие исследования привели к написанию книги «Visible Speech» [8], которая и по сей день служит источником информации о спектральных и временных свойствах речи. За годы, прошедшие с появления этой работы, многие исследователи речи вручную определяли по спектрограммам такие параметры речи, как частоты формант и основная частота.

Другим следствием изобретения спектрографа стало то обстоятельство, что по детальному анализу спектрограммы или «отпечатка голоса» произнесенной фразы, как оказалось, можно установить личность говорящего. И хотя вопрос надежности такой идентификации по спектрограмме остается открытым, она получила некоторое признание в юрисдикции.

Звуковой спектрограф долгое время оставался основным инструментом анализа в исследованиях речи. С появлением вычислительных машин, доступных при исследованиях речи, это изменилось. В предыдущих разделах этой главы показаны способы реализации кратковременного представления Фурье. Они гораздо более хитроумные, чем способы, реализуемые на аналоговой аппаратуре. Эти представления могут быть реализованы и в цифровой специализированной аппаратуре, и как программы для универсальной вычислительной машины. Воспользовавшись, например, методами, изложенными в § 6.3, можно получить $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ — комплексное двумерное представление речевого сигнала с дискретным временем и частотой и, кроме того, периодическое по частоте. Возникает, следовательно, задача визуального отображения такого представления. Часто ограничиваются только отображением $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$. Поскольку последовательность $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$ четна и периодична по k с периодом N, на практике необходимо отображать значения только из интервала $0 \leqslant k \leqslant N/2$.

В случаях, когда в вычислительной машине имеется осциллоскоп или графопостроитель, кратковременное преобразование Фурье можно отображать просто в виде последовательности графиков $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$ как функцию k при фиксированных n. Обычно значения n разносятся на интервалы, соответствующие частоте Найквиста в спектральных каналах. Для узкополосного анализа, например, разнос по времени может быть порядка 10—20 мс. На рис. 6.37 [11] приведена последовательность узкополосного спектра, вычисленного через интервалы 20 мс. Видно, что весь сегмент речи вокализован.

Альтернативой отображения спектра как сечений поверхности, определяемой $|X_n\left(e^{i\frac{2\pi}{N}k}\right)|$, может служить отображение этой поверхности в перспективе (рис. 6.38) [12].

Ясно, что такой график менее полезен при количественных измерениях. Однако он, как и спектрограмма, имеет преимущество отображения всей фразы в компактной форме.

Поскольку полезность и повсеместная принятость спектрограмм как основного инструмента уже продемонстрирована, постольку

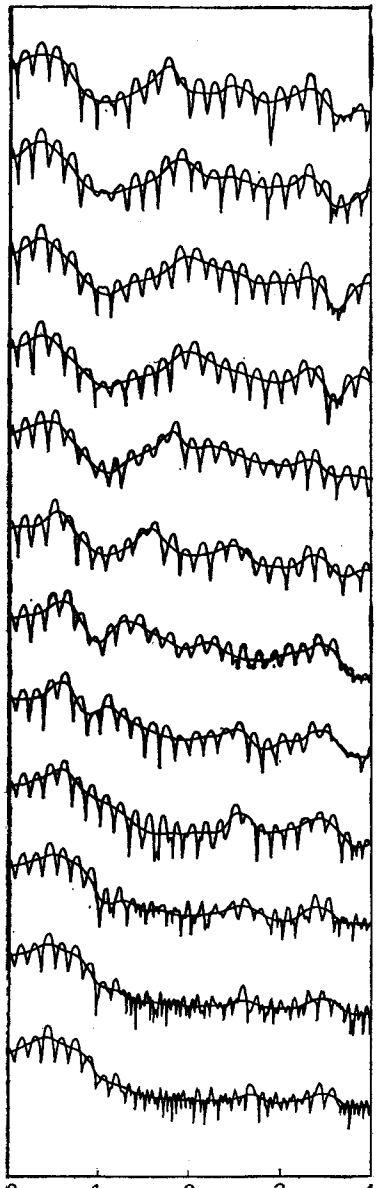

Рис. 6.37. Набор узкополосных спектров сегмента вокализованной речи [11]

по-видимому, цифровые спектрограммы полезнее любых других способов отображения. Если имеется телевизионный или электронно-лучевой дисплей для отображения дискретизованных изображений, то значение $|X_n(e^{i\frac{2\pi}{N}k})|$ из подходящего по размерам интервала можно рассматривать как дискретизованное изображение[1]. Некоторые исследователи рассмотрели такие возможности и нашли, что можно добиться полного сходства с аналоговыми спектрограммами. Действительно, поскольку электрочувствительная бумага имеет диапазон черного, равный 12 дБ [13], достаточно довольно грубого квантования значений $|X_n(e^{i\frac{2\pi}{N}k})|$, чтобы имитировать спектрограмму. У большинства цифровых отображающих устройств динамический диапазон гораздо шире. Следовательно, они могут дать гораздо больше спектральной информации по сравнению с аналоговыми системами.

Другое достоинство цифровых спектрограмм заключается в удобстве формирования спектра разнообразными способами, что увеличивает полезность дисплея. Примером может служить подчеркивание высоких частот, компенсирующее естественный спад спектра речи (это используется и в аналоговых спектрографах). Простой способ добиться подчеркивания высоких частот заключается в вычислении спектра первой разности входного сигнала (см. задачу 6.11). Другой, более гибкий способ, заключается в непосредственном формировании требуемого спектра до его отображения.

[1] Обычно этого добиваются, используя дополнительную память, которая служит для обновления изображения. Л. Р. Моррис изучил, однако, методы отображения спектрограмм, использующие только память и выходные устройства стандартного миникомпьютера (IEEE Tr., ASSP, June, 1975).

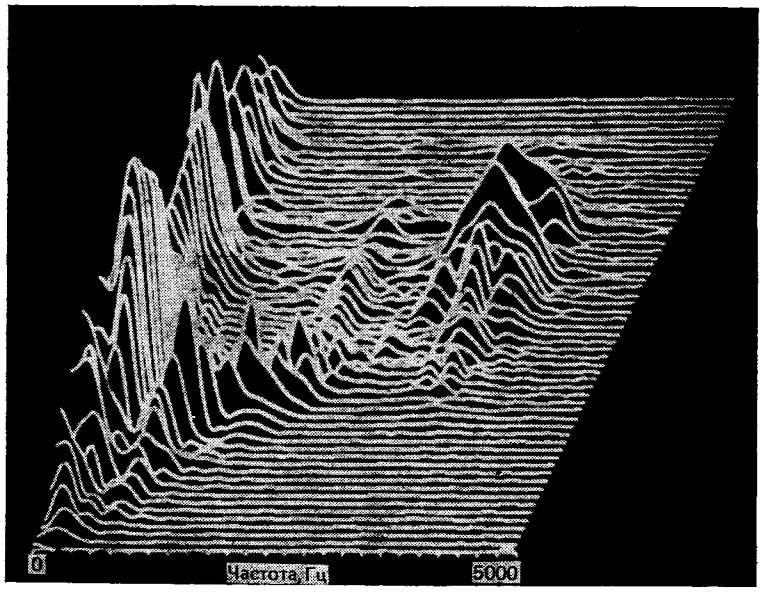

Рис. 6.38. Спектрограмма слова «read», вычисленная по сегментам речи длительностью 8 мс [12]

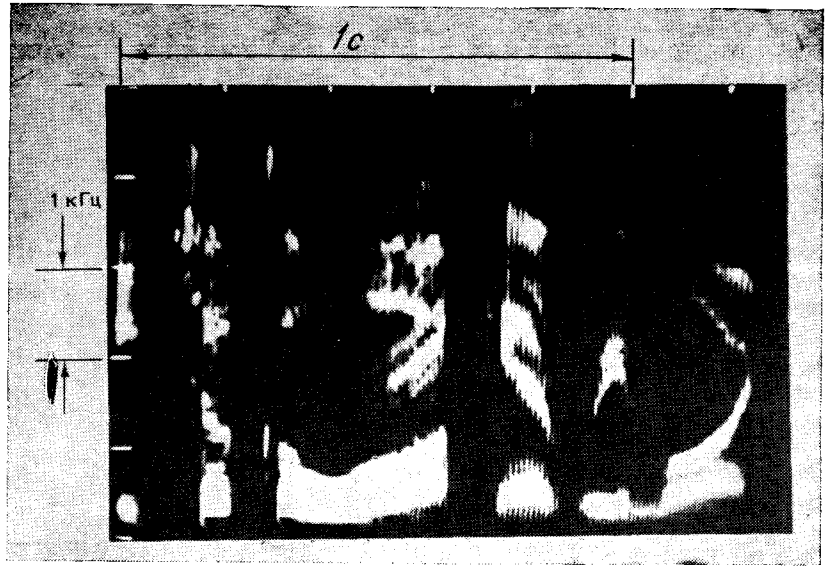

Рис. 6.39. Пример спектрограммы, полученной применением кратковременного анализа и графической системы ЭВМ [14]

Этот подход использован в [14] для создания с помощью ЭВМ спектрограмм, аналогичных показанным на рис. 6.39. Там же показано, что имеются широкие возможности отображения спектральной информации с преобразованием. Например, частотная и временна́я шкалы могут быть сжаты или растянуты по желанию.

Существует еще один подход к созданию спектрограмм с помощью ЭВМ, применяемый в тех случаях, когда нет другой возможности наглядно отобразить выходной сигнал. Если имеется печатающее устройство, допускающее повторную печать, можно получить шкалу черного, сравнимую со шкалой аналоговой спектрограммы, задавая каждый уровень набором накладываемых друг на друга печатных символов (рис. 6.40). Детали получения процедуры таких оттисков приведены в [15].

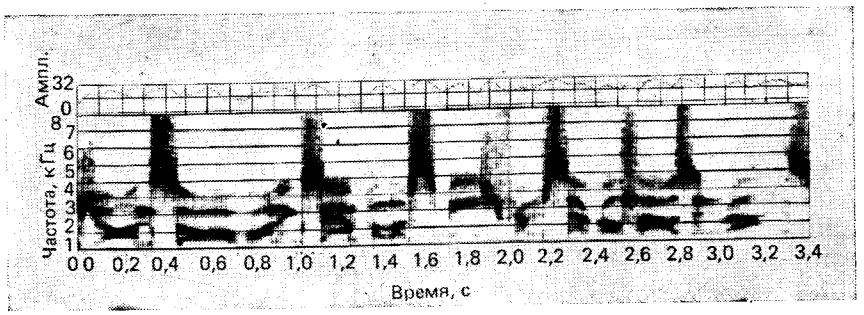

Рис. 6.40. 800-точечная ДПФ спектрограмма [15]

6.5. Выделение основного тона

При узкополосном кратковременном преобразовании Фурье возбуждение вокализованной речи проявляется узкими пиками на частотах, кратных основной частоте. Этот факт положен в основу ряда схем выделения основного тона. Рассмотрим выделитель основного тона, основанный на кратковременном спектральном анализе. Этот пример иллюстрирует как основные концепции использования кратковременного спектра для выделения основного тона, так и гибкость методов цифровой обработки. Внимательному читателю станет ясно, что существует еще много возможностей для использования кратковременного представления Фурье в задаче определения параметров возбуждения (другой пример предлагается в задаче 6.14).

Один из подходов связан с вычислением *произведения гармоник спектра*, определенного в [16] как

$$P_n(e^{i\omega}) = \prod_{r=1}^{K} |X_n(e^{i\omega r})|^2. \qquad (6.147)$$

Взяв логарифм, получим *логарифмическое произведение гармоник спектра:*

$$\hat{P}_n(e^{i\omega}) = 2\sum_{r=1}^{K} \log|X_n(e^{i\omega r})|. \qquad (6.148)$$

Видно, что $\hat{P}_n(e^{i\omega})$ представляет собой сумму K сжатых по частоте $\log|X_n(e^{i\omega})|$. Введение функции (6.148) мотивируется тем, что в вокализованной речи сжатие частотной шкалы в целые числа раз должно привести к совпадению гармоник основной частоты с ней самой. И на промежуточных частотах некоторые из сжатых по частоте гармоник будут совпадать, но всегда усиливаться они будут только на основной частоте. Схематически это приведено на рис. 6.41. Для непрерывной функции $|X_n(e^{i2\pi FTr})|$ пик на частоте F_0 становится все острее с ростом r. Следовательно, в сумме (6.148) острый пик будет на частоте F_0, возможно также появление меньших пиков на других частотах. Было найдено, что этот метод особенно устойчив к аддитивным шумам, поскольку вклад шумов в $X_n(e^{i\omega})$ не имеет коррелированной структуры, если рассматривается как функция частоты. Поэтому и в (6.148) шумовые компоненты в $X_n(e^{i\omega r})$ имеют тенденцию складываться некоррелированно. По той же причине невокализованная речь не даст явного пика в $\hat{P}_n(e^{i\omega})$. Другая важная особенность логарифмического произведения гармоник заключается в том, что на основной частоте пик вовсе не «обязан» явно присутствовать в $|X_n(e^{i\omega})|$, если он имеется в $\hat{P}_n(e^{i\omega})$. Этот метод, следовательно, привлекателен для работы с речью, пропущенной через фильтр высоких частот, как и получается в телефонной линии.

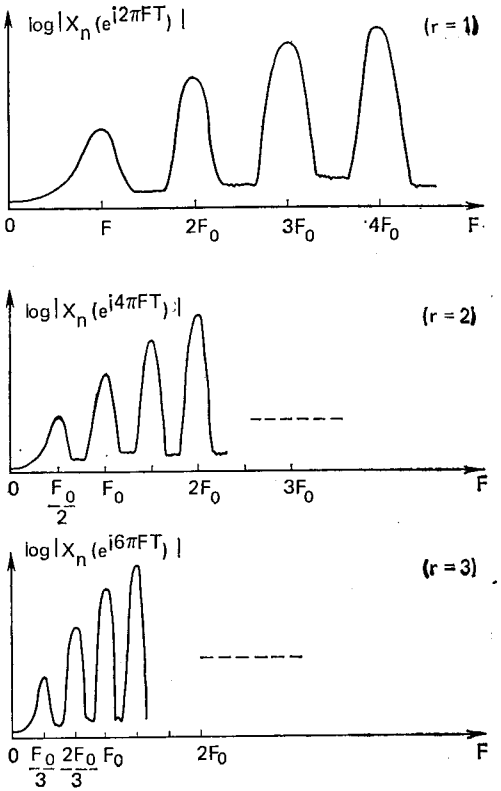

Рис. 6.41. Характер членов в логарифме произведения гармоник спектра

Пример использования метода приведен на рис. 6.42 [16]. Входная речь дискретизировалась с частотой 10 кГц и через каждые 10 мс сигнал умножался на окно Хемминга (400 отсчетов).

