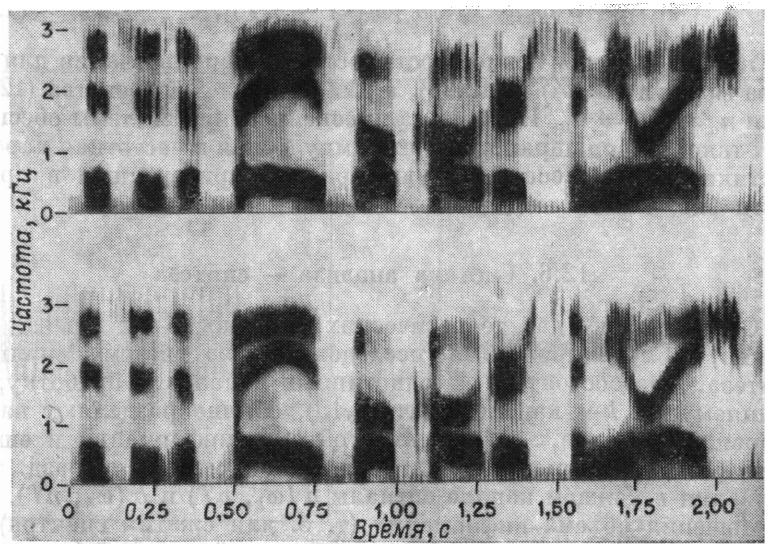


Фиг. 12.15. Полная схема системы анализа — синтеза, включая дискретизацию и квантование.

Подходящие значения T_1 и числа уровней квантования определяются из экспериментов по разборчивости речи. Блок синтеза аналогичен рассмотренному в разд. 12.5, за исключением интерполирующих ФНЧ, предназначенных для интерполяции принятых значений $a(\omega_k, nT)$ и $b(\omega_k, nT)$ к новому периоду дискретизации T_2 , отличному от T_1 , периода дискретизации при анализе.



Фиг. 12.16. Сравнение спектрограмм естественного и синтезированного высказываний.

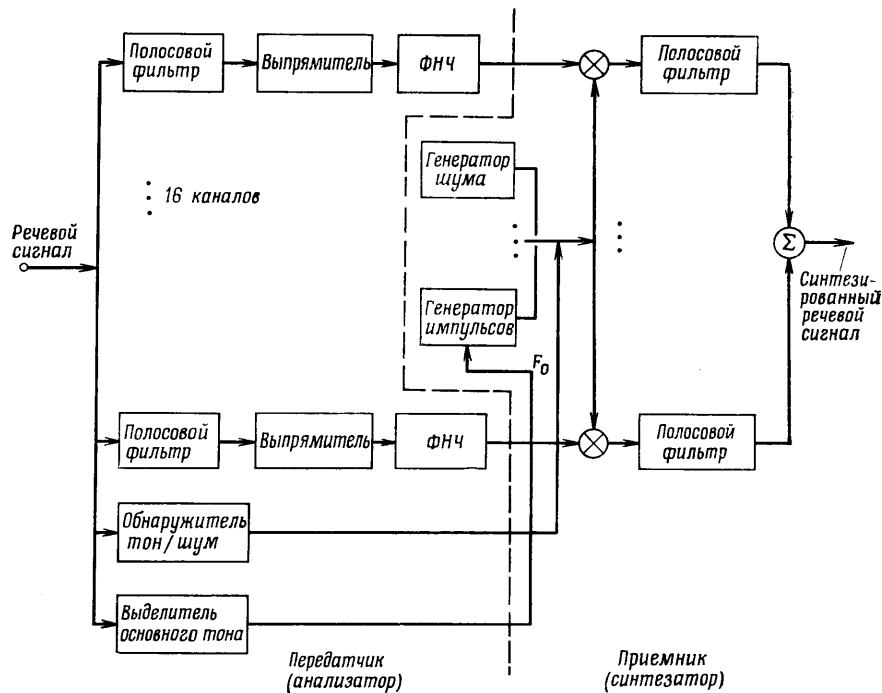
Эксперимент по моделированию системы анализа—синтеза подробно описан Шафером и Рабинером. Отметим лишь, что удовлетворительное качество речи получается при скорости передачи информации 15 000 бит/с, что примерно в четыре раза меньше, чем в системе с импульсно-кодовой модуляцией (ИКМ) при частоте дискретизации 8 кГц и семиразрядном логарифмическом кодировании. Если параметры спектра не квантованы по уровню, то спектрограммы исходного и синтезированного речевого сигналов (фиг. 12.16) трудно отличить.

