**Контрольная работа по теме:**

**Кодирование речи в цифровых системах связи**

**1. Постановка задачи**

Достоинства цифровых методов представления, обработки, передачи и хранения информации, бурное развитие цифровой элементной базы – все это способствует тому, что цифровые методы обработки и передачи информации стали основным направлением систем связи. Однако использование цифровых методов представления, обработки и передачи информации приводит к многократному увеличению занимаемой полосы частот. Решение проблемы лежит в области разработки эффективных методов сжатия (кодирования) различных сообщений. В первую очередь сказанное относится к речевым сообщениям.

Следует сразу разграничить задачи при передаче речи и данных:

- В случае передачи данных стремятся обеспечить максимальную скорость передачи в заданной полосе частот.

- В случае передачи речевых сообщений, наоборот, стремятся уменьшить скорость цифрового потока и тем самым уменьшить необходимую полосу частот.

Речь, в отличие от данных, обладает смысловой и сигнальной избыточностью. Различные методы устранения избыточности (кодирования) в процессе преобразования речевого сигнала представляют широкий диапазон возможных скоростей передачи (от 64 кбит/с до примерно 200 бит/с).

Для цифровой передачи речи необходимо произвести оцифровку аналогового сигнала и закодировать каждую выборку двоичным кодом. Для аналого-цифрового преобразования, очевидно, необходимо задаться частотой дискретизации и числом уровней квантования. Рассмотрим простейший пример преобразования аналогового речевого сигнала в цифровую форму без устранения избыточности:

- Согласно международному стандарту, для хорошего качества передачи речи, достаточна полоса частот 300 – 3400 Гц. Тогда частота дискретизации согласно теоремы Котельникова должна быть выбрана из условия . Учитывая неидеальность фильтра на входе АЦП частоту дискретизации выбирают с некоторым запасом .

- Для высокого качества передачи речи достаточно квантовать аналоговый сигнал с использованием 13 разрядов (бит) двоичного кода, т.е. с использованием уровней. Сказанное относится к случаю равномерного квантования.

- Тогда скорость цифрового потока на выходе такого простейшего кодера .

Много это или мало? Например в стандарте подвижной сотовой связи GSM используется частотная манипуляция с минимальным сдвигом ЧММН и гауссовым предмодуляционным фильтром – модуляция GMSK (Gauss minimum shift keying). Спектральная эффективность такой модуляции равна . Значит для организации одного канала передачи речи потребуется полоса частот примерно 104 кГц. А если добавить к полученному цифровому потоку еще контрольные биты (для обеспечения помехоустойчивого кодирования) и биты управления, то необходимая полоса частот еще больше увеличится. Такая огромная полоса частот – слишком большая цена за цифровое качество связи.

Из всего сказанного можно сделать вывод – задачей речевого кодера является минимизация скорости цифрового потока на своем выходе, разумеется при сохранении высокого качества передачи речи.

Все методы цифрового кодирования речи можно разделить на три категории:

1. Кодеры формы сигнала – это кодеры, имеющие на выходе скорости цифрового потока в диапазоне 64 – 16 кбит/с и использующие методы импульсно-кодовой модуляции (ИКМ), дифференциальной ИКМ (ДИКМ), дельта-модуляции (ДМ), а также модификации указанных методов.
2. Вокодеры (Voice Coder) – это кодеры, имеющие на выходе скорости цифрового потока до 1.2 кбит/с и обеспечивающие кодирование спектра речевого сигнала и восстановление аналогового сигнала, звучащего подобно оригиналу, но не повторяющего его формы.
3. Гибридные кодеры.

**2. Кодеры формы сигнала**

Кодеры формы сигнала не являются специфичными для речи в том смысле, что они с успехом работают с любой формой сигнала и их характеристики ограничены только диапазоном изменения амплитуд (динамическим диапазоном) и шириной полосы частот входного сигнала. Кодеры формы сохраняют огибающую формы сигнала.

**Импульсно-кодовая модуляция** (ИКМ либо PCM – Pulse Code Modulation) – первый мировой стандарт кодирования речи со скоростью 64 кбит/с.

ИКМ со скоростью 64 кбит/с в основном используется в широкополосных системах связи (как правило это проводная телефония с использованием витой пары, коаксиала, оптоволокна), а так же как предварительное звено более совершенных низкоскоростных речевых кодеров, поскольку ее характеристики считаются очень высококачественными. Данный вариант кодирования речи отличается от рассмотренного ранее тем, что вместо равномерного квантования применяется квантование с логарифмическим сжатием.