Затем вычислялись значения $X_n\left(e^{i\frac{2\pi}{N}k}\right)$ с помощью алгоритма БПФ с $N=2048$. На рис. 6.42*а* и *б* показана последовательность произведения гармоник и ее логарифма соответственно для (6.147) и (6.148) и при $K=5$. На рис. 6.42*в* и *г* приведены результаты вы-

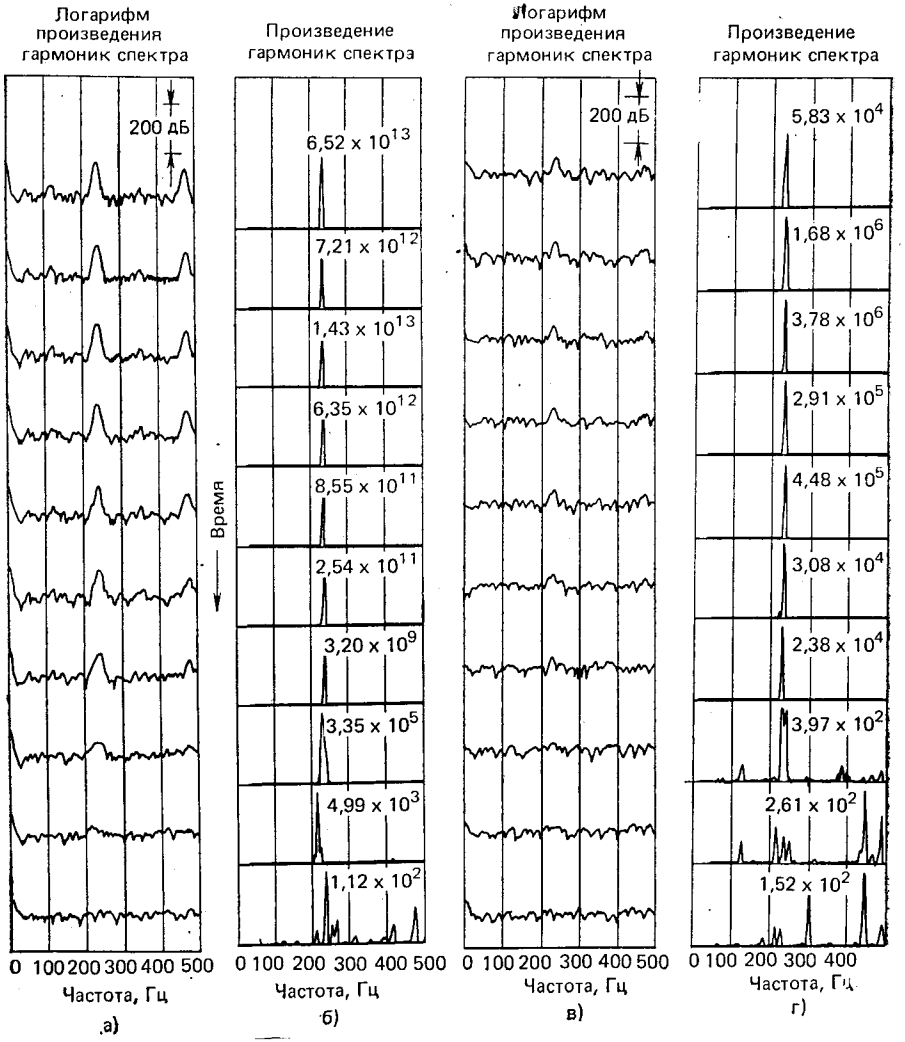

Рис. 6.42. Логарифм произведения гармоник спектра и произведение гармоник: *а*, *б*) — без шума; *в*, *г*) — отношение сигнал/шум равно 0 дБ [17]

числений с теми же параметрами, но во входной сигнал добавлен шум с отношением сигнал/шум, равным 0 дБ. Замечательна четкость, с которой проявляется основная частота. Из этих рисунков ясно, что на основе произведения гармоник можно получить простой алгоритм выделения основного тона. То, что такой алгоритм обладает великолепной устойчивостью к шуму, было проверено в [17].

6.6. Анализ через синтез

В § 6.4 и 6.5 показано, что основные параметры речи ясно проявляются в кратковременном преобразовании Фурье. В этом разделе рассмотрим метод, называемый анализом через синтез. Он оказался полезным при оценке частот формант и для оценивания глоттальных колебаний[1] вокализованной речи.

Основная идея анализа через синтез следующая. Предположим, во-первых, что имеется речевое колебание во времени или какое-либо другое представление речевого сигнала, например кратковременное преобразование Фурье. Допустим затем, что предлагается некоторая модель речеобразования. Эта модель (например, эквивалентная модель голосового тракта) имеет ряд параметров, изменяя которые можно получить различные звуки речи. Можно получить представление для модели, такое же, как и представление речевого сигнала. Если, например, речевой сигнал представлен кратковременным преобразованием Фурье, то можно также получить и кратковременное преобразование Фурье для модели. Меняя затем параметры модели некоторым систематическим образом, можно, например, найти такие значения параметров, при которых модель согласуется с речевым сигналом с минимальной ошибкой. Когда достигнуто такое согласование, считают, что параметры модели представляют собой параметры речи. Это весьма общий подход, не привязанный к кратковременному преобразованию Фурье. Принцип, однако, впервые использовался для анализа речи [18], затем был применен во временной области [19] и для кепструма [20], [21].

Одним из самых ранних описаний применения принципа анализа через синтез к речи было сделано группой в МТИ [22]. В этой работе кратковременное представление Фурье было получено с помощью гребенки аналоговых фильтров. Сигналы на выходах фильтров дискретизировались и подавались на ЭВМ. Результирующее грубое спектральное представление затем преобразовывалось по итеративной процедуре для подбора параметров в модели речеобразования. Параметры включали спектральные компоненты передаточной функции голосового тракта, форму глоттальных колебаний и сопротивление излучения. И хотя алгоритм подбора параметров модели не был полностью автоматическим, работа показала применимость метода анализа через синтез, дав отличные результаты при оценивании формант вокализованной речи [22].

Самым серьезным ограничением в схеме, описанной Беллом и другими [18], было использование аналоговой гребенки фильтров для анализа Фурье. Это ограничение снято в схеме, предложенной Мэтьюзом, Миллером и Дэйвидом [23]. Они начинали с отсчетов речевого сигнала и реализовали анализ Фурье на цифровой машине. Их подход привел к появлению еще одной новой концепции в спектральном анализе речи: понятию анализа речи, синхронного с основным тоном. И хотя работа появилась до того, как были развиты теория и практика применения дискретного анализа Фурье, воспользуемся преимуществами такого анализа для объяснения этого подхода.

6.6.1. Спектральный анализ, синхронный с основным тоном

В нашей модели короткий сегмент вокализованной речи совпадает с сегментом такой же длины периодической последовательности

$$\tilde{x}(n)=\sum_{m=-\infty}^{\infty} h_v(n+mN_p), \qquad (6.149)$$

[1] Имеется в виду форма колебания воздушного потока в голосовой щели. (*Прим. ред.*)

где $h_v(n)$ представляет свертку импульсной характеристики голосового тракта $v(n)$ с глоттальным импульсом $g(n)$ и с импульсной характеристикой сопротивления излучения $r(n)$. Иначе

$$h_v(n) = r(n) * v(n) * g(n). \tag{6.150}$$

Величина N_p есть период основного тона в отсчетах. Эффекты излучения, как правило, проявляющиеся как дифференцирование на нижних частотах, моделируются адекватно для большинства приложений просто вычислением первой разности, которой соответствует следующее z-преобразование:

$$R(z) = 1 - z^{-1}. \tag{6.151}$$

Голосовой тракт характеризуется передаточной функцией вида

$$V(z) = \frac{A}{\prod_{k=1}^{M} (1 - 2e^{-\sigma_k T}\cos(2\pi F_k T)\, z^{-1} + e^{-2\sigma_k T} z^{-2})}, \tag{6.152}$$

где число полюсов зависит от частоты дискретизации входной информации. И, наконец, глоттальный импульс имеет конечную продолжительность, что приводит к тому, что z-преобразование $g(n)$ есть полином по z вида

$$G(z) = \sum_{n=0}^{N_g} g(n) z^{-n} = B \sum_{n=1}^{N_g} (1 - z_n z^{-1}), \tag{6.153}$$

причем N_g меньше, чем N_p. Из (6.150) видно, что z-преобразование $h_v(n)$ есть

$$H_v(z) = R(z) V(z) G(z) \tag{6.154}$$

и что соответствующее преобразование Фурье имеет вид

$$H_v(e^{i\omega}) = R(e^{i\omega}) V(e^{i\omega}) G(e^{i\omega}). \tag{6.155}$$

Преобразование Фурье периодического сигнала $\tilde{x}(n)$ будет состоять из очень острых спектральных линий на частотах, кратных основной частоте.

Периодический сигнал $\tilde{x}(n)$ можно представить рядом Фурье:

$$\tilde{x}(n) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} \tilde{X}(k) e^{i\frac{2\pi}{N_p} kn}, \tag{6.156}$$

где

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right). \tag{6.157}$$

Подставив (6.156) и (6.157) в (6.1), легко показать, что

$$\tilde{X}_n(e^{i\omega}) = \frac{1}{N_p} \sum_{k=0}^{N_p-1} H_v\left(e^{i\frac{2\pi}{N_p} k}\right) W_n\left(e^{i\left(\omega - \frac{2\pi k}{N_p}\right)}\right), \tag{6.158}$$

где $W_n(e^{i\omega})$ — преобразование Фурье анализирующего окна $w(n-m)$. Характер кратковременного преобразования Фурье сильно зависит от длины и формы анализирующего окна. Вспомнив рис. 6.2 и 6.3, мы оказываемся с точки зрения оценивания параметров модели (отличных от основного тона) перед дилеммой. При узкополосном анализе (т. е. при широком анализирующем окне) информация об огибающей спектра скрывается за пиками основного тона; напротив, при широкополосном анализе (т. е. при узком анализирующем окне) пики формант окажутся сглаженными сверткой с преобразованием Фурье окна. Более того, из (6.158) следует, что, хотя $\tilde{x}(n)$ и периодична, $\tilde{X}_n(e^{i\omega})$ представляет собой функцию положения окна. В подходе, предложенном Мэтьюзом и другими, используется то, что коэффициенты ряда Фурье периодического сигнала, такого, как в (6.149), равны просто (6.157) и могут быть вычислены по одному периоду $\tilde{x}(n)$. Имеем

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right) = \sum_{n=0}^{N_p-1} \tilde{x}(n)\, e^{-i\frac{2\pi}{N_p} kn}, \quad 0 \leqslant k \leqslant N_p - 1. \tag{6.159}$$

Следовательно, выделив один период периодического сигнала, можно вычислить отсчеты $H_v(e^{i\omega})$ в N_p разноразнесенных точках. В случае, когда используется один период вокализованной речи вместо $\tilde{x}(n)$ в (6.159), результирующее кратковременное преобразование Фурье называют кратковременным преобразованием Фурье, синхронизированным сигналом основного тона. В общем этот подход к анализу вокализованной речи называют анализом, синхронизированным сигналом основного тона.

Этот подход полностью совместим с нашими рассуждениями о кратковременном анализе Фурье, правда, несколько измененными. Во-первых, моменты, в которых вычисляются значения $X_n(e^{i\omega})$, зависят от периода основного тона речи. Поскольку основной тон меняется со временем, нам необходима теперь неравномерная во времени дискретизация. Поскольку, кроме того, число получаемых значений частот зависит от периода основного тона, частота дискретизации в частотной области также оказывается зависящей от времени. Используемое в этом случае окно обычно выбирают прямоугольным, т. е. выделяется один период речевого колебания, затем он преобразуется с помощью (6.159). Как показано в задаче 6.15, это согласуется с (6.158), поскольку для прямоугольного окна шириной N_p нули $W(e^{i\omega})$ разнесены на интервалы, кратные $2\pi/N_p$. Следовательно, когда (6.158) вычисляется на частотах $\omega_k = 2\pi k/N_p$,

$$\tilde{X}_n\left(e^{i\frac{2\pi}{N_p} k}\right) = H_v\left(e^{i\frac{2\pi}{N_p} k}\right), \quad 0 \leqslant k \leqslant N_p - 1. \tag{6.160}$$

В рассматриваемом подходе не возникает трудностей, связанных с основным тоном в частотной области, так как они устраняются во временно́й. Следовательно, можно избежать временно́й

неопределенности узкополосного спектра, а смазывания в частотной области, присущего широкополосному анализу, можно избежать аккуратным оцениванием спектра только по N_p отсчетам.

6.6.2. Анализ полюсов и нулей модели с помощью анализа через синтез

Воспользовавшись спектром, синхронным с основным тоном, Мэтьюз и другие предложили итеративную процедуру вычисления параметров речи. Они пользовались эквивалентной моделью для передаточных функций сопротивления излучения, голосового тракта и глоттальных импульсов. Это привело к поправочному множителю с числом полюсов, большим, чем, вероятно, было бы нужно, если бы они пользовались соотношениями (6.151)—(6.154). Основной подход остается тем же независимо от конкретной функциональной формы модели речи, так что мы можем продолжить наше изложение, применив цифровую модель (по поводу используемых функций см. [23]).

Параметры $H_v(e^{i\omega})$ можно определить с помощью некоторой итеративной процедуры аппроксимации. Мэтьюз и другие ввели

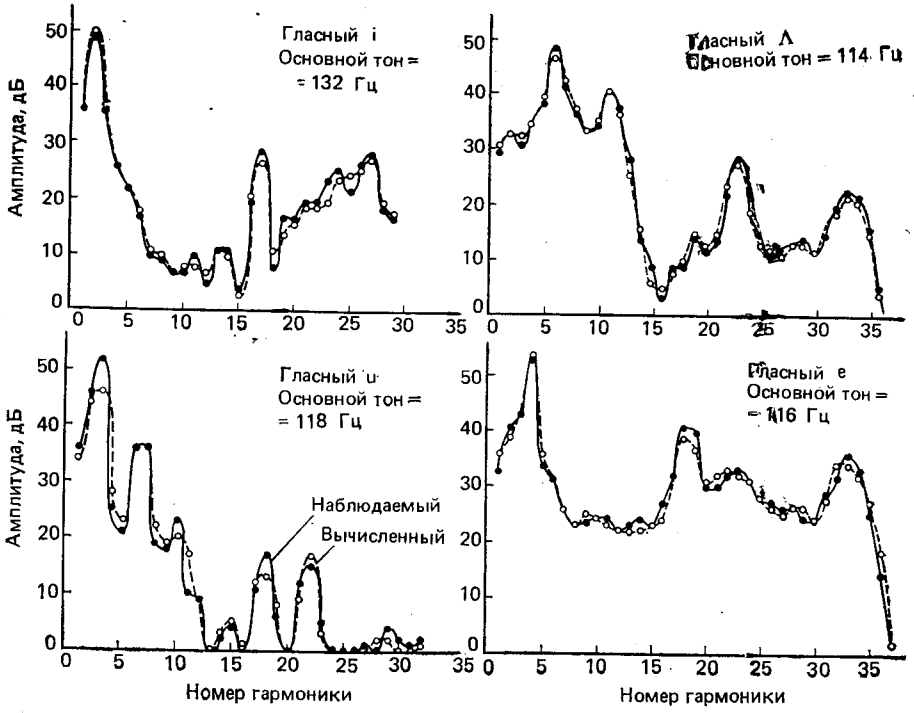

Рис. 6.43. Согласование спектра, синхронное с основным тоном: — наблюдаемый спектр;
— — — согласованный спектр [23]

ряд параметров, вычислили их значения на частотах $2\pi k/N_p$, а затем оценили функцию ошибки в виде

$$E=\sum_k Q(k)\left[\log\left|H_v\left(e^{i\frac{2\pi}{N_p}k}\right)\right|-\log\left|X_n\left(e^{i\frac{2\pi}{N_p}k}\right)\right|\right]^2, \quad (6.161)$$

где $Q(k)$ — весовая функция спектра, а $X_n\left(e^{i\frac{2\pi}{N_p}k}\right)$ — спектр речевого сигнала, синхронный с основным тоном. Параметры подбирались систематическим образом, так чтобы минимизировать функцию ошибки. Правила подбора полюсов $V(z)$ и нулей $G(z)$ рассмотрены в [23]. На рис. 6.43 показаны некоторые примеры согласования спектров, описанные в [23]. Когда ошибка минимизирована, значения полюсов $V(z)$ принимаются за оценки частот формант. Положение нулей дает информацию о глоттальных колебаниях.

