Глава1 Кодеры формы.

Кодеры формы характеризуются способностью сохранять основную форму речевого сигнала. Кодеры формы не являются специфичными для речи в том смысле, что они с успе­хом работают с любой формой входного сигнала, и их применение ограничено только пределами амплитуды и шириной полосы. Сохраняя огибающую формы сигнала, подобные кодеры работают по принципу выборка-выборка, и их характеристики эффективно измеряются отношением сигнал/шум (ОСШ), так как квантование является основным источником искажений формы выходного сигнала.

ИКМ — первый мировой стандарт кодирования речи со скоростью 64 кбит/с с логарифмическим сжатием (по μ-закону для Северной Америки и *А*-закону для Европы). ИКМ-кодер является кодером формы и все еще широко используется в цифровых системах. ИКМ со скоростью 64 кбит/с в основном используется как предварительное звено низкоскоростных речевых кодеров, поскольку ее характеристики считаются очень высоко­качественными. Позже был разработан стандарт на адаптивную дифференциальную ИКМ (АДИКМ) со скоростью 32 кбит/с. Снижение скорости цифрового потока наполовину было достигнуто благодаря использованию адаптивного предсказания и адаптивных квантователей для устранения избыточности речи. Некоторые другие кодеры, например с дельта-модуляцией и плавно изменяющейся крутизной (CVSDM — Continuous Variable Slope Delta Modulation) на скорости 32 кбит/с, используются для решения специфичных задач. Хотя подобные высокоскоростные алгоритмы кодирования малоэффективны, они, тем не менее, остаются самыми эксплуатируемыми системами и, возможно, будут оставаться таковыми еще некоторое время.

## Импульсно-кодовая модуляция ИКМ (РСМ – Pulse Code Modulation). Рекомендация G.711

При построении систем цифровой передачи непрерывных сообщений принципиальным моментом является определение полосы частот, требуемой для обеспечения заданного качества воспроизведения переданного сообщения. Вообще говоря, для высококачественной передачи речевого сообще­ния требуется полоса не менее 10 кГц.

Однако для достижения удовлетвори­тельного уровня разборчивости при пе­редаче речи по телефонным каналам достаточно передать спектр в полосе 300...3400 Гц. Именно такой спектр звуковых частот обычно передается в современных системах передачи рече­вой информации.

Как правило, максимальная частота передаваемого спектра аудио­сигнала выбирается равной ,а частота дискретизации (например, рекомендации G.711, G.721), хотя в ряде случаев с целью повышения качества передачи используются и более высокие значе­ния этих величин (например, рекомен­дация G.722).



При использовании ИКМ дискретизированное сообщение подвергается квантованию по *L* уровням (рис. 1.1), в результате чего каждому значе­нию ставится в соответствие чи­сло , , представленное *n*-разрядной комбинацией двоичного кода.



Для достижения приемлемого ка­чества восприятия восстановленного речевого сообщения при равномерном (простом) квантовании необходимо . Столь большое число уровней квантования при требует скорости передачи символов в канале не менее .



Рисунок 1.1

Однако в связи с тем, что при восприятии речи человеческим ухом в области больших мгновенных значений оказываются допустимыми значительно большие искажения сообщения, чем в области малых мгновенных значений, требуемое число уровней квантования может быть существенно снижено путем использования неравномерного квантования, используя компрессию исходного сообщения по логарифмическому закону с последующим равномерным квантованием при сравнительно малом числе уровней (например, при или путем соответствующего цифрового преобразования (цифровой компрессии) сообщения, предварительно преобразованного в цифровую форму при сравнительно большом исходном числе уровней квантования (например, при ).



Оптимальный квантователь имеет преимущества, если динамический диапазон входного сигнала фиксирован и достаточно мал. Характеристики квантователя быстро ухудшаются, мощность сигнала изменяется относительно значения, на которое он был рассчитан. Хотя этим процессом можно управлять, нормализуя входной сигнал и приводя его к единому диа­пазону, для правильного определения масштаба амплитуды восстановленно­го после квантователя сигнала потре­буется дополнительно несколько бит, необходимых для передачи динамиче­ского диапазона сигнала в определен­ные моменты времени.

Для обработки входных речевых сигналов с большим динамическим диапазоном используются два зако­на сжатия, называемые импульсно-кодовой модуляцией по закону (*А*-ИКМ) и по закону μ (μ-ИКМ). В обеих схемах характеристика от­ношения сигнал/шум квантования (ОСШкв) должна быть близка к харак­теристике для простого квантователя. Вместе с тем характеристики *А*-ИКМ и μ-ИКМ существенно не изменяют­ся и остаются сравнительно постоян­ными в большом диапазоне уровней входного сигнала. По сравнению с про­стыми квантователями (рис. 1.1) кван­тователи сжатия требуют меньше бит на входную выборку для определен­ного динамического диапазона сжатия и меньшего ОСШкв. В квантователях сжатия уровни квантования находятся в области малых амплитуд, которые увеличиваются при увеличении диа­пазона входного сигнала. Благодаря этому при квантовании речевых сигналов, у которых максимум функции распределения вероятностей находится в начале координат, наиболее ча­сто встречающиеся малые амплитуды квантуются с большей точностью, чем менее вероятные большие амплитуды, что приводит к значительно лучшим, Чем у простого квантователя, характер­истикам.