Возможность такого сжатия вызвана двумя факторами:

1. Чувствительность человеческого уха снижается при увеличении уровня звука. Значит, малое изменение уровня звука при большом его абсолютном значении фактически неразличимо. Поэтому без ущерба качеству можно сократить число уровней квантования в области больших амплитуд.
2. У речевых сигналов максимум функции распределения вероятности находится в начале координат, т.е. сигналы с малыми амплитудами встречаются чаще нежели с большими. А это наталкивает на мысль, что можно почти не ухудшая качество квантовать большие амплитуды с меньшей точностью, чем малые.

Неравномерное квантование реализуется следующим образом: входной аналоговый сигнал поступает на нелинейное устройство (компрессор), которое сжимает уровни сигнала.

Амплитудная характеристика компрессора имеет вид (рис. 1):

Рис. 1

Затем сигнал поступает на равномерный квантователь. В приемнике сигнал с выхода ЦАП поступает на экспандер. Комбинированную пару компрессор-экспандер называют также компандером.

Данные обстоятельства позволяют в итоге уменьшить необходимое число бит на выборку (при сохранении того же качества) до 8.

**Дифференциальная импульсно-кодовая модуляция** (ДИКМ либо DPCM – Differencial PCM).

В методе ДИКМ для снижения избыточности речевого сигнала используется наличие корреляции между соседними отсчетами. Наличие корреляции дает возможность предсказывать последующие отсчеты по предыдущим (рис. 2).

 

Предсказатель



Рис. 2

Здесь представляет собой -й отсчет входного сигнала,

 - предсказанное значение -го отсчета входного сигнала,

 - ошибка предсказания -го отсчета.

Функция предсказателя в этом случае – проанализировать предыдущие отсчетные значения сигнала и спрогнозировать последующие.

В системе с ДИКМ кодировке и передаче по каналу связи подлежат не сами отсчеты (выборки) сигнала, а ошибки предсказания. В случае, когда параметры устройства предсказания адекватны свойствам речевого сигнала, удается уменьшить динамический диапазон ошибок предсказания по сравнению с динамическим диапазоном речевого сигнала. А это позволяет уменьшить число уровней квантования и, соответственно, количество бит на выборку и скорость цифрового потока на выходе кодера.

В декодере производится обратная операция и восстановление отсчета речевого сигнала по ошибке предсказания (рис. 3).

 

Предсказатель



Рис. 3

В качестве устройства предсказания как правило используется нерекурсивный цифровой фильтр, или трансверсальный фильтр:

  

 

 



Рис. 4

Сигнал на выходе такого фильтра представляет собой предсказанное значение - го отсчета входного сигнала – взвешенную линейную комбинацию  отсчетов (рис. 4):



где - коэффициенты усиления фильтра или коэффициенты модели предсказания.

Величины коэффициентов  выбираются таким образом, чтобы минимизировать среднеквадратическую ошибку



При использовании для кодирования речи метода ДИКМ число бит квантования на выборку снижается до 4-6 (32-48 кбит/с).

Дальнейшее снижение скорости цифрового потока на выходе кодера достигается за счет использования адаптивного варианта ДИКМ.

**Адаптивная дифференциальная импульсно-кодовая модуляция** (АДИКМ либо ADPCM) – один из наиболее общепринятых и широко используемых стандартов сжатия (кодирования) речи со скоростью 24-32 кбит/с.

Кодеры на основе ДИКМ разрабатываются в предположении, что выход источника сообщений стационарен. В действительности речь – это нестационарный случайный процесс, т.е. ее дисперсия, спектр и автокорреляционная функция меняются во времени. Поэтому для эффективного кодирования необходимо постоянно отслеживать такого рода изменения. В методе АДИКМ в отличие от ДИКМ производится адаптивная настройка шага квантования сигнала ошибки предсказателя, а также автоматическая настройка коэффициентов нерекурсивного фильтра предсказателя в соответствии с изменением текущего спектра речевого сообщения. Данные меры позволяют минимизировать динамический диапазон ошибки предсказания и необходимое число уровней квантования.

При кодировании речи с использованием АДИКМ для високого качества восстановленной речи требуется использовать всего лишь 3-4 бит на выборку.

**Дельта-модуляция** (ДМ).

Дельта модуляцию можно рассматривать как разновидность ДИКМ, в которой используется двухуровневый квантователь в соединении с предсказателем первого порядка. Таким образом спрогнозированное значение – это просто задержанный на один такт прошлый отсчет.