6.6.3. Оценивание глоттальных колебаний, синхронное с основным тоном

Работа Мэтьюза, Миллера и Дэйвида связана, прежде всего, с распределением нулей. Были предприняты попытки связать разнесение нулей с формой глоттального импульса. В более поздней (неопубликованной) работе Миллера и Мэтьюза метод модифицирован так, чтобы получить оценки глоттального импульса. В этом случае модель имела вид

$$H_v(z)=R(z)\,G_f(z)\,V(z), \quad (6.162)$$

причем вклад в спектр глоттального колебания первоначально моделировался фиксированной передаточной функцией

$$G_f(z)=\frac{1}{(1-az^{-1})(1-bz^{-1})}. \quad (6.163)$$

Параметры $V(z)$ снова варьировались так, чтобы получить минимум критерия ошибки. Результирующее распределение полюсов служило оценкой частот формант. Чтобы получить форму глоттального колебания на анализируемом периоде речи, Миллер и Мэтьюз вычисляли величину

$$\tilde{G}(k)=\frac{X_n\left(e^{i\frac{2\pi}{N_p}k}\right)}{R\left(e^{i\frac{2\pi}{N_p}k}\right)V\left(e^{i\frac{2\pi}{N_p}k}\right)}, \quad 0\leqslant k\leqslant N_p-1. \quad (6.164)$$

Значения $\tilde{G}(k)$ при $0\leqslant k\leqslant N_p-1$ использовались в качестве коэффициентов Фурье глоттального импульса $g(n)$, который вычислялся с помощью обратного ДПФ:

$$\tilde{g}(n)=\frac{1}{N_p}\sum_{k=0}^{N_p-1}\tilde{G}(k)\,e^{i\frac{2\pi}{N_p}kn}. \quad (6.165)$$

Это выполнимо (поскольку $\tilde{g}(n)$ есть импульс конечной длительности) даже несмотря на то, что в общем случае наличия N_p отсчетов $H_v(e^{i\omega})$, получаемых синхронно с основным тоном, недостаточно для полного задания последовательности $h_v(n)$, которая, вообще говоря, длиннее чем N_p. Таким образом, с помощью модели речеобразования можно выделить из свертки компоненту конечной продолжительности. Этот метод с успехом применялся Розенбергом [24], изучавшим влияние формы глоттального импульса на качество звучания гласной. На рис. 6.44 показан пример речевого колебания и соответствующего глоттального колебания, выделенного указанной выше процедурой.

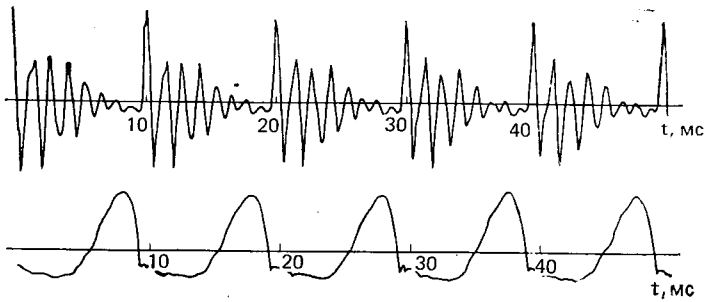

Рис. 6.44. Речь (вверху) и анализируемое колебание возбуждения (внизу) для гласной в слове «hod» [24]

Этот метод оценки глоттального импульса чувствителен к типу выбранной модели. В случаях, когда модель хорошо увязывается с речевым сигналом, как в случае установившихся гласных, получаются отличные результаты. В других ситуациях необходима более сложная модель. Другим фактором, влиявшим на результаты, был способ выделения из речевого колебания периода основного тона. Требуется большая осторожность при определении начала и конца периода. Речевой сигнал интерполировался с большей частотой дискретизации, чтобы облегчить поиск точного положения начала каждого цикла. Не удивительно, что это потребовалось, так как весьма мало вероятно, чтобы момент открытия (и смыкания) голосовой щели совпадал с моментом отсчета.

6.7. Системы анализа—синтеза

До сих пор в этой главе рассматривалась основная теория кратковременного анализа и синтеза Фурье. Было показано, что кратковременное преобразование Фурье может служить основой множества схем оценки параметров модели речеобразования. Однако мы еще не обсудили практического применения того факта, что речевой сигнал может быть точно восстановлен по его кратковременному Фурье-представлению. Этот факт лежит в основе схем кодирования речи, называемых вокодерами. Основная задача вокодеров состоит в цифровом представлении речи с гораздо меньшей скоростью, чем это возможно для схем непосредственного кодирования колебаний. В других случаях использование вокодеров позволяет удалять аддитивный шум и эффект реверберации, а также изменять основные параметры речи с преобразованием масштабов по времени и частоте.

В этом параграфе будут рассмотрены некоторые схемы кодирования речи, основанные на теоретических принципах § 6.1—6.3. Начнем с системы прямой реализации кратковременного анализа и синтеза Фурье, а затем обсудим такие системы, как полосный вокодер. Это не совпадает с историей развития вокодеров, однако при таком подходе станут яснее причины ухудшения качества, связанные с упрощением реализации.

6.7.1. Цифровое кодирование кратковременного преобразования Фурье

В § 6.1 и 6.2 было показано, что речь (практически, любой сигнал) можно точно представить набором полосовых каналов (рис. 6.45*а*). Центральные частоты ω_k и анализирующие окна

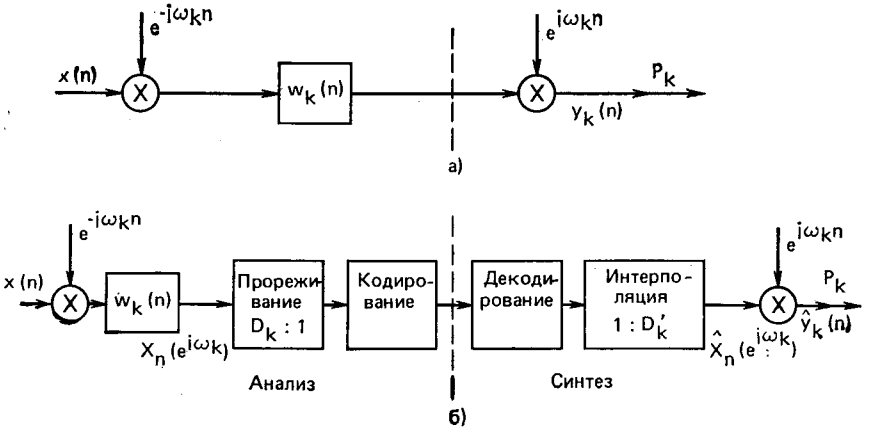

Рис. 6.45. Структурная схема кодирования одного канала процедуры анализ — синтез

$w_k(n)$ выбраны так, чтобы перекрыть нужную полосу частот, а комплексные константы $P_k = |P_k| e^{i\Phi_k}$ — чтобы общая характеристика суммы всех каналов была возможно ближе к идеальной со строго плоской амплитудно-частотной и линейной фазо-частотной характеристиками. В 6.13 было показано, что, поскольку $w_k(n)$ соответствует импульсной характеристике фильтра нижних частот, кратковременное преобразование Фурье на частотах ω_k может быть дискретизировано с частотой, меньшей, чем входной сигнал. Действительно, полное число необходимых отсчетов в секунду для $X_n(e^{i\omega_k})$ можно сделать равным частоте дискретизации входного сигнала. Поэтому чтобы снизить скорость вычислений, необходимую при реализации анализа, сигналы каналов дискретизируются с гораздо меньшей частотой, квантуются и кодируются для последующей передачи или хранения. Для одного канала это показано на рис. 6.45*б*. Операции анализа, показанные слева от пунктирной линии, выполняются модулятором, за которым следуют фильтр нижних частот, прореживатель, кодирую-

щее устройство. Когда $w_k(n)$ представляет собой последовательность конечной длины, операция прореживания просто совмещается с линейной фильтрацией, т. е. выход просто вычисляется для каждых D_k отсчетов на входе. Кодирование включает в себя квантование и собственно кодирование (см. гл. 5). При синтезе цифровое представление сначала декодируется, а затем вычисляется квантованная версия $X_n(e^{i\omega_k})$ с помощью интерполяции. Если опущены высокочастотные каналы, то можно при синтезе выходного колебания использовать частоту дискретизации, меньшую, чем исходная частота дискретизации входа. Следовательно, коэффициент интерполяции D_k может быть меньше коэффициента прореживания D_k. Квантованный канальный сигнал кратковременного преобразования Фурье $\hat{X}_n(e^{i\omega_k})$ модулирует комплексную синусоиду, образуя сигнал $\hat{y}_k(n)$, который складывается затем с другими:

$$\hat{y}(n)=\sum_{k=0}^{N-1} P_k\,\hat{y}_k(n). \qquad (6.166)$$

Чтобы иллюстрировать практические соображения по поводу такого кодирования речи, рассмотрим пример из [6]. Вычисление $X_n(e^{i\omega_k})$ реализуется с помощью БПФ (см. § 6.3). Поскольку программа БПФ требует, чтобы N было степенью двойки, вход дискретизировался с довольно необычной частотой 12195 отсч./с. Значение N равнялось 128, так что частотами анализа были $\omega_k=2\pi k/128$, что соответствует аналоговым частотам $F_k=(95,273k)$ Гц. Анализирующее окно $w(n)$ (одинаковое для всех каналов) имело импульсную характеристику КИХ-фильтров длиной 731 отсчет (с линейной фазо-частотной характеристикой). Оно было спроектировано методом частотной выборки [25]. Частотная характеристика фильтра приведена на рис. 6.46. Заметим,

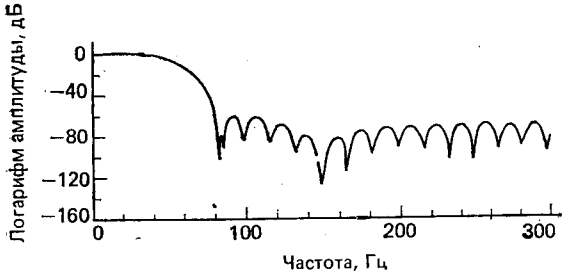

Рис. 6.46. Частотная характеристика анализирующего окна [6]

что выше 80 Гц затухание фильтра составляло не меньше 60 дБ. При подборе надлежащих комплексных констант P_k оказалось возможным получить общую характеристику, приведенную на рис. 6.47. Отметим, что $P_0=0$ и $P_k=0$ при $28<k<100$. Поскольку полоса перекрываемых синтезатором частот содержит частоты только до 2690 Гц, на выходе использовалась частота дискретизации 10004 отсч./с. (На выходе можно было бы использовать еще меньшую частоту дискретизации — примерно 6000 отсч./с,

если бы применялись соответственно более избирательные аналоговые фильтры.)

Результат проектирования показан на спектрограммах (рис. 6.48). На рис. 6.48*а* приведена фраза, поступающая на вход, а на рис. 6.48*б* — выходной сигнал системы анализа — синтеза, состоящей из 28 каналов (см. рис. 6.18*б*). Комплексные константы равны единице при $1\leqslant k\leqslant 28$ и $100\leqslant k\leqslant 127$ и нулю во всех остальных случаях, т. е. никакой специальной фазовой компенсации не применялось. Канальные сигналы дискретизировались с частотой Найквиста (т. е. 160 раз в секунду), что обеспечило точное восстановление на стадии синтеза. Квантование не проводилось, т. е. $X_n(e^{i\omega_k})$ представлялась с точностью 16 разрядов. Сравнение широкополосных спектрограмм (которые обладают хорошим разрешением по времени) выявляет эффект, согласующийся с видом импульсной характеристики рис. 6.25. Нечеткость спектрограммы, изображенной на рис. 6.48*б*, связана с задержкой энергии сигнала эха; искажения такого рода воспринимаются как реверберация. Прослушивание отчетливо выявило эффект «пустой бочки» в звучании сравниваемых фраз, соответствующих рис. 6.48*а* и *б*. Когда же фразы были должным образом подобраны (см. 6.2.1 и рис. 6.47), спектрограмма на выходе (рис. 6.48*в*) неотличима от спектрограммы на входе (рис. 6.48*а*); соответственно речевые сигналы на входе и выходе оказались неразличимыми при восприятии.

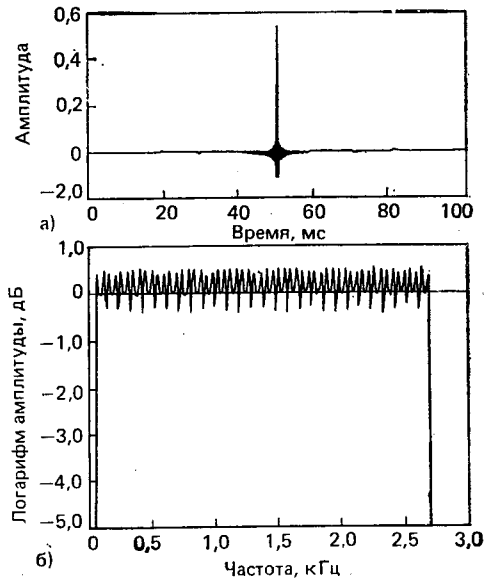

Рис. 6.47. Временна́я и частотная характеристики гребенки фильтров [6]

Уведившись в том, что можно действительно восстановить речевой сигнал по его кратковременному преобразованию Фурье, обратимся к способам кодирования канальных сигналов для цифровой передачи или хранения. В гл. 5 было установлено, что при кодировании любого колебания имеются два основных параметра: частота дискретизации и число бит, требуемых на отсчетах. Произведение этих двух величин дает скорость, которую, как правило, минимизируют, применяя минимальную допустимую частоту дискретизации и минимальное число бит на отсчет. В нашем случае общая информационная скорость представляет собой сумму скоростей для каждого из канальных сигналов в каналах.