12.7. Полосный вокодер

Полосный вокодер — это система анализа — синтеза речи, основанная на знании механизмов ее образования и восприятия. В частности, в полосном вокодере используется нечувствительность органов слуха к фазе сигнала и воспроизводится только кратковременный энергетический спектр речевого сигнала (это эквивалентно использованию амплитуды кратковременного фурье-преобразования без учета его фазы). Огибающая спектра речи измеряется с помощью гребенки полосовых фильтров, причем предполагается, что ее форма определяется характеристикой фильтра, образованного голосовым трактом. Источник возбуждения считается шумовым или импульсным квазипериодическим. (Таким образом, в вокодере непосредственно используется модель образования речи с независимыми источником возбуждения и голосовым трактом.) Существуют различные методы восстановления речи на основе измеренных параметров. В данном разделе описано несколько схем вокодера и рассмотрены факторы, влияющие на выбор их конструктивных параметров.

В типичном полосном вокодере (фиг. 12.17) исходный речевой сигнал $x(n)$ анализируется гребенкой полосовых фильтров (в данном случае их 16), неравномерно перекрывающих диапазон, существенный для восприятия речи (обычно от 0 до 3 кГц). Особенности проектирования этих фильтров будут рассмотрены ниже. Колебания на выходах полосовых фильтров детектируются и проходят через ФНЧ, выходные сигналы которых $y_k(n)$ в той или иной степени представляют огибающую спектра речи. Параметры, характеризующие источник возбуждения, получаются с помощью обнаружителя тон—шум, определяющего, является ли звук звонким (голосовые связки вибрируют) или глухим. В первом случае выделитель основного тона определяет основную частоту вибрации связок F_0 .

Шестнадцать канальных сигналов, сигнал тон—шум и значение высоты основного тона кодируются и передаются по каналу связи к приемнику. Предположим, что передача происходит без ошибок. Тогда задача приемника сводится к восстановлению речи



Фиг. 12.17. Блок-схема типичного полосного вокодера.

на основе переданных параметров. Источником возбуждения служит либо генератор импульсов, частота которого синхронизируется сигналом F_0 , либо генератор шума. В зависимости от сигнала тон—шум один из них подключается к гребенке фильтров, идентичных фильтрам анализатора, и возбуждает их. Протектированные сигналы огибающей спектра используются для модуляции колебаний на выходах соответствующих полосовых фильтров, за счет чего создается звуковая мощность в каждой из частотных полос. Синтезированный речевой сигнал получается после суммирования всех промодулированных полосовых колебаний.

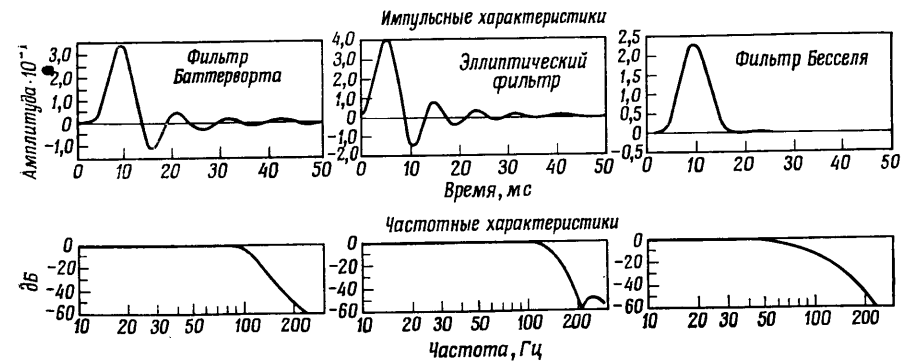
Одним из назначений вокодеров является передача речи удовлетворительного качества при малой скорости передачи информации. Во многих работах рассматривалось, какой должна быть скорость передачи, необходимая для полного представления параметров, используемых в полосном вокодере. Хотя в вокодерах различных типов требуется разная скорость, было показано, что для большинства из них она лежит в пределах от 2400 до 9600 бит/с. Вообще качество речи монотонно зависит от скорости передачи.