Сжатие по *А*-закону определяется зависимостью:

(1.1)



где *A* — параметр сжатия с типовыми значениями 86 (Северо-Американская ИКМ) и 87,56 (Европейская ИКМ) для семибитных речевых квантователей.

Сжатие по μ-закону определяется выражением

(1.2)



где *V0* задается формулой , в которой *L* – нагрузочный фактор, a – среднеквадратическое значение входного речевого сигнала.



Типовое значение фактора сжатия μ равно 255. Выражение (1.1) пока­зывает, что *А*-закон — это комбина­ция логарифмической кривой, исполь­зуемой для больших амплитуд, и ли­нейного участка, используемого на ма­лых амплитудах. μ-закон не являет­ся в точности линейным или логариф­мическим ни в одном диапазоне, од­нако является приблизительно линей­ным для малых амплитуд и прибли­зительно логарифмическим для боль­ших амплитуд. Сравнение между квантователем по μ-закону и опти­мальным квантователем показало, что оптимальный квантователь дает вы­игрыш 4 дБ, однако может иметь бо­лее высокий уровень фонового шума, когда канал свободен, и его динамиче­ский диапазон сведен к минимальному диапазону входного сигнала. Поэтому наиболее предпочтителен логарифми­ческий квантователь.

Цифровое преобразование непре­рывного речевого сообщения в соответ­ствии с рекомендацией G.711 (рис. 1.2) используется наиболее часто.



Рисунок 1.2

При этом ; частота дискретизации . После равномерного квантования при числе уровней и предварительного кодирования производится цифровая компрессия, в результате чего длина кодовой комбинации уменьшается до разрядов. Результатом преобразования является двоичная последовательность, передаваемая со скоростью 64 кбит/с.



Из различных систем адаптивной ИКМ (АИКМ) наибольшее распространение получила система блочной ИКМ (БИКМ), оторую часто называют системой с почти мгновенным компандированием (NIC — Near Instantaneous Companding).

Отсчеты *n*-разрядного АЦП разбивают на блоки по *N* отсчетов. В каждом блоке находят отсчет с макси­мальным для данного блока уровнем. Этому уровню соответствует определенный номер старшего значащего раз­ряда (*j*), и все старшие разряды в ком­бинациях этого блока будут нулевыми. Записанный в двоичном коде но­мер этого разряда образует масштабную информацию, которая из-за своей важности, как правило, защищается помехоустойчивым кодом. В результате масштабная информация вместе с проверочными символами образует *m*-значную комбинацию, которую добавляют к основной информации.

Основная информация формируется выбором *k* разрядов из *n* исходных разрядов, причем первым (старшим) разрядом является разряд с номером, описанным в масштабной информации.

Основная информация для каждого из блоков объединяется с масштабной в единый цифровой поток. Результирующая скорость цифрового потока на выходе системы БИКМ . На практике, как правило, используют следующие параметры: .



При одинаковых условиях передачи БИКМ дает лучшее качество, чем ИКМ. Поэтому можно снизить скорость передачи до 32.. .56 кбит/с.

## Дифференциальная импульсно-кодовая модуляция ДИКМ (DPCM – Differencial Pulse Code Modulation)

Наряду с ИКМ применяются и более эффективные цифровые методы передачи речи. В частности, с целью снижения требований к пропускной способности канала можно использовать наличие корреляции между от­четными значениями передаваемого сообщения. Такой метод называется передачей с предсказанием. При этом последовательность значений поступает на один вход вычитающего устройства (рис. 1.3,а), в то время как на другой вход поступает предсказанное значение , полученное тем или иным методом в устройстве предска­зания на основе анализа как преды­дущих отсчетных значений сообщения, так и текущих передаваемых значений на входе вычитающего устройства.



Рисунок 1.3

На приемном конце значения сооб­щения восстанавливаются путем добавления принятого сигнала ошибки предсказания к предсказываемому значению (рис. 1.3,б).



В системе с дифференциаль­ной импульсно-кодовой модуляци­ей (ДИКМ) отсчетные значения ошибки предсказания подвергаются квантованию с переходом к значениям аналогично тому, как это делает­ся при использовании обычной ИКМ, однако при существенно меньшем числе уровней квантования. Таким обра­зом, при одинаковом качестве переда­чи речи метод ДИКМ позволяет ис­пользовать меньшее число разрядов *n* в кодовых комбинациях по сравне­нию с ИКМ. При этом существует большое число различных вариантов реализации метода ДИКМ, наиболее типичный из которых представлен на рис. 1.4.



Рисунок 1.4

При этом имеют место соотноше­ния:

(1.3)



Классификационными признака­ми кодеров ДИКМ считаются нали­чие блока линейного предсказания ав­торегрессионных последовательностей (предсказателя) и использование мно­гоуровневого (больше двух уровней) квантователя. Блок линейного пред­сказания может состоять из двух ча­стей — долговременного и кратковре­менного предсказателей. В канал пе­редается разность истинного и пред­сказанного значений сигнала (сигнал-остаток, он же – погрешность пред­сказания). Системы с ДИКМ обеспе­чивают такое качество восстановления сигнала, которое сопоставимо с предоставляемым ИКМ, и на порядок более высокую помехоустойчивость.

Эффективность метода ДИКМ может быть повышена путем пере хода к адаптивной дифференциальной импульсно-кодовой модуляции АДИКМ.