Прежде чем перейти к квантованию, полезно рассмотреть эффекты снижения частоты дискретизации канальных сигналов. Вспомним, что действительная и мнимая части $X_n(e^{i\omega_k})$ представляют собой выходы фильтров нижних частот. Следовательно, в

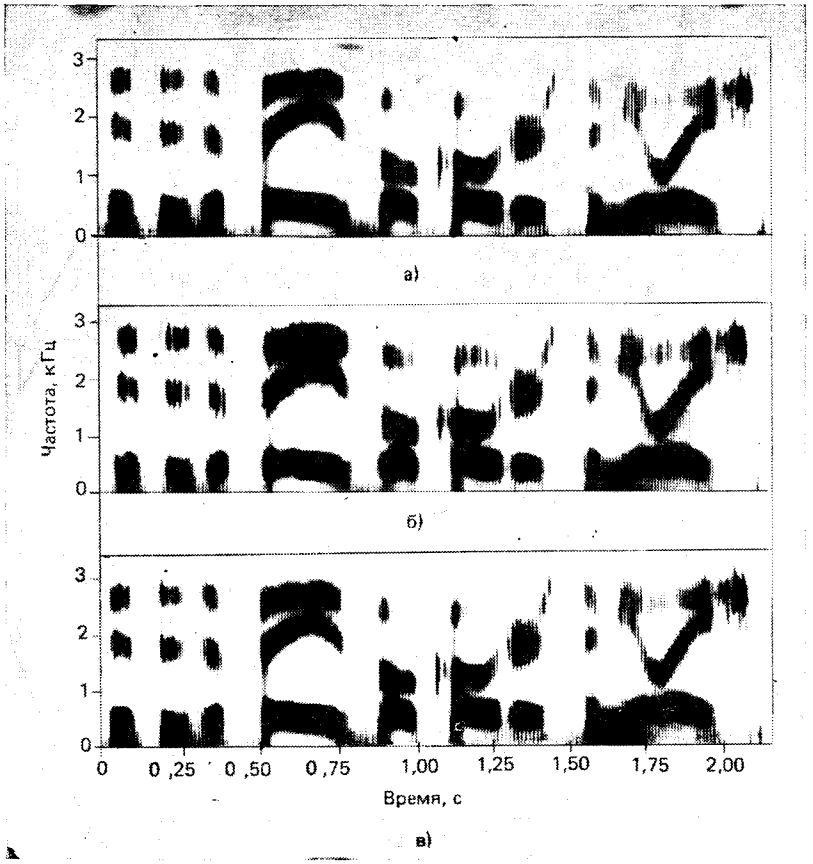

Рис. 6.48. Иллюстрация операции без квантования при $1/T_1 = 160$ Гц; *а*) речь на входе; *б*) речь на выходе без подстройки фазы; *в*) речь на выходе при наилучшей подстройке фазы [6]

рассматриваемом примере (частотная характеристика которого приведена на рис. 6.46) возникнут лишь пренебрежимо малые наложения при частотах дискретизации не меньше 160 Гц, поскольку характеристика фильтра имеет затухание по крайней мере 60 дБ на частотах выше 80 Гц. Если использовать меньшую частоту дискретизации без соответствующего сокращения полосы фильтров, то возникнут наложения во временно́й области. Если уменьшить полосу, не уменьшая разноса каналов, то синтезированная речь окажется более реверберирующей, поскольку характеристики отдельных каналов не будут перекрываться и в спектре синтезированной речи появятся «дыры». Уменьшить разнос каналов без увеличения их числа невозможно. Поэтому, пытаясь сни-

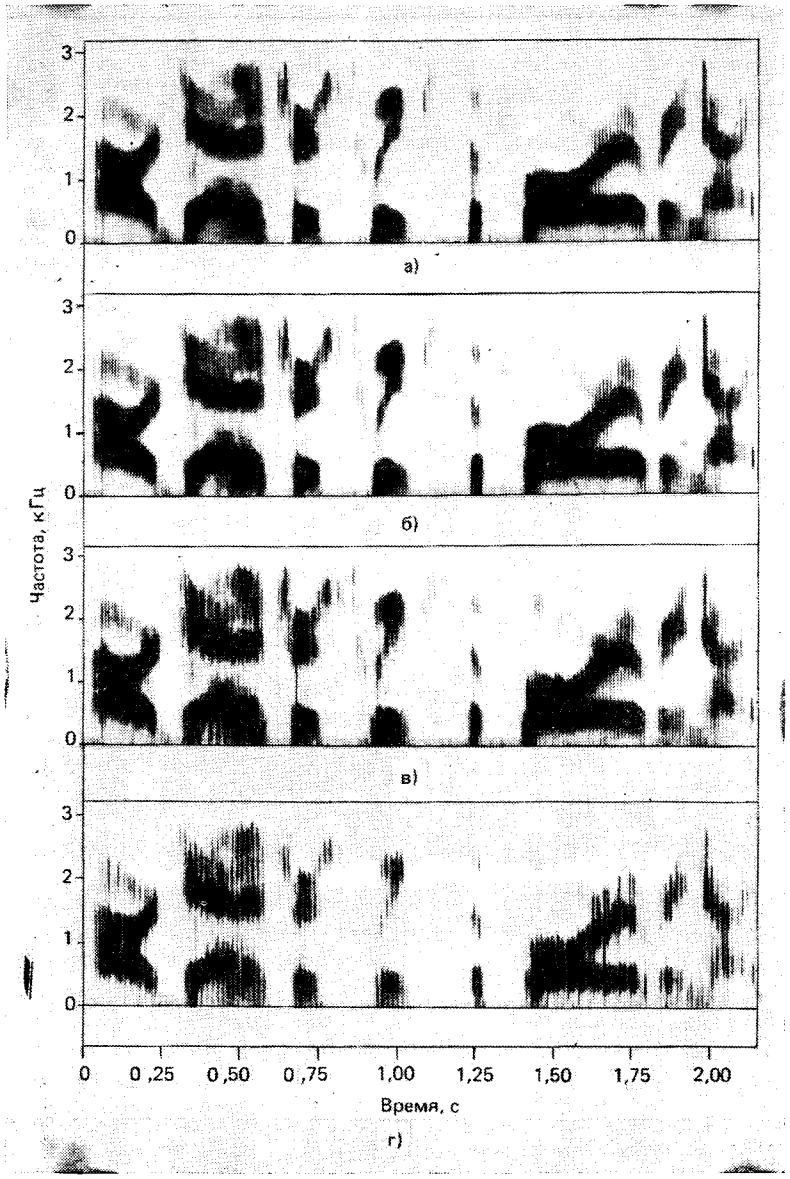

Рис. 6.49. Иллюстрация эффекта наложения при ИКМ (нижняя частота среза 80 Гц, без квантования) при $1/T_1 = 160$ Гц (*а*), 100 Гц (*б*), 80 Гц (*в*) и 60 Гц (*г*) [6]

зить информационную скорость уменьшением частоты дискретизации канальных сигналов, мы должны быть готовы к тому, чтобы смириться либо с искажениями из-за наложений во временно́й области, возникающими вследствие понижения частоты дискретизации, либо с повышенной реверберацией, вызванной сужением полос фильтров.

Оба эти эффекта показаны на рис. 6.49 и 6.50. Первый из них иллюстрирует возникновение эффекта наложения из-за слишком

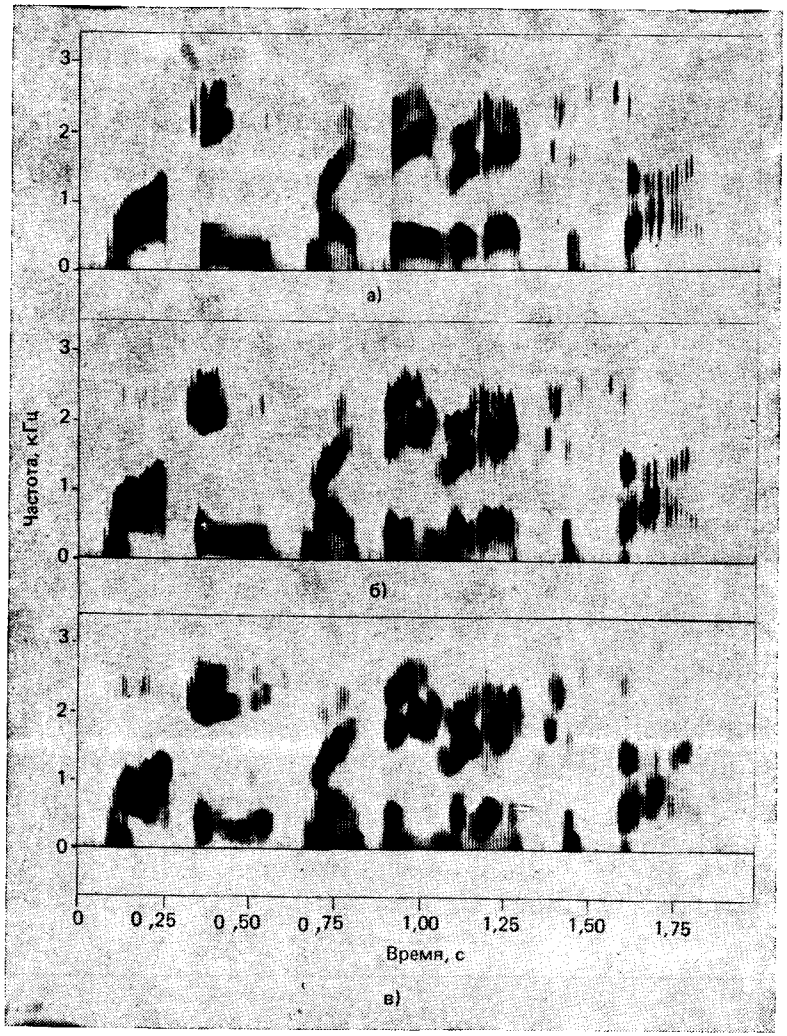

Рис. 6.50. Иллюстрация эффекта узкополосных анализирующих фильтров в ИКМ без квантования при $1/T_1 = 160$ Гц и нижней частоте среза 80 Гц (*а*), 53 Гц (*б*) и 11 Гц (*в*) [6]

малой частоты дискретизации кратковременного спектра. В случаях рис. 6.49*в* и *г* наблюдаются значительные искажения, которые менее заметны в случае рис. 6.49*б*. Сравнив рис. 6.49*г* со спектрограммой исходной речевой фразы (рис. 6.48*а*), видим, что в случаях серьезных искажений из-за наложений во временно́й области основной тон сильно искажается, тогда как частоты формант остаются почти без изменений. Рисунок 6.50 является иллюстрацией эффекта сужения полосы анализирующего фильтра при постоянном разнесении частот каналов. Во всех случаях частота дискретизации канальных сигналов равнялась 160 Гц. В случае, соответствующем рис. 6.50*а*, фильтр был таким же, как и при получении спектра, изображенного на рис. 6.46, т. е. выше частоты 80 Гц затухание фильтра было не меньше 60 дБ. На рис. 6.50*б* соответствующая частота среза равнялась 53 Гц, а на рис. 6.50*в* — 36 Гц. Как и ожидалось, спектрограммы на рис. 6.50*б* и *в* демонстрируют существенное ухудшение качества, что можно отнести за счет реверберации, вызванной расширением эффективной ширины окна. Кроме этого, можно видеть, что, хотя основной тон сигнала остался неизменным, траектории формант сильно пострадали из-за реверберации, вызванной узкополосными фильтрами. По разборчивости искажения, вызванные наложениями во временно́й области, предпочтительнее реверберции, вызванной сужением полос фильтров.

Для того чтобы определить информационную скорость, требуемую для представления речи с помощью кратковременного преобразования Фурье, необходимо выбрать схему квантования. Для квантования действительной и мнимой частей комплексных сигналов можно применить бо́льшую часть схем гл. 5. Два примера, рассматриваемые в [6], используют адаптивную дельта-модуляцию и ИКМ. В системе адаптивной дельта-модуляции [26] 28 каналов кодировались битом на отсчет. Полная скорость системы оказалась в 56 раз больше скорости (частоты) дискретизации канальных сигналов, поскольку нужно было кодировать и действительную и мнимую части канальных сигналов. Так как в системе адаптивной дельта-модуляции частота дискретизации в 5—10 раз превышает частоту Найквиста, можно ожидать, что для достижения хороших результатов потребуется скорость 20—30 кбит/с. На рис. 6.51 приведены примеры кодирования с помощью адаптивной дельта-модуляции для ряда скоростей. На рис. 6.51*а* полная скорость 28 кбит/с соответствует частоте дискретизации 500 отсч./с, на 6.51*б* полная скорость 21 кбит/с соответствует частоте дискретизации 375 отсч./с и на рис. 6.51*в* полная скорость 14 кбит/с соответствует частоте дискретизации 250 отсч./с. Из рис. 6.51 видно, что хорошее качество передачи достигается при скорости 28 кбит/с и что оно быстро ухудшается при меньших скоростях. Альтернативой кодированию АДМ может служить кодирование АИКМ, которое использовалось Крошьером [27] при неравномерном анализе и позволяло осуществлять пе-

редачу с хорошим качеством по четырем-пяти каналам при скорости около 16 кбит/с.

Как пример кодирования ИКМ, та же система с 28 каналами использовалась с частотой дискретизации канальных сигналов 100 отсч./с, т. е. допускался небольшой уровень искажений из-за

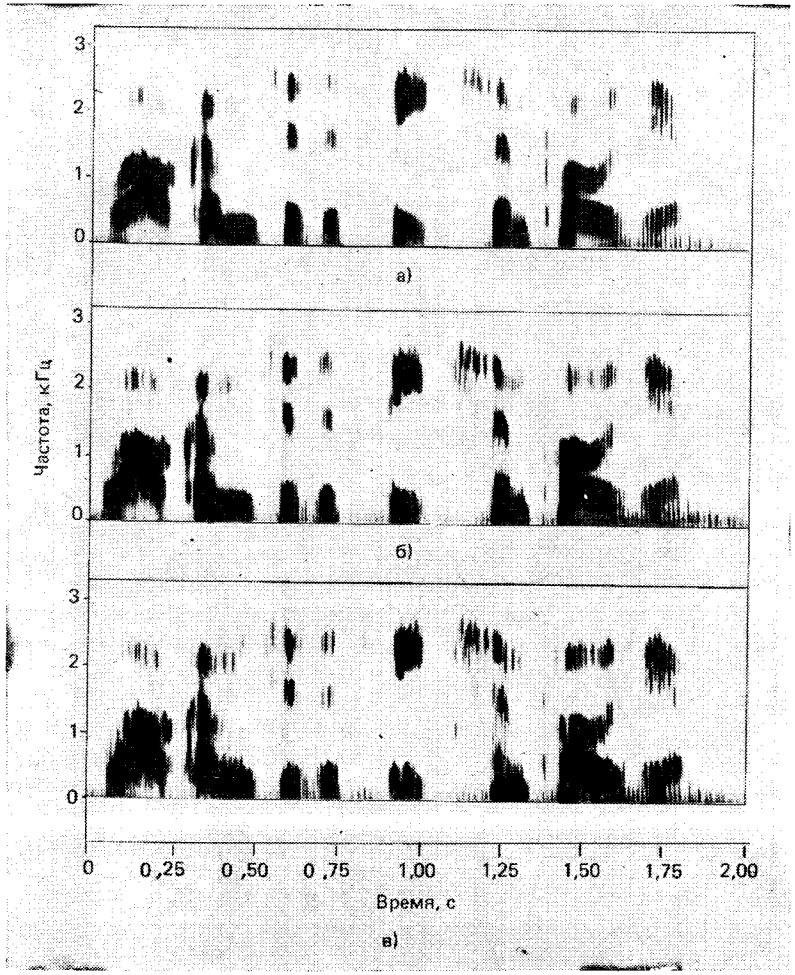

Рис. 6.51. Кодирование параметров спектра адаптивной дельта-модуляцией: *а*) 28 кбит/с, $1/T_1=500$ Гц; *б*) 21 кбит/с, $1/T_1=375$ Гц; *в*) 14 кбит/с, $1/T_1=250$ Гц [6]

наложения с тем, чтобы понизить частоту дискретизации. Квантование логарифмов модуля и фазы кратковременного преобразования Фурье было признано предпочтительным по сравнению с кодированием действительной и мнимой частей комплексных ка-

нальных сигналов. Для того чтобы воспользоваться слабой чувствительностью слуха на высоких частотах, сигналы низкочастотных каналов отображались точнее, чем сигналы высокочастотных каналов. При этом скорость передачи составляла 16 кбит/с; в каналах 1—10 использовалось 3 бит на логарифм амплитуды и 4 на фазу, а в каналах 11—28 соответственно 2 и 3 бит. На рис. 6.52*а* показана широкополосная спектрограмма входного речевого сигнала, на рис. 6.52*б* — результат кодирования со скоростью 16 кбит/с.