Так, лучше всего вокодеры звучат при скорости 9600 бит/с, а при 2400 бит/с речь для большинства слушателей становится почти неразборчивой.

12.8. Анализаторы вокодеров. Особенности обработки сигналов

Измерение спектра речи с помощью гребенки фильтров осложняется тем, что речь напоминает либо шум, либо периодическое колебание, основная частота которого может меняться в широких пределах. Дополнительные затруднения связаны с нестационарностью речи. Исходя из статических частотных характеристик, можно предположить, что фильтры с крутыми скатами дадут наилучшие оценки спектра. На практике оказывается, что такие фильтры имеют сравнительно длинные импульсные характеристики. Поэтому их использование приводит к временному смазыванию быстрых изменений спектра и, как уже обсуждалось, к появлению реверберации в синтезированном речевом сигнале. Для иллюстрации этого явления рассмотрим импульсные и частотные характеристики (фиг. 12.18) фильтра Баттерворта восьмого порядка с частотой среза 100 Гц, эллиптического фильтра с переходным отношением 0,5 и фильтра Бесселя. Частотные характеристики фильтра Баттерворта и эллиптического фильтра имеют более крутые скаты, чем характеристики фильтра Бесселя, однако длина их импульсных характеристик составляет почти 50 мс, а у фильтра Бесселя — менее 20 мс. Поэтому из трех рассмотренных только фильтр Бесселя может быть использован в полосном вокодере.

Временные искажения второго вида появляются, когда полосовые фильтры имеют неравные временные задержки, например если разные каналы имеют неодинаковые полосы. На практике полосы высокочастотных каналов делают более широкими, чтобы



Фиг. 12.18. Импульсные и частотные характеристики фильтров нижних частот Баттерворта, эллиптического и Бесселя (по Голдену).

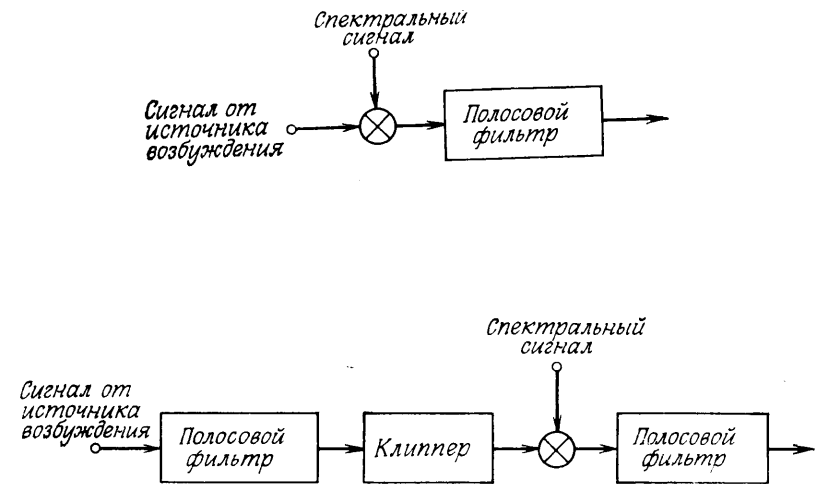
приблизить зависимость разрешения от частоты к характеристикам человеческого уха. Например, в низкочастотных каналах имеет смысл применять полосы шириной около 125 Гц, тогда как в наиболее высокочастотном канале часто используют полосу около 400 Гц. Существует несколько способов выравнивания задержек в фильтрах с неодинаковыми полосами. Простой метод состоит в увеличении порядка фильтра пропорционально его полосе; при этом особое внимание уделяется модулю и фазе частотной характеристики в области перекрытия с соседними каналами, чтобы сохранить суммарную частотную характеристику гребенки фильтров. Следует отметить, что при неравных полосах в более высокочастотные фильтры обязательно попадет несколько гармоник основного тона речевого сигнала. Вследствие этого в синтезированном сигнале появляются искажения, которые, как оказалось, слушатели почти не замечают.