## Адаптивная дифференциальная импульсно-кодовая модуляция (ADPCM — Adaptive Differencial Pulse Code Modulation). Рекомендации G.721 и G.726

ADPCM – один из наиболее об­щепринятых и давно используемых ал­горитмов сжатия речи, который регла­ментируется стандартом G.726, был принят в 1984 г. Этот алгоритм да­ет практически такое же качество вос­произведения речи, как и РСМ, однако для передачи информации при его ис­пользовании требуется всего 32 кбит/с. Метод основан на том, что в анало­говом сигнале, передающем речь, не­возможны резкие скачки интенсивно­сти. Поэтому, если кодировать не са­му амплитуду сигнала, а ее измене­ние по сравнению с предыдущим зна­чением, то можно обойтись меньшим числом разрядов. В ADPCM изменение уровня сигнала кодируется четырехразрядным числом, при этом частота измерения амплитуды сигнала сохраняется неизменной.

Все методы кодирования, основанные на определенных предположениях о форме сигнала, плохо работают в ситуации, когда сигнал может передаваться с резкими скачками амплитуды. Именно такой вид имеет аудиосигнал, генерируемый модемам или факсимильными аппаратами. Современные системы обмена информацией, поддерживающие цифровые ли­нии связи, умеют распознавать фак­симильный обмен и передают соответ­ствующие сигналы непосредственно в цифровом виде, не преобразуя их в ау­диосигнал.

Нелинейный 15-уровневый адаптивный квантователь используется для квантования разностного сигна­ла . Перед квантованием сигнал логарифмируется по основанию 2 и масштабируются посредством коэф­фициента , который вычисляется с помощью блока адаптации масштаб­ного коэффициента.



Для определения квантованного уровня используются четыре дво­ичных символа (три для амплитуды и один для знака). Четырехбитовый вы­ход квантователя образует выход­ной цифровой сигнал со скоростью 32 кбит/с, который одновременно подает­ся на инверсный адаптивный кванто­ватель и блок управления скоростью адаптации масштабного коэффициента квантователя.



Квантованная версия разностного сигнала формируется путем мас­штабирования с использованием спе­циальной величины , выделяемой из нормализованной характеристики квантователя, и дальнейшей транс­формации результата из логарифмиче­ского представления.



Блок адаптации масштабного ко­эффициента квантователя вычисляет — масштабный коэффициент для квантователя и инверсного квантовате­ля. На его входы подаются четырехби­товые выходные сигналы квантователя и параметр управления скоростью адаптации .



Основной принцип, реализуемый при масштабировании, заключается в бимодальной адаптации:

– быстрой – для сигналов (напри­мер, речевых), которые дают разност­ные сигналы с большими флуктуациями;

–медленной – для сигналов (например, данных в диапазоне тональ­ных частот, тонов), которые дают разностные сигналы с малыми флуктуациями.

Управление скоростью адаптации производится с помощью комбинации быстрого и медленного масштабных коэффициентов.

Быстрый (нефиксированный) масштабный коэффициент вычисляется рекурсивно в логарифмиче­ском представлении с основанием 2 из результирующего логарифмическо­го масштабного коэффициента:



(1.6)



Как правило, лежит в пределах . Дискретная функция определяется таблич­ным образом. Множитель (1 – 2-5) вводит ограниченную память в процесс адаптации таким образом, что состо­яния кодера и декодера сходятся при ошибках передачи.



Медленный (фиксированный) мас­штабный коэффициент получа­ется из с помощью операции фильтрации нижних частот:



(1.7)



Затем быстрый и медленный мас­штабные коэффициенты объединяются для получения результирующего мас­штабного коэффициента:

(1.8)



где .



Управление скоростью адапта­ции. Предполагается, что управляю­щий параметр может принимать значения в диапазоне [0, 1]. Для рече­вых сигналов он стремится к единице, Для сигналов, данных в диапазоне то­нальных частот и одночастотных сигналов он стремится к нулю. Величи­на коэффициента определяется мерой скорости изменения величины разност­ного сигнала.



Адаптивный предсказатель и калькулятор восстановленного сигна­ла. Первоначальная функция ада­птивного предсказателя заключается в вычислении оценки разностного сигнала . Используются две структуры адаптивного предсказате­ля – каскад первого порядка, модели­рующий нули, и каскад второго поряд­ка, моделирующий полюсы во входном сигнале.



Детектор тона и перехода. С целью улучшения рабочих характери­стик для сигналов, поступающих с вы­ходов модемов с частотной манипуля­цией, работающих в режиме кодовых комбинаций, определен двухступенча­тый процесс декодирования. Снача­ла производится детектирование сиг­нала с ограниченной полосой (напри­мер, тона), в результате чего квантова­тель может быть переведен в быстрый режим адаптации.

Упрощенная и развернутая струк­турные схемы декодера АДНКМ при­ведены на рис. 1.6,а и 1.7,б соответ­ственно. Декодер включает схему, идентичную цепи обратной связи коде­ра, преобразователь линейной ИКМ в сигнал по законам *А* или μ и устрой­ство установки синхронного кодирова­ния.

Устройство установки синхрон­ного кодирования предотвращает нако­пление искажений, имеющих место при синхронном последовательном кодиро­вании (АДИКМ-ИКМ-АДИКМ, дру­гие цифровые соединения). Установ­ка синхронного кодирования достига­ется путем подстройки проходного ко­да ИКМ таким образом, чтобы попы­таться устранить искажения квантова­ния в следующем каскаде кодирования АДИКМ.