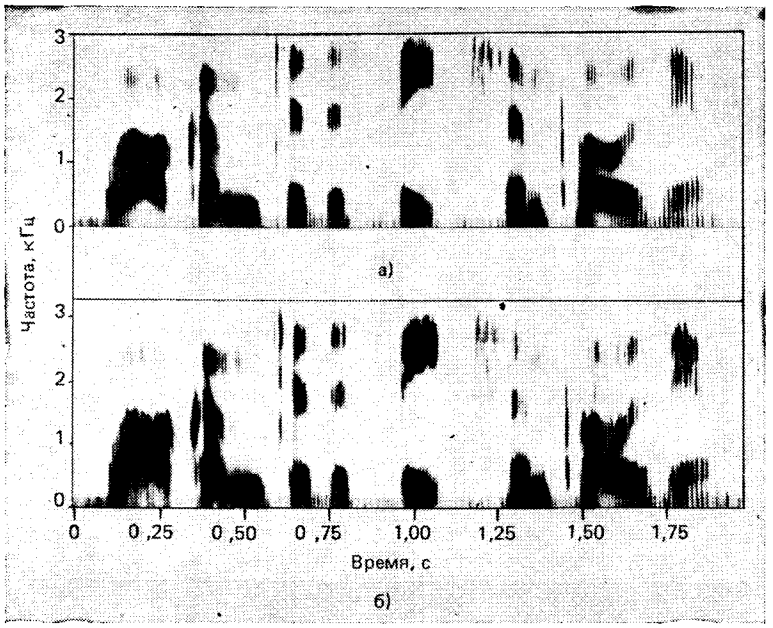

Рис. 6.52. Операции с квантованием:
а) речь на входе; *б*) $1/T_1=100$ Гц (полная скорость 16 кбит/с) [6]

Информационные скорости при использовании цифрового кодирования кратковременного преобразования Фурье сравнительно велики и сравнимы со скоростями непосредственного кодирования речевых колебаний методами адаптивного квантования. Сложность кратковременного преобразования Фурье, конечно, гораздо выше, чем в большинстве систем кодирования непрерывных сигналов. Основным преимуществом представления с помощью кратковременного преобразования Фурье является наличие гибкости, с которой можно манипулировать параметрами речевого сигнала. Это станет очевидным из последующих рассуждений.

Соберем воедино все, что уже изложено относительно схем кодирования речи, основанных на кратковременном преобразовании Фурье. Во-первых, при дискретизации канальных сигналов

с достаточно большой частотой и без квантования отсчетов (достаточно 12 бит/отсч.) можно достичь безукоризненного на слух воспроизведения речи. Скорость, необходимая для передачи такого представления, однако, довольно велика. Действительно, в примере, который был приведен ранее, высококачественное воспроизведение сигнала с полосой частот 3 кГц требовало гораздо меньшей скорости (около 100 бит/с). Скорость передачи можно снизить двумя способами. Во-первых, допустимо более грубо квантовать канальные сигналы и снижать частоту их дискретизации. В этом случае удается достигнуть скорости передачи 16 кбит/с при незначительном ухудшении качества восприятия. Во-вторых, можно использовать свойства речи, удалив часть избыточности сигнала. Ухудшение качества восприятия речи получается, кроме всего прочего, из-за того, что для упрощения реализации системы анализа—синтеза вводится ряд приближений. Такого рода ухудшение воспринимается как изменение разборчивости, отличное от искажений в системах прямого кодирования, представимых, как правило, аддитивным (возможно коррелированным с сигналом) шумом. Следовательно, измерения отношения сигнал/шум (см. гл. 5) не имеют смысла в системах типа вокодеров. По этой причине приходится описывать качество восприятия речи в вокодерах, сравнивая спектрограммы, или по субъективным оценкам искажений, воспринимаемых слушателями.

6.7.2. Фазовый вокодер [1]

Фазовый вокодер представляет собой интересный новый подход к анализу, основанному на кратковременном спектре [28]. Чтобы понять, как работает эта система, рассмотрим отклик в одном канале. Для этого удобно представить систему, изображенную на рис. 6.45*а*, через действительные операции, как это сделано на рис. 6.53. Вспомнив, что обычно выбираются $\omega_{N-k}=2\pi-\omega_k$ и $P_k=P^*_{N-k}$, видим, что мнимые части сокращаются и остаются только действительные, которые легко представить в виде

$$R_e[P_k y_k(n)] = |P_k|\left|X_n\left(e^{i\omega_k}\right)\right|\cos[\omega_k n+\theta_n(\omega_k)+\gamma_k]. \qquad (6.167)$$

Следовательно, общий сигнал на выходе образуется как сумма сигналов. Такие сигналы можно интерпретировать как дискретные косинусоидальные колебания, модулированные по амплитуде и по частоте кратковременным преобразованием Фурье сигналов в канале. Значение $|P_k|$ равно, вообще говоря, единице или нулю, в зависимости от того, участвует ли в суммировании соответствующий канал. Фазовые константы введены для того, чтобы добиться максимально-плоской общей характеристики.

[1] Фазовый вокодер был предложен и исследован Фланаганом и Голденом [28]. Результаты настоящего раздела основаны на этой работе.

Вводя понятие мгновенной частоты, получим полезную интерпретацию (6.167). Удобно рассмотреть непрерывное зависящее от времени преобразование Фурье:

$$X_a(t,\ \Omega_k) = |X_a(t,\ \Omega_k)|\, e^{i\theta_a(t,\ \Omega_k)} = \qquad (6.168а)$$

$$= a_a(t,\ \Omega_k) - i\, b_a(t,\ \Omega_k), \qquad (6.168б)$$

где

$$|X_a(t,\ \Omega_k)| = [a_a^2(t,\ \Omega_k) + b_a^2(t,\ \Omega_k)]^{1/2} \qquad (6.169а)$$

и

$$\theta_a(t,\ \Omega_k) = -\mathrm{tg}^{-1}\left[\frac{b_a(t,\ \Omega_k)}{a_a(t,\ \Omega_k)}\right]. \qquad (6.169б)$$

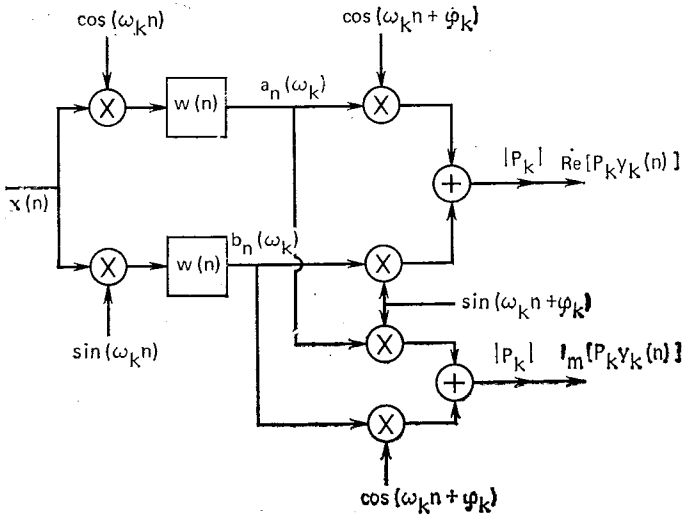

Рис. 6.53. Реализация одного канала фазового вокодера

Это зависящее от времени преобразование Фурье с непрерывным временем можно определить как

$$X_a(t,\ \Omega_k) = \int_{-\infty}^{\infty} x_a(\tau)\, w_a(t-\tau)\, e^{-i\Omega_k \tau}\, d\tau, \qquad (6.170)$$

где $x_a(\tau)$ — речевой сигнал с непрерывным временем, а $w_a(\tau)$ — анализирующее окно с непрерывным временем, или, что то же, импульсная характеристика аналогового фильтра нижних частот. Величина

$$\dot{\theta}_a(t,\ \Omega_k) = \frac{d\,\theta_a(t,\ \Omega_k)}{dt}, \qquad (6.171)$$

называемая фазовой производной, представляет собой мгновенную девиацию частоты от центральной частоты Ω_k в k-м канале.

Производную фазы можно выразить через $a_a(t, \Omega_k)$ и $b_a(t, \Omega_k)$:

$$\dot{\theta}_a(t,\ \Omega_k) = \frac{b_a(t,\ \Omega_k)\,\dot{a}_a(t,\ \Omega_k) - a_a(t,\ \Omega_k)\,\dot{b}_a(t,\ \Omega_k)}{a_a^2(t,\ \Omega_k) + b_a^2(t,\ \Omega_k)}, \qquad (6.172)$$

где точки сверху означают дифференцирование по времени. В случае обработки сигнала с дискретным временем предполагается, что $x_a(t)$ и $X_a(t, \Omega_k)$ ограничены по частоте и что $X_n(e^{i\omega_k})$ представляет собой дискретизованное кратковременное преобразование Фурье с непрерывным временем:

$$X_n\left(e^{i\omega_k}\right) = X_a(nT,\ \omega_k/T). \qquad (6.173)$$

Аналогичным образом, фазовая производная $X_n(e^{i\omega_k})$ определяется как дискретизованная версия $\dot{\theta}_a(t, \Omega_k)$:

$$\dot{\theta}_n(\omega_k) = \frac{b_n(\omega_k)\,\dot{a}_n(\omega_k) - a_n(\omega_k)\,\dot{b}_n(\omega_k)}{a_n^2(\omega_k) + b_n^2(\omega_k)}. \qquad (6.174)$$

В этом случае $\dot{a}_n(\omega_k)$ и $\dot{b}_n(\omega_k)$ предполагаются последовательностями, полученными при дискретизации соответствующих производных с непрерывным временем и ограниченных по частоте. Эти сигналы производных можно получить цифровой фильтрацией последовательностей $a_n(\omega_k)$ и $b_n(\omega_k)$ (см. задачу 6.16).

Чтобы понять, почему представляют интерес сигналы производной фазы, рассмотрим случай, когда центральные частоты каналов мало разнесены. В частности, возьмем случай, когда основной тон постоянен и всего одна гармоника основной частоты попадает в полосу k-го канала. Легко видеть, что $|X_n(e^{i\omega_k})|$ отражает медленно меняющуюся амплитудно-частотную характеристику голосового тракта на частоте вблизи ω_k. Производная фазы будет константой, равной отклонению частоты гармоники от центральной частоты. Если теперь характеристика голосового тракта и основной тон медленно меняются, так, как это бывает при обычной речи, допустимо предположить, что медленно меняются и амплитудный спектр, и производная фазы. Действительно, для восприятия эффекты наложения при дискретизации переменного амплитудного спектра и производной фазы окажутся менее заметными, чем аналогичные эффекты при дискретизации действительной и мнимой частей кратковременного преобразования Фурье [28].

При синтезе $\theta_a(t, \Omega_k)$ получается из $\dot{\theta}_a(t, \Omega_k)$ интегрированием

$$\theta_a(t,\ \Omega_k) = \int_{t_0}^{t} \dot{\theta}_a(\tau,\ \Omega_k)\,d\tau + \theta_a(t_0,\ \Omega_k). \qquad (6.175)$$

Из приведенного равенства следует, что $\theta_n(\omega_k)$, представляющее собой дискретизованную версию $\theta_a(t, \Omega_k)$, окажется более сглаженной, чем $\dot{\theta}_n(\omega_k)$. Поэтому можно предположить, что $\theta_n(\omega_k)$ допустимо дискретизовать с еще меньшей частотой, чем $\dot{\theta}_n(\omega_k)$. Однако мы пренебрегаем тем, что величина $\theta_n(\omega_k)$ не ограничена и, следовательно, непригодна для квантования. (В этом нетрудно убедиться, рассмотрев случай постоянного основного тона.) Ограниченную фазу можно получить, вычисляя главное значение, т. е. ограничив $\theta_n(\omega_k)$ значениями в интервале от 0 до 2π или от $-\pi$ до π. Главное значение фазы окажется, к сожалению, «разрывным» (т. е. главное значение $\theta_a(t, \Omega_k)$ будет разрывной функцией t), а следовательно, не будет сигналом, спектр которого ограничен по частоте в фильтре нижних частот. Разрывность главного значения фазы не означает того, что фазу нельзя квантовать, поскольку единственное, что требуется, это восстановить при подходящей частоте дискретизации соответствующие действительную и мнимую части $X_n(e^{i\omega_k})$. Поэтому частота дискретизации для $\theta_n(\omega_k)$ должна быть не меньше частот, необходимых для $a_n(\omega_k)$ и $b_n(\omega_k)$. В действительности именно главное значение фазы квантовалось в 6.7.1.

Применение производной фазы в системе анализа—синтеза помимо достоинства — гладкости — обладает и рядом недостатков, что видно из (6.175), так как для восстановления $\theta_n(\omega_k)$ по $\dot{\theta}_n(\omega_k)$ необходимы начальные условия. Обычно подобные начальные условия нам неизвестны, а положив произвольно начальную фазу равной нулю, мы получим систематические ошибки по фазе φ_k. Это может привести к существенному отклонению общей характеристики системы анализа—синтеза от идеальной (с плоской амплитудно-частотной и линейной фазо-частотной характеристиками), а синтезированная речь будет звучать с сильными искажениями типа реверберации.