При разработке ФНЧ анализатора учитываются следующие соображения. ФНЧ должен обеспечить надежное подавление на частотах от 50 Гц (наименьшей основной частоте мужского голоса) и выше, чтобы устранить все компоненты речевого сигнала на выходе детектора. Поскольку речевой спектр меняется медленно, частота среза ФНЧ не должна превышать 25 Гц. Для устранения реверберации импульсная характеристика фильтра не должна иметь «звона». Отсюда ясно, что условия проектирования ФНЧ весьма сложные. Обычно применяют фильтры, имеющие приблизительно линейную фазовую характеристику, например фильтры Бесселя или БИХ-фильтры Лернера, хотя КИХ-фильтры с линейной фазовой характеристикой тоже можно использовать.

12.9. Синтезаторы вокодеров. Особенности обработки сигналов

Методика расчета элементов синтезатора достаточно проста. Как уже обсуждалось, полосовые фильтры синтезатора в большинстве случаев идентичны фильтрам анализатора. Часто, однако, вместо стандартной схемы канального синтезатора применяют другой вариант, называемый каналом с выравниванием спектра (фиг. 12.19). Перед модуляторами включены жесткие (двухуровневые) ограничители, на выходе которых образуется сигнал типа меандра, принимающий в зависимости от знака входного сигнала, значения +1 или -1. Эти прямоугольные колебания постоянной мощности поступают на входы модуляторов.

В установленном режиме сигнал возбуждения тонового (звонкого) звука создает набор «меандров» с частотами, близкими к гармоникам основного тона. Таким образом, поведение канала с выравниванием спектра должно совпадать с поведением стандартного канала, если через первый полосовой фильтр проходит только



Фиг. 12.19. Сопоставление стандартной и модифицированной схем канального синтезатора вокодера.

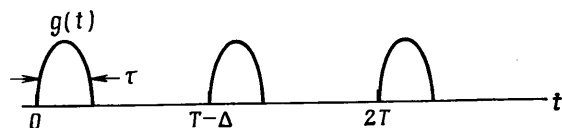
одна гармоника, поскольку третья и высшие гармоники, образующиеся при жестком ограничении, отфильтровываются выходным полосовым фильтром. Однако эти две схемы дают разное звучание даже при высокой основной частоте, когда в каждый фильтр попадает заведомо только одна гармоника.

Для объяснения этого явного несоответствия следует учесть, что спектр источника звука в гортани не постоянен, а флуктуирует во времени. Эти флуктуации вызваны как изменением сечения гортани, так и непостоянством частоты вибрации голосовых связок. Анализатор вокодера выделяет только частоту, но не форму колебаний голосовых связок. Тот факт, что выравнивание спектра существенно влияет на качество синтезированной речи, означает, что флуктуации спектра речи по крайней мере отчасти вызваны различием формы соседних звуковых импульсов, образующихся в гортани.

Быстрые флуктуации спектра, вызванные изменением частоты повторения импульсов возбуждения, можно исследовать количественно, вычислив энергетический спектр трех импульсов (фиг. 12.20). Расчет дает

$$G(\omega) \approx \frac{\sin^2(\omega\tau/2)}{(\omega\tau/2)^2} [3 + 4 \cos(\Delta\omega) \cos(\omega T) + 2 \cos(2\omega T)]. \quad (12.29)$$

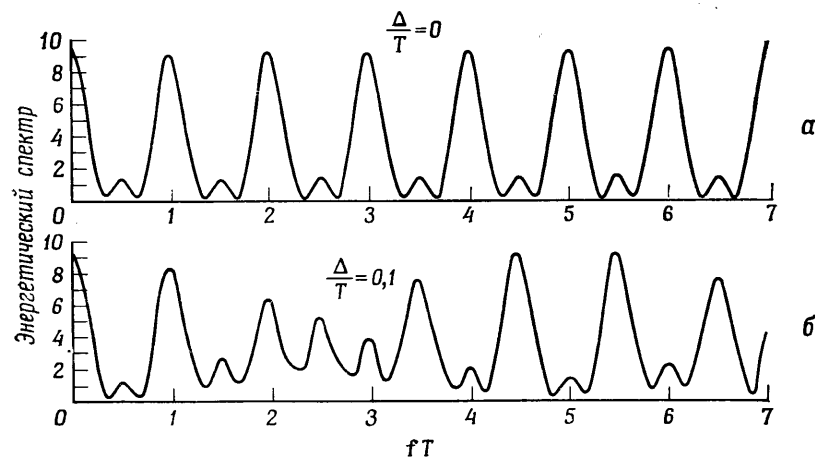
Если $\tau/T \ll 1$ (т. е. ширина импульса мала по сравнению с расстоянием между импульсами), то $\sin^2(\omega\tau/2)/(\omega\tau/2)^2 \approx 1$ в рассматриваемом диапазоне частот. Спектры $G(\omega)$ для $\Delta = 0$ и $\Delta = 0,1T$ изображены на фиг. 12.21.