Функции основных блоков декоде­ра и кодера совпадают и поэтому ниже не рассматриваются.

Вокодеры

Вокодер (от английских слов voice – голос и coder – кодировщик) пред­ставляет собой устройство, осуще­ствляющее параметрическое компандирование речевых сигналов. Ком­прессия речевых сигналов на переда­ющем конце канала связи производит­ся в анализаторе, выделяющем из ре­чевого сигнала медленно меняющиеся составляющие, которые передаются по каналу связи в виде кодовых посылок. На приемном конце с помощью мест­ных источников сигналов, управляе­мых принятыми параметрами, синте­зируется речевой сигнал.

Работа вокодеров основана на моделировании человеческой речи с учетом ее характерных особенностей. Вместо непосредственного измерения амплитуды вокодер преобразует вход­ной сигнал в некий другой, похожий на исходный. Причем измеряемые харак­теристики речевого сигнала использу­ются для подгонки параметров в при­нятой модели речевого сигнала. Имен­но эти параметры и передаются прием­нику, который по ним восстанавливает исходный речевой сигнал. По суще­ству, речь идет о синтезе речи. Есте­ственно, что измерение искажений от­ношения сигнал/шум бесполезно для вокодеров, и, следовательно, необхо­димы другие субъективные оценки, та­кие, как средняя экспертная оценка, диагностический рифмованный тест, диагностическая оценка приемлемости и др. Вокодеры можно разделить на два класса: речеэлементные и параметри­ческие.

В речеэлементных вокодерах при передаче распознаются произне­сенные элементы речи (например, фо­немы) и передаются только их но­мера. На приеме эти элементы со­здаются по правилам речеобразования или берутся из памяти устройства. Область применения фонемных воко­деров – линии командной связи, ре­чевое управление и говорящие автома­ты информационно-справочной служ­бы. Практически в таких вокодерах происходит автоматическое распозна­вание слуховых образов, а не опреде­ление параметров речи.

В параметрических вокодерах из речевого сигнала выделяют два ти­па параметров:

параметры, характеризующие оги­бающую спектра речевого сигнала, (фильтровую функцию);

параметры, характеризующие ис­точник речевых колебаний (генератор­ную функцию), – частота основного тона, ее изменение во времени, момен­ты появления и исчезновения основно­го тона, шумового сигнала.

По этим параметрам на приеме синтезируют речь.

По принципу определения параме­тров фильтровой функции речи разли­чают вокодеры:

• полосные канальные (channel);

• формантные;

• ортогональные;

• липредеры (с линейным предска­занием речи);

• гомоморфные.

В полосных вокодерах спектр речи делится на 7-20 полос (каналов) аналоговыми или цифровыми полосо­выми фильтрами. Большее число ка­налов в вокодере дает большую нату­ральность и разборчивость. С каждо­го полосового фильтра сигнал поступа­ет на детектор и фильтр низких частот с частотой среза *Fcp*. Таким образом, сигналы на выходе каждого канала из­меняются с частотой менее *Fcp*. Их пе­редача возможна в аналоговом или ци­фровом виде.

В формантных вокодерах оги­бающая спектра речи описывается комбинацией формант (резонансных частот голосового тракта). Основные параметры формант – центральная частота, амплитуда и ширина полосы частот.

В ортогональных вокодерах огибающая мгновенного спектра рас­кладывается в ряд по выбранной си­стеме ортогональных базисных функ­ций. Вычисленные коэффициенты это­го разложения передаются на прием­ную сторону. Распространение полу­чили гармонические вокодеры, исполь­зующие разложение в ряд Фурье.

Вокодеры с линейным пред­сказанием (LPC — Linear Prediction Coding, или липредеры, основаны на оригинальном математическом аппа­рате. Они получили наибольшее рас­пространение и будут ниже рассмотре­ны более подробно.

Гомоморфная обработка позво­ляет разделить генераторную и филь­тровую функции, образующие речевой сигнал.

Из-за сложности определения па­раметров генераторной функции по­явились полувокодеры (VE — Voice Excited Vocoder), в которых вместо сигналов основного тона и тон-шума используется полоса речевого сигна­ла. Полоса частот до 800. .. 1000 Гц кодируется АДИКМ, АДМ (адаптивная дельта модуляция) или с помощью линейного предсказания малого порядка, а в некоторых моделях передается в аналоговом виде. Известны разные типы полувокодеров-липредеров: VELP — Voice Excite Linear Prediction; RELP — Residue Excited Linear Prediction.

Вокодеры VELP используют голосовое возбуждение и коэффициент линейного предсказания (КЛП). В вокодерах RELP по исходному сигнал также вычисляются КЛП. Так как КЛП описывает фильтровую функцию, то сигнал ошибки (остатка) предсказания содержит информацию о генераторной функции речи и передается на приемную сторону (возможно ее сжатие методами АДИКМ, АДМ или помощью линейного предсказания малого порядка).

Характеристики вокодеров. Качество речи вокодеров являет функцией скорости передачи, производительности и задержки обработки. Если вокодеры предназначены для телефонии по Интернет, разработчики продукции должны учитывать эти характеристики, между которыми существует строгая зависимость. Например, низкоскоростные вокодеры обычно имеют большую задержку и более низкое качество речи, чем высокоскоростные.