Данный вид кодирования является эффективным в случае, когда разность амплитуд между соседними отсчетами мала. Этого достигают стробируя аналоговый речевой сигнал с частотой в 4-6 раз большей, нежели частота Найквиста. Тогда скорость выдачи информации кодером составляет 32-48 кбит/с.

Речевому кодеру на основе Дельта-модуляции свойственно два вида искажений:

- перегрузка по наклону

Вызывается размером шага квантования, который слишком мал, чтобы отслеживать сигнал с резким наклоном.

- шум дробления

Вызывается размером шага квантования, который слишком велик, чтобы отслеживать сигнал с малым наклоном.

Для устранения указанных недостатков используют переменный шаг квантования, т. е. изменяют его адаптивно тем либо иным методом в направлении минимизации общего среднего квадрата ошибки от двух указанных факторов.

**3. Дискретная модель речеобразования**

Исследование статистических характеристик речевых сигналов базируется на математическом описании акустического процесса речеобразования, который в свою очередь основывается на физических процессах речеобразования.

Модель

возбуждения

Модель

излучения

Рис. 5

В представленной модели можно выделить две системы – модель возбуждения и модель излучения (рис.5).

В случае вокализованных звуков источник возбуждения должен формировать квазипериодическую последовательность импульсов с частотой основного тона. Поэтому модель возбуждения для случая вокализованных звуков можно представить в виде (рис. 6):

Генератор

импульсов

Частота основного тона

Фильтр модели голосовой щели 

Усилитель



Рис. 6

В случае невокализованных источник возбуждения формирует случайное шумовое колебание. Тогда модель возбуждения для случая невокализованных звуков представима в виде (рис. 7):

Генератор шума (ГШ)

Усилитель



Рис. 7

Коэффициенты усиления  и  определяют интенсивность голосового возбуждения (громкость звука).

Аналогичным образом модель излучения может быть представлена в виде (рис. 8):

Модель голосового тракта



Фильтр модели излучения



Параметры голосового тракта

Рис. 8

Модель голосового тракта полностью характеризуется передаточной функцией , полюсы которой соответствуют резонансам (формантам) речевого сигнала, а нули - антирезонансам (провалам) в спектре речевого сигнала. Причем модель голосового тракта с учетом только лишь полюсов уже дает в большинстве случаев довольно точное описание голосового тракта для большинства звуков речи.

Таким образом результирующая передаточная функция процесса речеобразования записывается в следующем виде:



Для вокализованных звуков, медленно изменяющихся во времени, рассмотренная модель речеобразования оказывается достаточно точной. Для невокализованных звуков, быстро изменяющихся во времени (поскольку они шумоподобные), данная модель будет адекватной только для очень коротких во времени реализаций (фрагментов) речи. В любом случае параметры модели речеобразования (частоту основного тона, коэффициенты усиления, параметры голосового тракта) обновляют в течение каждых 10-20 мс. Для этого используется так называемый кратковременный анализ речи.

Основной задачей большинства систем синтеза и анализа речи (в том числе и систем кодирования речи с использованием вокодеров) является оценка параметров в модели речеобразования по реальной речи.

**4. Метод кратковременного анализа**

Как уже было сказанно, речь – это нестационарный случайный процесс. Однако на интервалах 10-20 мс ее можно рассматривать как локально стационарный случайный процесс. Согласно методу кратковременного анализа поток отсчетов речевого сигнала разбивается на временные окна, называемые также сегментами или окнами анализа. Эти временные окна могут в принципе соответствовать как отдельным фонемам, так и слогам и даже целым словам, но для обеспечения большей эффективности анализа выбираются в пределах стационарности речи (10-20 мс). Каждый сегмент речи подвергается анализу, в процессе которого вычисляются такие его характеристики как энергия, число пересечений нуля, текущий спектр.

Например, энергия фрагмента речи, состоящего из  отсчетов определяется выражением:

,

где  - значение амплитуды -го отсчета.

Число пересечений нуля может быть вычислено по формуле:

,

где



На основе вычисления энергии фрагмента речи и числа пересечений нуля каждое окно анализа можно отнести к одному из трех типов:

1. вокализованное – окно V
2. невокализованное – окно N
3. пауза – окно P

Алгоритмы разделения временных окон по типам V, N, P как правило основываются на сравнении полученных значений энергии и числа пересечений с некоторыми пороговыми значениями, полученными экспериментально.

Например:

1. Если , то окно типа P.
2. Если , то если  - окно типа V

Иначе – окно типа N.

После разбиения потока отсчетов речевого сигнала на окна V, N, P в пределах каждого окна вычисляются параметры модели речеобразования, которые используются для кодирования речи в передатчике и для синтеза речи в приемнике.