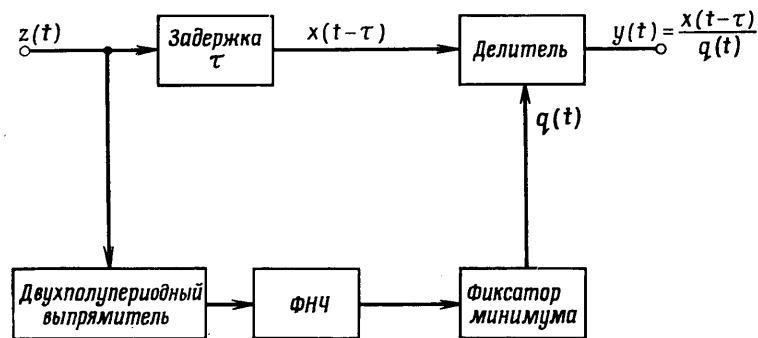
Фиг. 12.20. Импульсы возбуждающего сигнала для тоновой речи.

Если приведенные рассуждения справедливы, то флуктуации спектра, вызванные нестабильностью основного тона, проявятся как при спектральном анализе, так и при выделении высоты основного тона. Выравнивание спектра эффективно устраняет (в случае, если в каждый фильтр попадает не более одной гармоники) флуктуации спектра, связанные с нестабильностью сигнала возбуждения.

Если основная частота мала, так что в фильтры попадают по несколько гармоник, то схема выравнивания спектра не будет полностью сглаживать спектр возбуждения. Это связано со спецификой ограничения суммы двух гармоник: более слабая из них подавляется. Рассмотренные трудности можно в значительной степени преодолеть, заменив ограничитель системой автоматической регулировки усиления (АРУ), изображенной на фиг. 12.22. В этой системе задержанный узкополосный сигнал делится на собственный модуль, т. е. в предельном случае система АРУ переходит в двухуровневый ограничитель. При умеренных задержках (порядка 10 мс) спектральные искажения весьма малы,



Фиг. 12.21. Спектры двух различных импульсных последовательностей.



Фиг. 12.22. Автоматическая регулировка усиления в вокодеру.

даже если присутствует не одна гармоника. Короче говоря, выравнивание спектра методом АРУ сохраняет амплитудные соотношения в полосе фильтра и устраняет флуктуации спектра при переходе от фильтра к фильтру. Поскольку при этом не возникает (или почти не возникает) третьей гармоники, можно вообще не применять выходной полосовой фильтр в каждом канале.

12.10. Другие схемы вокодеров

Различными авторами были предложены и подробно изучены многие другие варианты анализаторов и синтезаторов для полосных вокодеров. Читатель может ознакомиться с устройством конкретных систем по многочисленным подробным публикациям.

12.11. Выделение высоты основного тона и обнаружение тон — шум

В заключительной части раздела о канальных вокодерах следует упомянуть задачи выделения высоты основного тона и обнаружения тон — шум, т. е. различения звонких и глухих звуков. Существует очень много алгоритмов измерения периода основного тона. Для иллюстрации рассмотрим весьма эффективный алгоритм, основанный на обработке во временной области и использующий параллельную обработку. Обе рассматриваемые задачи по сути являются сочетанием обработки сигналов и распознавания образов. Но, поскольку выделители высоты основного тона входят во многие системы обработки речи, целесообразно рассмотреть их здесь. В разд. 12.12 в связи с гомоморфной обработкой речи рассмотрен совершенно другой алгоритм выделения высоты тона.