Скорость. Так как вокодер совместно использует канал связи и часто перегруженную сеть предприятия или Интернет с другими информационными потоками, максимальная скорость должна была бы быть как можно ниже, особенно для приложений малых офисов. В настоящее время большинство вокодеров работают на фиксированной скорости вне зависимости от характеристик входного сигнала, однако целью современных разработок являются вокодеры с переменной скоростью. Для приложений по одновременной передаче речи и дан­ных компромиссом является создание алгоритмов сжатия пауз в качестве части стандарта кодирова­ния. Общим решением является ис­пользование фиксированной скорости для речи и низкой скорости для фо­новых шумов. Способ выполнения ме­ханизма сжатия пауз важен для повы­шения качества передачи речи, одна­ко часто выигрыш от компрессии пауз не реализуется. Проблемой является то, что при больших фоновых шумах сложно провести различия между ре­чью и шумом. Другая проблема за­ключается в том, что если механизм сжатия пауз неправильно выявил состояние речи, начало речи может быть «отрезано», что значительно ухудшает разборчивость кодированной речи.

|  |  |
| --- | --- |
| Алгоритм | Описание |
| Детектор активности речи (VAD) | Определяет, является ли входной сигнал речью или фоновым шумом. Если сигнал считается речью, он кодируется на полной фиксированной скорости; если сигнал считается шумом, он кодируется на более низкой скорости |
| Генерация комфортного шума | Механизм действует на стороне приемника для воссоздания основной характеристики фонового шума |

Способ генерации комфортного пума должен быть таким, чтобы кодер и декодер оставались синхронизи­рованными, даже если в течение некоторого интервала времени передача данных не осуществляется. Это позволяет сгладить переходы между сегментами активной и неактивной речи.

Производительность алгоритмa. Вокодеры частот выполняются на основе цифровых сигнальных процес­соров (ЦСП). В соответствии с ком­пьютерной терминологией их произво­дительность может быть измерена в млн. операций в секунду, объеме па­мяти с произвольным доступом ОЗУ и объеме ПЗУ. Производительность определяет стоимость вокодера, поэто­му при определении типа вокодера для тех или иных приложений разработ­чик должен сделать соответствующий выбор. В случаях, когда вокодер со­вместно использует процессор с дру­гими приложениями, разработчик дол­жен решить, сколько ресурсов мож­но выделить для вокодера. Вокодеры, использующие менее 15 млн. опера­ций/с, считаются низкопроизводитель­ными. Использующие 30 или более млн. операций/с – высокопроизводи­тельными.

Увеличение производительности приводит к увеличению стоимости и большим затратам энергии. Энерге­тические затраты важны для прило­жений в портативной аппаратуре, так как при больших затратах энергии со­кращается время между подзарядками батарей или возникает необходимость использовать батареи большей емко­сти, что, в свою очередь, приводит к большей цене и весу.

При синтезе и исследовании по­лосных вокодеров и полосных вокоде­ров с ЛПК используются различные модели речевого процесса. Наиболее точная модель речи представляет со­бой нестационарный случайный про­цесс с медленно меняющейся дисперси­ей и спектральной плотностью. При использовании подобной модели мож­но получить наиболее точный резуль­тат оценки качества вокодера.

## Ортогональные вокодеры

Речевой сигнал можно промо­делировать откликом на возбуждаю­щий сигнал линейной системы с им­пульсной характеристикой с пере­менными параметрами, так что выход­ной сигнал равен свертке возбу­ждающего сигнала и импульсного от­клика голосового тракта при условии, что форма голосового тракта неизмен­на.



Все разнообразие звуков получа­ется путем изменения формы голосо­вого тракта. Если форма голосового тракта изменяется медленно, то на ко­ротких временных интервалах аппрок­симация выходного сигнала сверткой возбуждающего сигнала и импульсно­го отклика голосового тракта справед­лива. Если на коротком отрезке вре­мени входной сигнал является перио­дическим, с постоянной входной часто­той, то выходной сигнал также являет­ся периодическим. Такая модель спра­ведлива для описания звонких звуков. Аналогично временному, преобразова­ние речи может быть описано в частот­ной области, поскольку преобразова­ние Фурье речевого сигнала равно про­изведению преобразований Фурье воз­буждающего сигнала и импульсного отклика голосового тракта.

Частотная характеристика голо­сового тракта является гладкой функ­цией частоты и характеризуется аку­стическими резонансами, называемы­ми формантными частотами.

Поскольку при изменении различ­ных звуков форма голосового тракта изменяется, то с течением времени бу­дет изменяться и огибающая спектра речевого сигнала. Так, в частности, при изменении периода сигнала, воз­буждающего звонкие звуки, частотный разнос между гармониками спектра бу­дет также изменяться.

Таким образом, для адекватного описания речевого сигнала надо не только знать вид его спектра, но и то, как он изменяется во времени.

Основным параметром речевого сигнала, возбуждающего звонкий звук, является разнос гармоник основно­го тона, а характеристики голосового тракта достаточно полно определяют­ся частотами формант.

Изложенное позволяет сделать вывод об удобстве частотного метода описания и преобразования речевых процессов на основе кратковременно­го спектрального анализа.

Рассмотрим основные принципы, положенные в основу построения орто­гональных вокодеров.