На рис. 6.54 изображен анализатор вокодера, основанный на использовании амплитудного спектра и производной фазы. Здесь

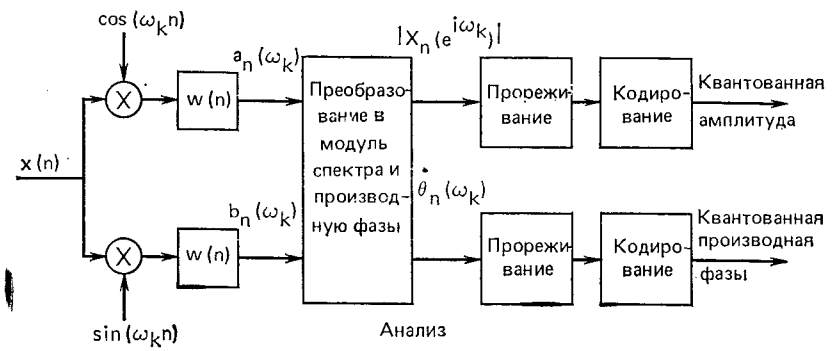

Рис. 6.54. Отдельный канал анализатора фазового вокодера

приведен один канал анализатора с частотой $0 < \omega_k < \pi$. Все остальные каналы строятся точно так же, если не считать возможных различий в прореживании и интерполяции. Операции, требуемые для преобразования $a_n(\omega_k)$ и $b_n(\omega_k)$ в $|X_n(e)|$ и $\theta_n(\omega_k)$, приведены на рис. 6.55. Один из подходов к синтезу по амплитуд-

ному спектру и производной фазы приведен на рис. 6.56. Операции, требуемые для преобразования сигналов амплитудного спектра и производной фазы в действительную и мнимую части, показаны на рис. 6.57. Видно, что сигнал производной фазы нуж-

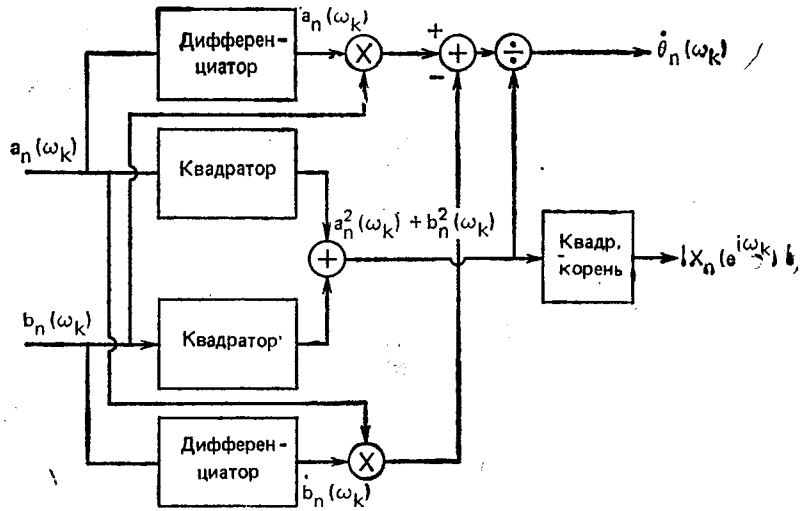

Рис. 6.55. Переход от a и b к $\dot{\theta}$ и $|X(e^{i\omega})|$

но проинтегрировать, чтобы получить сигнал фазы. Косинус и синус фазового угла умножаются затем на значение амплитудного спектра, что и дает действительную и мнимую части. На

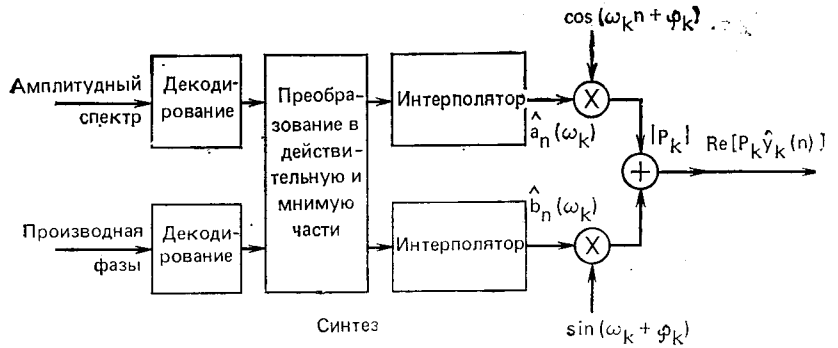

Рис. 6.56. Канал синтезатора фазового вокодера

рис. 6.58 приведен другой подход к синтезу, он позволяет избежать преобразований. В этом случае интерпретируются сигналы амплитудного спектра и производной фазы; результирующие последовательности амплитудного спектра и фазы используются для модуляции синусоиды по амплитуде и фазе. Следовательно, вместо преобразователя амплитудного спектра и производной фазы в

действительную и мнимую части требуется фазовый модулятор. Ясно, что, если реализация такого цифрового фазового модулятора не слишком сложна, схема рис. 6.58 значительно проще схемы рис. 6.56 и 6.57.

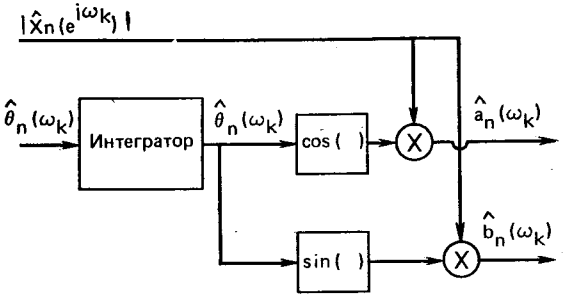

Рис. 6.57. Переход от $|X(e^{i\omega})|$ и θ к a и b

Тщательное исследование методов дискретизации и квантования амплитудного спектра и производной фазы в фазовом вокодере было проведено Карлсом [29]. В этом исследовании был реализован вокодер с 28 каналами и разнесением между ними в

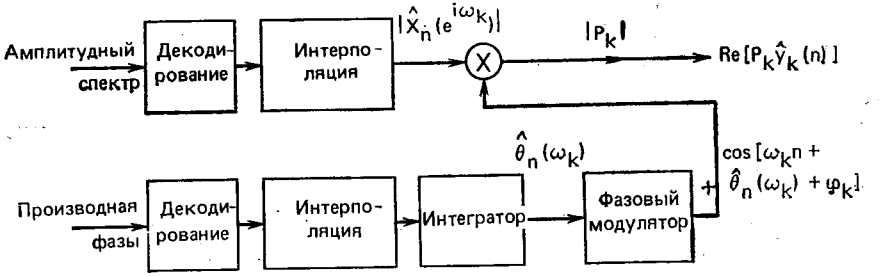

Рис. 6.58. Другая форма синтеза

100 Гц. Для производной фазы использовались линейные квантователи, а для амплитудного спектра — логарифмические. Распределение бит по каналам было неравномерным — больше бит отводилось под низкочастотные каналы и меньше — под высокочастотные. Кроме того, большее число бит отводилось под производную фазы в сравнении с амплитудным спектром. Полученная скорость составляла 7,2 кбит/с при дискретизации сигналов амплитудного спектра и производной фазы с частотой всего 60 раз/с, при этом на сигналы амплитудного спектра отводилось 2 бита в низкочастотных каналах и 1 бит — в высокочастотных каналах, и на сигналы производной фазы: 3 бита — в низкочастотных каналах и 2 бита — в высокочастотных. Неофициальные тесты показали, что такое кодирование речи обеспечивает качество восприятия, сравнимое с качеством восприятия в системе логарифмической ИКМ со скоростью передачи в 2—3 раза выше.

Отметим характерную особенность фазовых вокодеров и вокодеров вообще — большую гибкость в манипуляции параметрами речевого сигнала. По сравнению с представлением в виде колебаний, где изменения в речевом сигнале отображаются некоторой последовательностью чисел, в вокодере сигнал представляется параметрами, теснее связанными с параметрами речеобразования. Например, как уже отмечалось, в случае фазового вокодера можно считать, что амплитуда комплексного канального сигнала несет главным образом сведения о передаточной характеристике голосового тракта, тогда как сигнал производной фазы — сведения о возбуждении. На рис. 6.58 показан простой способ, которым с помощью фазового вокодера можно изменить основные параметры речи. Допустим, что сигнал производной фазы установлен равным нулю, так что выходной сигнал образован произведением модуля кратковременного преобразования Фурье и косинуса фиксированной частоты ω_k. В случае равноразнесенных каналов общий выходной сигнал будет похожим на периодический сигнал с основной частотой, численно равной интервалу частот, на который разнесены каналы. Выходной сигнал не будет строго периодическим, поскольку амплитудный спектр медленно меняется со временем. При таком синтезе выходной сигнал будет иметь монотонное звучание. Если же сигнал производной фазы изменять хаотически, можно ожидать, что синтезированная речь будет похожа на шепот.

Другим более полезным использованием гибкости, присущей системам фазового вокодера, является преобразование временно́го и частотного масштабов сигнала, как это описано в [28]. Обращаясь опять к рис. 6.58, вспомним, что мгновенная частота косинуса равна $[\omega_k+\dot{\theta}_n(\omega_k)]$, следовательно, сигнал с уменьшенной частотой можно получить, просто разделив ω_k и $\dot{\theta}_n(\omega_k)$ на константу q. Если синтезировать каждый канал таким образом, то в результате получится сигнал, сжатый по частоте в q раз. Частотный масштаб результирующего сигнала можно восстановить, записав сигнал на одной скорости и воспроизводя его со скоростью, большей в q раз. Другой способ состоит в использовании преобразователя код—аналог с частотой синхронизации, большей в q раз частоты дискретизации на входе. В любом случае сжатие временно́го масштаба компенсирует сжатие частотного масштаба при синтезе. В результате получается сигнал с обычным частотным масштабом, но со сжатым временны́м. Аналогичными операциями можно растянуть масштаб времени. В этом случае центральная частота ω_k и фазовый угол умножаются на множитель q, а полученный растянутый масштаб частот восстанавливается воспроизведением записанного сигнала с меньшей скоростью. В результате получается растянутый во времени сигнал с обычным частотным масштабом. На рис. 6.59 [28] показано, как описанным способом образуется речь, сжатая и растянутая во времени в $q=2$ раз.

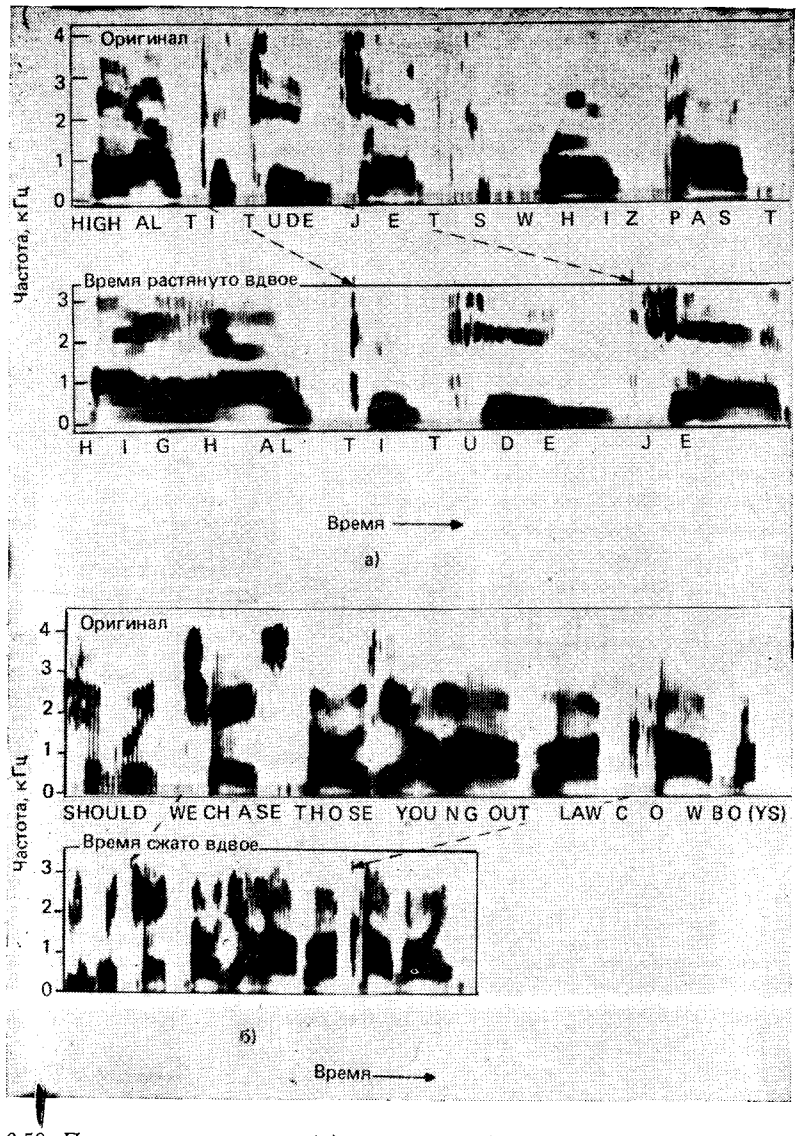

Рис. 6.59. Примеры растяжения (*а*) и сжатия (*б*) временно́го масштаба в фазовом вокодере [28]

6.7.3. Полосный вокодер

Полосный вокодер, изобретенный Дадли [30], представляет собой самое первое устройство для кодирования речи. Он во многих отношениях аналогичен системам, которые уже рассмотрены в этом параграфе. Основные отличия состоят в том, что полосный вокодер теснее связывает схемы анализа и синтеза с моделью

речи, а также в том, что в схемы кратковременного анализа и синтеза Фурье введен для упрощений ряд приближений. Чтобы понять, как связан полосный вокодер с рассмотренными представлениями кратковременного преобразования Фурье, вернемся к (6.167). Вспомним, что это выражение отражает вклад k-го канала в общий выход. Мы интерпретировали это выражение как представление для косинуса с номинальной центральной частотой ω_k, модулированного по фазе и амплитуде, причем по амплитуде — амплитудой (модулем) кратковременного преобразования Фурье, а по фазе — сигналом, соответствующим фазовому углу (аргументу) кратковременного преобразования Фурье. Каждый канал анализа можно представлять себе как полосовой фильтр с центральной частотой ω_k. Отсюда можно предположить, что амплитуду кратковременного преобразования Фурье можно получить детектором огибающей на выходе полосового фильтра с центральной частотой ω_k. Это показано на рис. 6.60, где за полосовым фильтром с импульсной характеристикой $w(n)\cos(\omega_k n)$ следует двухполупериодный выпрямитель (блок амплитуды) и снова фильтр нижних частот.

Рис. 6.60. Метод аппроксимации кратковременного спектра

Двухполупериодный выпрямитель и фильтр нижних частот служат детектором огибающей. Такая система представляет собой базовый блок полосного вокодера. Анализатор состоит из набора таких каналов, причем частоты анализа распределены по частотному диапазону речи. Однако сигнал речи не может быть представлен только амплитудным спектром — сигнал производной фазы содержит информацию о возбуждении. Если приравнять сигнал производной фазы нулю, то результирующая речь окажется полностью вокализованной и монотонной. Для того чтобы отразить должным образом возбуждение, полосной вокодер имеет дополнительное анализирующее устройство определяющее тип возбуждения — вокализованное или невокализованное, и, если вокализованное, основную частоту речевого сигнала. Эти параметры дискретизуются и квантуются для передачи или хранения в цифровой системе. Полностью анализатор полосного вокодера приведен на рис. 6.61. В системе нужно предусмотреть значительные изменения для синтеза сигнала полосного вокодера (рас. 6.62).

Основной принцип синтеза в полосном вокодере можно сформулировать просто. Сигналы каналов отражают амплитуду вклада каждого из каналов, тогда как сигнал возбуждения управляет структурой в заданном канале. Сигнал тон/шум задает нужный тип возбуждения — случайный шум для невокализованной речи

или периодические импульсы в случае вокализованной речи. Основная частота импульсного генератора управляется сигналом основного тона. Следовательно, общий выходной спектр строится по отдельным сегментам, в которых амплитуда в заданной полосе

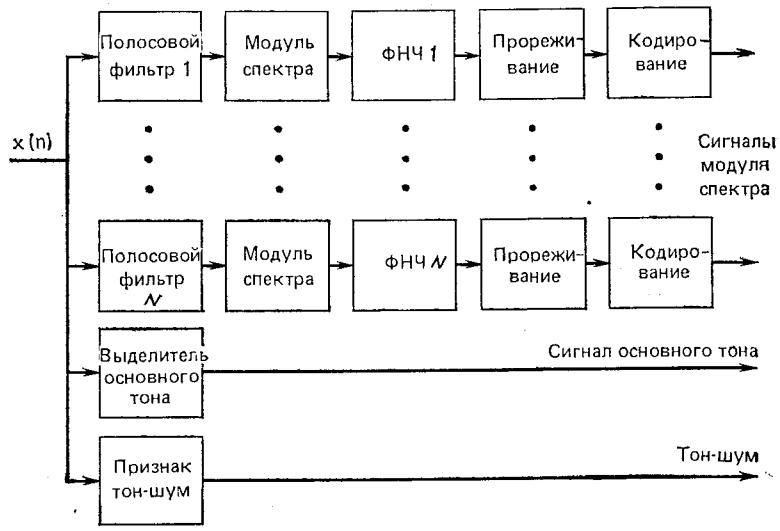

Рис. 6.61. Структурная схема анализатора полосного вокодера

частот, грубо говоря, постоянна. Действительно, амплитуда в каждой полосе частот сохраняет форму, обусловленную частотно-избирательными свойствами используемых при синтезе полосовых

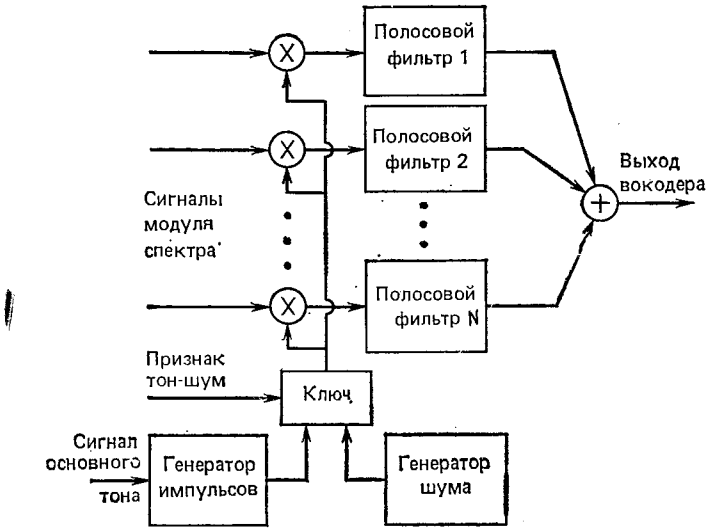

Рис. 6.62. Структурная схема синтезатора полосного вокодера

фильтров. Когда возбуждение вокализованно, выходной сигнал образуется из смежных полос частот, причем тонкая структура спектра характеризуется периодичностью.

Для невокализованного возбуждения спектр непрерывно меняется в каждой из полос частот. Результатом будет речь с сильной реверберацией, что вызвано полным отсутствием контроля над интерференцией смежных частотных полос. На рис. 6.63 [31]

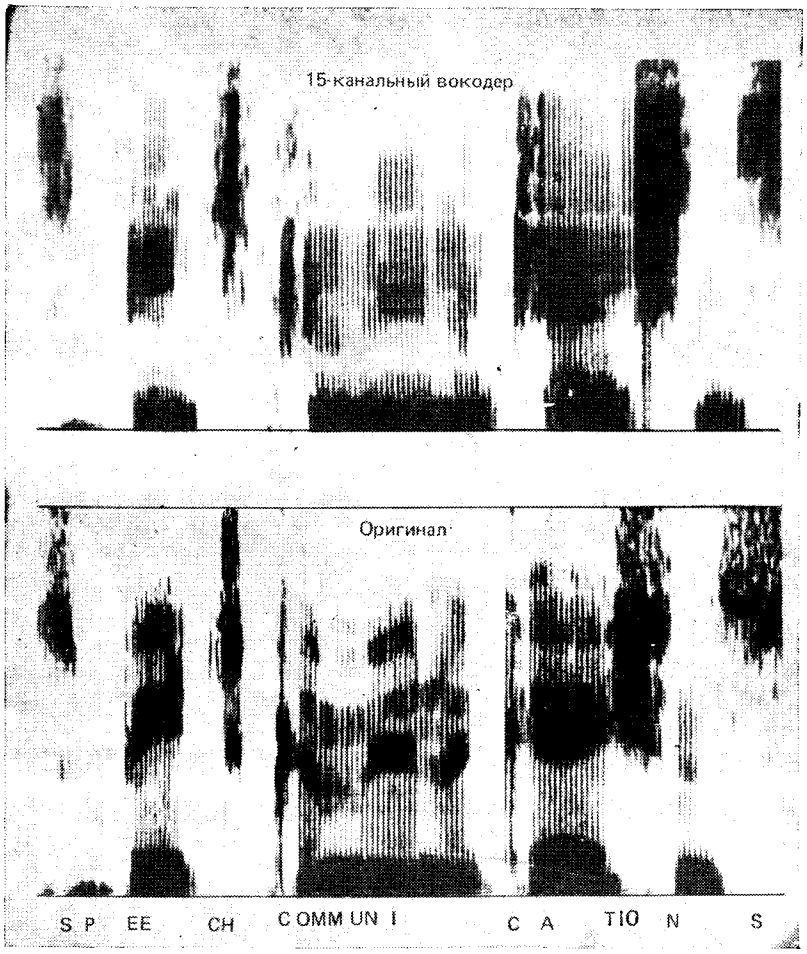

Рис. 6.63. Пример сигнала в 15-канальном вокодере [31]

сравниваются спектрограммы речевого сигнала на входе 15-канального полосного вокодера со спектрограммой соответствующего выхода. Видно, что из-за грубого разнесения каналов формантная структура сильно квантованна, причем частоты формант в

некоторых случаях весьма значительно изменены. Устройства, изображенные на рис. 6.61 и 6.62, обеспечивают значительное снижение скорости передачи информации с сопутствующим ростом искажений.

Полосные вокодеры обычно работают со скоростями 1200—9600 бит/с, причем около 600 бит/с отводится под информацию об основном тоне и типе возбуждения, а оставшиеся — под канальные сигналы. Полосный вокодер еще больше, чем фазовый, допускает модификацию речевого сигнала, поскольку информация о возбуждении и голосовом тракте представляется порознь. Легко понять, например, как можно изменить информацию об основном тоне независимо от информации о голосовом тракте. Если, например, импульсный генератор всегда выдает одну и ту же основную частоту, т. е. информация об основном тоне не используется вовсе, результатом будет монотонная речь. Если возбуждение импульсным генератором не используется, а вместо этого возбуждение всегда представляет собой случайный шум, результатом будет шепот. Используя полосный вокодер, можно также получить независимые изменения временно́й и частотной шкал. Это достигается простым масштабированием центральных частот полосовых фильтров и периода основного тона.

Выделение информации об основном тоне и типе возбуждения обеспечивает основной вклад в снижение скорости, достигаемое в полосном вокодере. Это, однако, является и слабым местом таких систем, поскольку выделение основного тона представляет собой трудную задачу. Следовательно, фазовый вокодер или, точнее, представления, вытекающие из теории, изложенной в предыдущих параграфах этой главы, обладают тем преимуществом, что не требуют отслеживания основного тона.

Полосный вокодер был предметом интенсивных исследований как вследствие традиции, так и вследствие большого числа факторов, влияющих на качество его работы (например, число и тип фильтров, их разнос и др.). Результатом этих исследований оказалось большое число остроумных решений реализации полосных вокодеров [31—34].

6.8. Заключение

В этой главе проведен анализ кратковременного преобразования Фурье применительно к речевым сигналам. Показано, как такое представление может быть эффективно использовано для оценки основных параметров речи, таких, как период основного тона и частота формант. Рассмотрено также применение кратковременного преобразования Фурье при проектировании фазовых и полосных вокодеров.

Задачи

6.1. Пусть кратковременное преобразование Фурье задано в виде $X_n(e^{i\omega}) = a_n(\omega) - ib_n(\omega) = |X_n(e^{i\omega})|e^{i\theta_n(\omega)}$. Доказать, что если $x(n)$ действительна, то:

а) $a_n(\omega) = a_n(2\pi - \omega) = a_n(-\omega)$;

б) $b_n(\omega) = -b_n(2\pi - \omega) = -b_n(-\omega)$;

в) $|X_n(e^{i\omega})| = |X_n(e^{i(2\pi-\omega)})| = |X(e^{-i\omega})|$;
г) $\theta_n(\omega) = -\theta_n(2\pi-\omega) = -\theta_n(-\omega)$.

6.2. Пусть кратковременное преобразование Фурье сигнала задано соотношением $X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m}$.

Показать, что справедливы следующие свойства:
а) линейность[1] — если $v(n) = x(n) + y(n)$, то $V_n(e^{i\omega}) = X_n(e^{i\omega}) + Y_n(e^{i\omega})$;
б) свойство сдвига — если $v(n) = x(n-n_0)$, то $V_n(e^{i\omega}) = X_{n-n_0}(e^{i\omega})\,e^{-i\omega n_0}$;
в) масштабирование — если $v(n) = \alpha x(n)$, то $V_n(e^{i\omega}) = \alpha X_n(e^{i\omega})$;
г) экспоненциальное взвешивание — если $v(n) = a^n x(n)$, то $V_n(e^{i\omega}) = X_n(a^{-1}e^{i\omega})$;
д) сопряженная симметрия — если $x(n)$ действительна, то $X_n(e^{i\omega}) = X^*_n(e^{-i\omega})$.

6.3. По определению $X_n(e^{i\omega}) = a_n(\omega) - ib_n(\omega) = |X_n(e^{i\omega})|\,e^{i\theta_n(\omega)}$:
а) Выразить $|X_n(e^{i\omega})|$ и $\theta_n(\omega)$ через $a_n(\omega)$ и $b_n(\omega)$.
б) Выразить $a_n(\omega)$ и $b_n(\omega)$ через $|X_n(e^{i\omega})|$ и $\theta_n(\omega)$.

6.4. Пусть $x(n)$ и $w(n)$ имеют в качестве обычного преобразования Фурье $X(e^{i\omega})$ и $W(e^{i\omega})$ соответственно. Показать, что кратковременное преобразование Фурье $X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m}$ можно представить в виде

$$X_n(e^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{i\theta})\,e^{i\theta n}\,X(e^{i(\omega+\theta)})\,d\theta,$$

т. е. $X_n(e^{i\omega})$ представляет собой сглаженную оценку спектра $X(e^{i\omega})$ на частоте ω.

6.5. Определим кратковременную спектральную плотность мощности сигнала посредством кратковременного преобразования Фурье:

$$S_n(e^{i\omega}) = |X_n(e^{i\omega})|^2$$

и зададим кратковременную автокорреляционную функцию сигнала

$$R_n(k) = \sum_{m=-\infty}^{\infty} w(n-m)\,x(m)\,w(n-k-m)\,x(m+k).$$

Показать, что если

$$X_n(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,w(n-m)\,e^{-i\omega m},$$

то $R_n(k)$ и $S_n(e^{i\omega})$ связаны преобразованием Фурье, т. е. $S_n(e^{i\omega})$ есть преобразование Фурье от $R_n(k)$ и наоборот.

6.6. Допустим, что используемая для кратковременного анализа Фурье последовательность окна $w(n)$ физически реализуема и имеет рациональное z-преобразование вида

$$W(z) = \sum_{r=0}^{Nz} b_r z^{-r} / \left(1 - \sum_{k=1}^{Np} a_k z^{-k}\right).$$

[1] Начинающему полезно иметь в виду, что в математических работах (а также в работах, претендующих на «совместимость» терминологии) свойство а) принято называть аддитивностью, а свойство в) — однородностью. При этом линейность определяют как аддитивность и однородность. Постарайтесь понять, почему все же свойство а) названо здесь линейностью (совет: покажите, что из а) следует однородность для рациональных α). (Прим. перев.)

а) Какими свойствами должна обладать последовательность $W(z)$ или, что то же, $W(e^{i\omega})$, чтобы она годилась для указанного применения?
б) Получить рекуррентную формулу для $X_n(e^{i\omega})$ через сигнал $x(n)$ и предыдущие значения $X_n(e^{i\omega})$.
в) Рассмотрим случай $W(z) = 1/(1-az^{-1})$. Как следует выбрать a, чтобы получить разрешение по частоте около 100 Гц при частоте дискретизации 10 кГц?
г) Использование требуемого в в) значения a может привести к трудностям, если зависящий от времени анализ Фурье узкополосен и реализуется рекурсивно. Рассмотреть природу этих трудностей.

6.7. Доказать, что

$$\sum_{k=0}^{N-1} e^{i\frac{2\pi}{N}kn} = \begin{cases} N\sum_{r=-\infty}^{\infty}\delta(n-rN); \\ N, \quad n = rN, \quad r = 0, \pm 1, \ldots; \\ 0, \quad \textit{в противном случае}. \end{cases}$$

При доказательстве воспользуйтесь тождеством $\sum_{k=0}^{N-1}\alpha^k = \frac{1-\alpha^N}{1-\alpha}$.

6.8. Реализуя зависящее от времени Фурье-представление, можно применить дискретизацию как по времени, так и по частоте. В этой задаче исследуем эффекты обоих типов дискретизации. Рассмотрим последовательность $x(n)$ с обычным преобразованием Фурье

$$X(e^{i\omega}) = \sum_{m=-\infty}^{\infty} x(m)\,e^{-i\omega m}.$$

а) Для периодической функции $X(e^{i\omega})$, дискретизируемой на частоте $\omega_k = 2\pi k/N$, $k = 0, 1, \ldots, N-1$, имеем

$$\tilde{X}(k) = \sum_{m=-\infty}^{\infty} x(m)\,e^{-i\frac{2\pi}{N}km}.$$

Такие отсчеты можно представлять себе как дискретное преобразование Фурье последовательности $\tilde{x}(n)$, задаваемой соотношением

$$\tilde{x}(n) = \frac{1}{N}\sum_{k=0}^{N-1}\tilde{X}(k)\,e^{i\frac{2\pi}{N}kn}.$$

Показать, что $\tilde{x}(n) = \sum_{r=-\infty}^{\infty} x(n+rN)$.