Запишем кратковременное пре­образование Фурье дискретизированного речевого процесса в виде



(2.1)



Здесь весовая функция, сдвига­емая во времени.



Соотношение (2.1) может быть пе­реписано в двух формах. Первая фор­ма имеет вид свертки

(2.2)



где – обозначает свертку.



Реализация (2.2) может быть представлена в виде рис. 2.2,а и озна­чает, что спектр речевого процесса со­ответствует свертке весовой функции с сигналом , промодулированным колебанием .



Другая форма записи (2.1) получается, если ее переписать в виде

(2.3)



Система (2.3) может быть реализована в виде, представленном на рис. 2.2,б и означает преобразование речевого сигнала полосовым фильтром центральной частотой ω и импульсной характеристикой .



Реализации, представленные на рис. 2.2,а,б, отличаются тем, что первом случае используется цифровой фильтр нижних частот с импульсной характеристикой , а во втором – полосовой фильтр, что удобно при параллельном измерении на нескольких частотах ω.



Используя алгоритмы БПФ, оценку кратковременного преобразования Фурье на равноотстоящих частотах можно записать в виде



(2.4)



где



Формулу (2.4) можно преобразовать к виду



где

(2.5)



можно рассматривать как характери­стику комплексного ПФ с центральной частотой .



Исходный речевой сигнал можно восстановить, сложив сигналы



на всех выходах гребенки ПФ так, что

(2.6)



Формула (2.4) является осно­вополагающим уравнением анализа с кратковременным преобразованием Фурье, а формула (2.6) – основным уравнением синтезатора.

## Гомоморфные вокодеры

В основе гомоморфных вокодеров лежит метод нелинейной (гомоморфной) фильтрации. Общая структу­ра гомоморфных систем, предназначенных для инверсной фильтрации речевых сообщений, представлена на рис. 2.4.

Свойства системы определяется соотношением , где и – Z-преобразования и соответственно.



Сигнал на выходе системы обычно называют комплексным кепстром. Система является линейной, а система – обратной к системе .



Удобства подобных преобразова­ний для анализа и синтеза рече­вых процессов обусловлены рядом свойств комплексного кепстра. В част­ности: комплексный спектр последова­тельностей, имеющих Z-преобразование, в основном сосредоточен вблизи нуля; последовательность, состоящая из равноотстоящих импульсов, имеет комплексный кепстр того же вида; для вычисления комплексного кепстра последовательности с мини­мальной фазой можно обойтись ло­гарифмом действительной, а не ком­плексной функции.

Выше было показано, что отрез­ки речевых сигналов могут быть пред­ставлены откликом линейной системы.

Так, в случае звонких звуков возбу­ждаемый сигнал имеет вид последова­тельности импульсов. В случае глу­хих звуков возбуждающий сигнал мо­жет быть смоделирован в виде шума.

Обычно предполагают, что пере­даточная функция линейной системы, имитирующей голосовой тракт, опи­сывается рациональной функцией Z. В результате, согласно перечисленным выше свойствам кепстра, комплексный кепстр импульсного отклика голосово­го тракта сосредоточен вблизи нуля.

В случае звонкого звука комплекс­ный кепстр возбуждающего сигнала состоит из импульсов, повторяющихся с периодом основного тона. То есть в случае звонкой речи комплексные кепстры возбуждающего сигнала и им­пульсного отклика голосового тракта занимают неперекрывающиеся времен­ные сигналы и могут быть извлечены из общего кепстра с помощью линей­ной системы L.

Механизм восстановления речево­го сигнала с помощью кепстров может быть пояснен следующим образом.

Поскольку спектр звонкого звука формируется умножением огибающей, характеризующей состояние голосово­го тракта, на функцию, описывающую тонкую структуру спектра возбужда­ющего сигнала, то логарифм спектра равен сумме логарифмов огибающей спектра и спектра возбуждающего сиг­нала.

Логарифм спектра возбуждающе­го сигнала изменяется с ростом часто­ты гораздо быстрее логарифма огиба­ющей спектра. Кроме того, он пери­одичен. В результате обратное пре­образование Фурье от логарифма оги­бающей спектра сконцентрировано по оси времени вблизи нуля, в то вре­мя как обратное преобразование от ло­гарифма спектра возбуждающего сиг­нала является линейчатым, отража­ющим его периодичность в частотной области.

Для выделения логарифма огиба­ющей спектра из полного спектра ло­гарифма его «взвешивают» окном, от­крытым только в начальном участке кепстра (вблизи нуля). Эту процедуру называют «сглаживанием кепстра».

В системе анализа-синтеза, осно­ванной на гомоморфной фильтрации, начальные значения кепстра служат параметрами, описывающими состоя­ние голосового тракта или огибающую спектра речевого процесса.

Значения кепстра при больших значениях времени используются для оценки параметров возбуждающего сигнала.

Таким образом, основная идея го­моморфной обработки заключается в разделении или обратной свертке сег­мента речевого сигнала с компонента­ми, представляющими собой импульс­ную характеристику и источник возбу­ждения. Это достигается путем линей­ной фильтрации обратного преобразо­вания Фурье логарифма спектра сигна­ла (кепстра). Гомоморфные вокодеры, как и любые другие вокодеры, в ко­торых осуществляется разделение па­раметров речи на сигнал возбуждения и параметры речевого тракта, позволя­ют достигнуть малой скорости переда­чи и дополнительной гибкости при об­работке речи ценой усложнения алго­ритмов преобразований.