б) При каких условиях на $x(n)$ при дискретизации $X(e^{i\omega})$ не возникают искажения из-за «наложений» во временно́й области.
в) Рассмотрим теперь «дискретизацию» последовательности $x(n)$, т. е. сформируем новую последовательность $y(n) = x(nM)$, состоящую из M-х отсчетов $x(n)$. Показать, что

$$Y(e^{i\omega}) = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{i(\omega-2\pi k)/M})$$

есть преобразование Фурье от $y(n)$. При доказательстве вы, возможно, захотите начать с рассмотрения последовательности $v(n) = x(n)p(n)$, где $p(n) = \sum_{r=-\infty}^{\infty}\delta(n+rM)$. После этого заметьте, что $y(n) = v(nM) = x(nM)$.
г) Какие ограничения надо наложить на $X(e^{i\omega})$, чтобы при дискретизации $x(n)$ не возникало наложений в частотной области?

6.9. Рассмотрим окно $w(n)$ с преобразованием Фурье $W(e^{i\Omega T})$, ограниченным по частоте интервалом $0 \leqslant \Omega \leqslant \Omega_c$. Мы хотим показать, что

$$\sum_{r=-\infty}^{\infty} w(rR-n) = W(e^{i0})/R$$

не зависимо от n для достаточно малого (ненулевого) целого R.

а) Пусть $w(r)=w(rR-n)$. Получить выражение для $W(e^{i\Omega T'})$ через R и $W(e^{i\Omega T})$, где T — период дискретизации $w(n)$, а $T'=RT$ — период дискретизации $\hat{w}(r)$ (указание: вспомните задачу прореживания сигнала в отношении $R:1$ или задачу 6.8в).

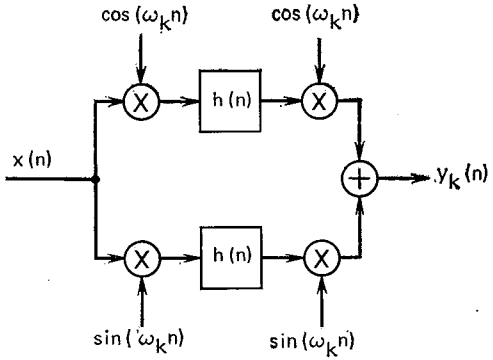

Рис. 3.6.1

б) Допустим, что $W(e^{i\Omega T})=0$ при $|\Omega|>\Omega_c$, получить выражение для максимального значения R (как функции от Ω_c) такого, что $\hat{W}(e^{i0})=W(e^{i0})/R$.

в) Вспомнив соотношение $\sum_{r=-\infty}^{\infty} \hat{w}(r)e^{-i\Omega T'r}=\hat{W}(e^{i\Omega T'})$, покажите, что если выполнены условия пункта б), то справедливо и соотношение, приведенное в формулировке задачи.

6.10. а) Показать, что импульсная характеристика системы, изображенной на рис. 3.6.1, имеет вид $h_k(n)=h(n)\cos(\omega_k n)$.

б) Найти частотную характеристику системы, изображенной на рис. 3.6.1.

6.11. Подчеркивания высокочастотной части спектра часто добиваются переходом к вычислению первой разности. В этой задаче исследуем влияние эффекта от такой операции на кратковременное преобразование Фурье.

а) Пусть $y(n)=x(n)-x(n-1)$. Показать, что $Y_n(e^{i\omega})=X_n(e^{i\omega})-e^{-i\omega}\times X_{n-1}(e^{i\omega})$.

б) При каких условиях можно считать, что $Y_n(e^{i\omega})\approx(1-e^{-i\omega})X_n(e^{i\omega})$? Вообще говоря, можно считать $x(n)$ линейно отфильтрованной: $y(n)=\sum_{k=0}^{N-1} h(k)x(n-k)$.

в) Показать, что $Y_n(e^{i\omega})$ связано с $X_n(e^{i\omega})$ соотношением $Y_n(e^{i\omega})=X_n(e^{i\omega})*h_\omega(n)$. Выразить $h_\omega(n)$ через $h(n)$.

г) Оправдано ли считать, что $Y_n(e^{i\omega})=H(e^{i\omega})X_n(e^{i\omega})$.

6.12. Гребенка из N фильтров характеризуется следующими свойствами: полосы расположены симметрично вокруг $\omega=\pi$, т. е. $\omega_k=2\pi-\omega_{N-k}$, $P_k=P^*_{N-k}$, $w_k(n)=w_{N-k}(n)$; существует канал, для которого $\omega_k=0$. Для четных и нечетных N:

а) прикинуть, как расположены N полос фильтров;

б) выразить общую импульсную характеристику гребенки фильтров через $w_k(n)$, ω_k, P_k, N.

6.13. Чтобы проиллюстрировать эффект реверберации, возникающий в гребенках БИХ-фильтров, рассмотрим общую импульсную характеристику $h(n)=\alpha_1\delta(n)+\alpha_2\delta(n-N)+\alpha_3\delta(n-2N)$, где представлены эхо, разнесенные на N отсчетов.

а) Определить функцию $H(e^{i\omega})$ системы и показать, что ее квадрат можно записать в следующем виде: $|H(e^{i\omega})|^2=(\alpha_2+(\alpha_1+\alpha_3)\cos(\omega N))^2+(\alpha_1-\alpha_3)^2\times\sin^2(\omega N)$.

б) Показать, что фазовую характеристику можно записать в виде

$$\theta(\omega) = -\omega N + \operatorname{tg}^{-1}\left[\frac{(\alpha_1-\alpha_3)\sin(\omega N)}{\alpha_2+(\alpha_1+\alpha_3)\cos(\omega N)}\right].$$

в) Определить, где расположены максимумы и минимумы амплитуды, для чего продифференцировать $|H(e^{i\omega})|^2$ по ω и положить результат равным нулю. Показать, что в случае $|\alpha_1+\alpha_3|\ll|\alpha_2|$ максимумы и минимумы расположены в точках $\omega=\pm k\pi/N$, $k=0, 1, 2, \ldots$

г) Воспользовавшись результатами п. в), показать, что максимальную амплитудную пульсацию (в децибелах) можно записать как

$$R_A = 20\log_{10}\left[\frac{|\alpha_2+\alpha_1+\alpha_3|}{|\alpha_2-\alpha_1-\alpha_3|}\right].$$

д) Найти R_A для

$$\alpha_1=0{,}1;\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}2;$$
$$\alpha_1=0{,}15,\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}15;$$
$$\alpha_1=0{,}1,\quad \alpha_2=1{,}0,\quad \alpha_3=0{,}1$$

е) Дифференцируя $\theta(\omega)$ по ω, можно показать, что максимумы и минимумы θ расположены в тех значениях ω, для которых $\cos(\omega N)=-[(\alpha_1+\alpha_3)/\alpha_2]$. Показать, что максимальная пульсация фазы задается равенством

$$R_p = 2\operatorname{tg}^{-1}\left[\frac{\alpha_1-\alpha_3}{(\alpha_2^2-(\alpha_1+\alpha_3)^2)^{1/2}}\right].$$

ж) Найти R_p для случаев, перечисленных в п. д.). Определить, как влияют изменения α_1 и α_3 на R_A и R_p.

6.14. Предлагается цифровое устройство выделения основного тона, состоящее из гребенки цифровых полосовых фильтров с нижними частотами среза $F_k=2^{k-1}F_1$, $k=1, 2, \ldots, M$, и верхними частотами среза $F_{k+1}=2^kF_1$, $k=1, 2, \ldots, \ldots, M$. При таком выборе частот среза гребенка фильтров обладает следующим свойством: если вход периодичен с основной частотой F_0 такой, что $F_k<F_0<F_{k+1}$, то энергия на выходе фильтров в полосах от 1 до $k-1$ будет мала, выходной сигнал в полосе k будет содержать основную частоту, а в полосы от $k+1$ до M попадут гармоники. Поэтому, если поместить за каждым фильтром выделитель чистого тона, получится хороший индикатор наличия основного тона.

а) Определить такие F_1 и M, чтобы указанный метод использовался для частот основного тона 50—800 Гц.

б) Сделать набросок необходимой частотной характеристики для каждого из M полосовых фильтров.

в) Что Вы можете предложить для реализации обнаружителя тона, необходимого на выходе каждого из фильтров?

г) Какие трудности можно предвидеть при реализации описываемого метода неидеальными полосовыми фильтрами?

д) Что произойдет, если на вход поступает речь, ограниченная полосой 300—3000 Гц, т. е. входной сигнал телефонной линии? Можно ли предложить какие-либо усовершенствования в этом случае?

6.15. Рассмотрим периодическую последовательность

$$\tilde{x}(n) = \sum_{r=-\infty}^{\infty} h_v(n+rN_p),$$

представляющую сегмент вокализованной речи.

а) Показать, что $\tilde{x}(n)$ можно разложить в ряд Фурье:

$$\tilde{x}(n) = \frac{1}{N_p}\sum_{k=0}^{N_p-1} \tilde{X}(k)\, e^{i\frac{2\pi}{N_p}kn},$$

где коэффициенты Фурье $\tilde{X}(k)$ представляют собой отсчеты преобразования Фурье вокализованной речи, т. е.

$$\tilde{X}(k) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right) \quad \text{(см. задачу 6.8).}$$

б) Показать, что кратковременное преобразование Фурье от $\tilde{x}(n)$ можно представить в виде

$$\tilde{X}(e^{i\omega}) = \frac{1}{N_p}\sum_{k=0}^{N_p-1} H_v\left(e^{i\frac{2\pi}{N_p}k}\right) W_n(e^{i(\omega-2\pi k/N_p)}),$$

где $W_n(e^{i\omega})$ — преобразование Фурье от $w(n-m)$.

в) Сколько различных значений принимает $X_n(e^{i\omega})$ при фиксированном ω? Для прямоугольного окна

$$w(n) = \begin{cases} 1, & 0 \leqslant n \leqslant N_p - 1; \\ 0, & \textit{в противном случае} \end{cases}$$

найти функцию $W_n(e^{i\omega})$.

г) Для каких значений N_p справедливо равенство

$$\tilde{X}_n\left(e^{i\frac{2\pi}{N_p}k}\right) = H_v\left(e^{i\frac{2\pi}{N_p}k}\right)$$

для прямоугольного окна шириной n?

6.16. Займемся анализом и синтезом сигнала $x(n)=\cos(\omega_0 n)$. Схема для анализа приведена для k-го канала на рис. 3.6.2*а*.

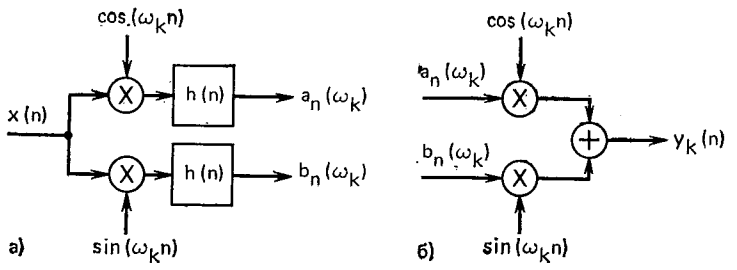

Рис. 3.6.2

а) Определить $a_n(\omega_k)$ и $b_n(\omega_k)$ для заданного входного сигнала.

б) Предположив, что $h(n)$ соответствует узкополосному фильтру нижних частот, упростить полученные выражения для $a_n(\omega_k)$ и $b_n(\omega_k)$ в предположении, что $(\omega_0-\omega_k)$ спадает в полосе фильтра и что $H(e^{i\omega})\approx 1$ для этих частот.

в) Сигналы $a_n(\omega_k)$ и $b_n(\omega_k)$ дают в комбинации мгновенное значение $M_n(\omega_k)$ и производную фазы $\dot{\varphi}_n(\omega_k)$. Определить $M_n(\omega_k)$ и $\dot{\varphi}_n(\omega_k)$ для нашего примера.

г) Показать, что для схемы синтеза, показанной на рис. 3.6.2*б*, выходной сигнал по существу совпадает с входным.

д) Производная фазы $\varphi_n=(\omega_k)$ вычисляется по формуле

$$\dot{\varphi}_n(\omega_k) = \frac{b_n(\omega_k)\dot{a}_n(\omega_k) - a_n(\omega_k)\dot{b}_n(\omega_k)}{[a_n(\omega_k)]^2 + [b_n(\omega_k)]^2}.$$

Найти $\dot{\varphi}_n(\omega_k)$ для нашего примера и сравнить результат с результатом в).

ж) Предположить, что производная фазы из п.д.) вычисляется с помощью первой разности, т. е. $\varphi_n(\omega_k)\approx\frac{1}{T}(a_n(\omega_k)-a_{n-1}(\omega_k))$, где T — период дискретизации во временной области. Найти $\varphi_n(\omega_k)$ по этой формуле и сравнить результат с полученным в п.в). При каких условиях они близки?

7
Гомоморфная обработка речи

7.0. Введение

Одно из основных предположений, сделанных в этой книге, состоит в том, что речевой сигнал трактуется как сигнал на выходе линейной системы с медленно изменяющимися параметрами. Это предположение позволяет считать, что на коротких сегментах речевой сигнал можно рассматривать как сигнал на выходе линейной системы с постоянными параметрами, возбуждаемой либо последовательностью импульсов, либо случайным шумом. Как уже отмечалось, проблема анализа речевого сигнала сводится к измерению параметров модели и оценке изменения этих параметров с течением времени. Поскольку сигнал возбуждения и импульсная характеристика фильтра взаимодействуют через операцию свертки, задача анализа речи может рассматриваться как задача разделения компонент, участвующих в операции свертки. Такая задача иногда называется задачей обратной свертки[1]. В гл. 6 был рассмотрен метод ее решения на основе представления речевого сигнала в виде переменного во времени преобразования Фурье. В данной главе на основе использования теории, изложенной в гл. 6, развивается другой подход к задаче, названный гомоморфной фильтрацией. После краткого введения в общую теорию гомоморфных систем будут рассмотрены различные способы применения методов гомоморфной обратной свертки в области анализа речевых сигналов.

7.1. Гомоморфные относительно свертки системы

Гомоморфные относительно свертки системы удовлетворяют обобщенному принципу суперпозиции. Принцип суперпозиции, если его записать для обычных линейных систем, имеет вид

$$\begin{aligned} L[x(n)] &= L[x_1(n)+x_2(n)] = \\ &= L[x_1(n)]+L[x_2(n)] = \\ &= y_1(n)+y_2(n) = y(n) \end{aligned} \tag{7.1а}$$

и

$$L[ax(n)] = aL[x(n)] = ay(n), \tag{7.1б}$$

где L — линейный оператор. Принцип суперпозиции устанавливает, что если сигнал на входе является линейной комбинацией элементарных сигналов, то и сигнал на выходе будет представлен в виде линейной комбинации соответствующих сигналов. Этот принцип иллюстрируется на рис. 7.1, где символ «+» на входе и выходе

[1] Операцию, обратную свертке (deconvolution), в переводной литературе также называют «разверткой». (*Прим. ред.*)