## Кодирование с линейным предсказанием (LPC — Linear Predictive Coding). Рекомендации G.728, G.729, G.723

При кодировании с линейным предсказанием моделируются различные параметры человеческой речи, которые передаются вместо отсчетов или их разности, требующих значительно большей пропускной способности канала. Следует заметить, что буферы, необходимые для хранения потоков данных, увеличивают задержку кодирова­ния.

Первые реализации LPC, такие как LPC-вокодер, были предназначены ля передачи данных на низких скоростях – 2,4 и 4,8 кбит/с. На скорости 2,4 кбит/с обеспечивался приемлемый уровень разборчивости речи, однако качество, естественность и узнаваемость речи недостаточны. Поскольку этот метод сильно зависит от точного воспроизведения человеческой речи, его реализации, такие как LPC-вокодер, не подходят для сигналов неречевого происхождения, например сигналов модема.

Широко используемый в настоящее время метод кодирования с линейным предсказанием работает с блоками отсчетов, для каждого из которых вычисляется и передается частота основного тона, его амплитуда и информация о типе возбуждающего воздействия.

Структура синтезатора речи с линейным предсказанием показана на рис. 2.5. Здесь управляющий вход или сигнал возбуждения смоделирован в виде последовательности импульсов на частоте основного тона (для вокализо­ванной речи) или случайный шум (для невокализированной речи).

Комбинированные спектральные составляющие потока от голосовых связок, голосового тракта и звукообра­зования за счет губ могут быть пред­ставлены цифровым фильтром с изме­няющимися параметрами и передаточ­ной функцией

(2.7)



где



Параметрами, характеризующими голосовой тракт, являются коэффици­енты знаменателя и масштабный мно­житель G.

Преобразуя уравнение (2.7) во вре­менную область, можно получить раз­ностное уравнение для импульсной ха­рактеристики , соответствующей :



(2.8)



Уравнение (2.8) называют раз­ностным уравнением LPC. Оно устана­вливает, что текущее значение выход­ного сигнала может быть опреде­лено суммированием взвешенного те­кущего входного значения и взвешен­ной суммы предыдущих выходных вы­борок. Следовательно, в LPC анализе проблема может быть сформулирова­на так: даны измерения сигнала, требуется определить параметры пере­даточной функции системы .



Линейное предсказание при ана­лизе речевых сигналов обычно исполь­зуется в двух направлениях. Одно из них – проведение кратковременно­го спектрального анализа речи. Вто­рое направление – построение систем анализа-синтеза.

Параметры, входящие в функцию предсказания, через формулу (2.7) определяют параметры передаточной функции голосового тракта. Может быть предложено несколько вариантов структуры анализатора, пригодных для построения синтезатора и реализующих передаточную функцию голосового тракта. Структуру прямой формы можно получить непосредственно по коэффициентам функции предсказания. С другой стороны, дробь (2.7) можно преобразовать в произведение и получить структуру каскадной формы.

Во всех случаях параметры синтезатора непрерывно обновляются при смене анализируемых кадров речи. Чтобы избежать эффектов, связанных со скачками значений параметров, необходимо плавно изменять параметры с помощью интерполяции при переходе от одного участка речи к другому. При прямой форме синтеза может возникать ситуация, соответствующая неустойчивому фильтру, хотя исходные значения относились к устойчивому фильтру. В каскадной структуре устойчивость обеспечивается проще.Определение параметров возбу­ждающего сигнала в системе анализа-синтеза с линейным предсказанием, как правило, основывается на иссле­довании сигнала ошибки, получаемого пропусканием исходного речевого сиг­нала через фильтр с характеристикой, обратной той характеристике, которая аппроксимирует передаточную функ­цию голосового тракта. Полученный сигнал ошибки является аппроксима­цией сигнала, возбуждающего речевое колебание. Для определения параме­тров возбуждающего сигнала можно применить один из известных алгорит­мов различения звонкой и глухой ре­чи, а также оценки периода основного тона, например на основе рассмотрен­ного выше корреляционного анализа сигналов во временной области.

## Кодирование речи методами анализа через синтез (AbS)

При классификации методов ко­дирования речи на скоростях 4,8... ...16 кбит/с выделяют две основные группы — методы анализа и синтеза (AaS — Analysis-and-Synthesis) и ме­тоды анализа через синтез (AbS – Analysis-by-Synthesis). Хотя такие схе­мы AaS, как RELP, АРС, АТС и SBC успешно работают на скоростях 9,6... 16 кбит/с, при скоростях ниже 9,6 кбит/с они не могут обеспечивать хорошее качество речи. Это объясняется двумя причинами: 1) кодируе­мая речь не анализируется на предмет эффективности, т.е. не производится коррекция искажений в восстановлен­ной речи; 2) ошибки, накопленные в предыдущих фреймах, не учитывают­ся в момент анализа текущего фрейма и беспрепятственно переходят в следу­ющие фреймы. В схемах AbS, особенно в AbS-LPC, эти факторы, как правило, учтены. В этих схемах используется процедура оптимизации типа «замкну­тая петля» для нахождения возбужда­ющего сигнала, который при возбужде­нии моделирующего фильтра создает оптимальный речевой сигнал. Это по­зволяет схемам AbS более успешно ра­ботать на скоростях 4,8.. .9,6 кбит/с.

Методы AbS подходят не только для кодирования речи, но могут так­же использоваться для оценки и идентификации. Основная идея AbS такова. Во-первых, допускается, что сигнал можно исследовать и представить в какой-либо форме, например в виде временных или частотных доменов. Затем созданная модель сигнала подвергается оптимизации (подгонке), как показано на рис. 2.6.

Модель имеет несколько параметров, изменение которых приводит к изменению формы моделируемого сигнала. Для нахождения модели сигнала, которая имеет ту же форму, что и модель истинного сигнала, используют процедуры минимизации ошибки. Путем изменении параметров модели находят такой их набор, при котором синтезированный сигнал с минимальной погрешностью совпадает с реальным. Следовательно, когда достигнуто такое совпадение, параметры модели принимаются за параметры истинного сигнала.

Базовая структура системы кодирования AbS-LPC представлена на рис. 2.7. В этой модели есть три компонента, которые можно изменять, добиваясь максимального подобия синтезированного сигнала с исходным:

1) нестационарный фильтр;

2) возбуждающий сигнал;

3) процедура минимизации, основанная на восприятии.

Так как эта модель требует ча­стого обновления параметров для по­лучения хорошего совпадения с исход­ным сигналом, процедура анализа вы­полняется поблочно, т.е. входной рече­вой сигнал разбивается на блоки вы­борок. Длина анализируемых блоков (фреймов) и периодичность их обно­вления определяют скорость передачи (емкость) схемы кодирования. Алго­ритм работы AbS-LPC следующий:

1. Инициализировать LPC и то­нальный фильтры (нестационарные фильтры), т.е. установить уровень ну­ля или минимального случайного шу­ма;

2. Фрейм выборок речи заносится в буфер и на основании LPC-анализа вычисляется набор LPC-коэффициентов;

3. Используя вычисленные LPC-коэффициенты, формируется инверс­ный LPC-фильтр для вычисления пер­вого восстановленного после квантова­ния остатка. Если для поиска подхо­дящего тона используется «замкнутая петля», надобность в этом шаге отпа­дает.

4. Так как LPC фрейм обычно слишком велик для эффективного ана­лиза, при определении возбуждения фрейм разделяется на целое число подфреймов;

5. Для каждого подфрейма:

а) рассчитываются параметры тонального фильтра (долгосрочного предсказателя), такие, как задержка и связанный с ней коэффициент мас­штабирования;

б) тональный фильтра вместе с LPC-фильтром образуют каскадный фильтр, с помощью которого опре­деляется наилучшее вторичное возбу­ждение, т.е. такое, которое минимизи­рует разницу между синтезированной и исходной речью.

6. Окончательно синтезирован­ная речь получается при пропускании оптимального вторичного возбуждения через каскадный фильтр, параметры которого остались от синтеза предыду­щего подфрейма.

7. Повторение шагов 2-6 для сле­дующего фрейма последовательности.

Таким образом, и на стороне коде­ра, и на стороне декодера синтезирует­ся речь, что необходимо для обновле­ния содержимого памяти нестационар­ных фильтров. В результате и кодер, и декодер имеют идентичное содержа­ние памяти. В противном случае для общей синхронности содержимое памя­ти пришлось бы передавать на деко­дер. Действительно, главный вопрос в схемах AbS-LPC — как сохранить это одинаковое состояние в кодере и деко­дере, когда средства передачи несовер­шенны, например, в системах подвиж­ной радиосвязи, где очень высока доля ошибок.

Может показаться, что схема AbS-LPC не является полноценной схемой «анализа через синтез». Это связа­но с тем, что в действительности про­цедуры последовательны, т.е. сначала вычисляются параметры фильтра, ко­торые фиксируются, и только затем следует вычисление методом «анали­за через синтез» вторичного возбужде­ния. Хотя вторичное возбуждение вы­полняется по исходному сигналу, оно ограничено оптимальностью использу­емых фильтров. Поэтому, в идеале, требуется наилучшая комбинация как возбуждения, так и фильтров, которая означает одновременную оптимизацию всех параметров. Эта процедура очень сложна, насыщена вычислениями, по­этому ее обычно разбивают на после­довательные этапы.

Главное отличие классических во­кодеров от кодеров AbS-LPC состоит в том, что в классических вокодерах воз­буждение разделяется на вокализован­ные (импульсное возбуждение) и не­вокализованные (возбуждение случай­ным шумом), что является первопри­чиной точности модели. В AbS-LPC такое деление не явно, и поэтому воз­буждающий сигнал может носить лю­бой характер — от псевдоимпульсного до шумоподобного, что позволяет син­тезировать речь более высокого каче­ства.

## Векторное квантование и кодовые книги

Когда набор значений амплитуд, дискретизированных по времени, кван­туется совместно как единичный век­тор, такой процесс называется век­торным квантованием (VQ – vector quantisation), известный также как блочное квантование.

Будем считать, что *N*-мерный вектор с дей­ствительными значениями («т» озна­чает транспонирование); – случайным образом меняющий­ся компонент с непрерывной амплиту­дой. При векторном квантовании век­тору ставится в соответствие другой *N*-размерный вектор , имеющий дей­ствительные значения и дискретную амплитуду. Таким образом, кванту­ется как . Другими словами, ис­пользуется для представления .



Обычно выбирается из конечного набора значений , где – размер кодовой книги, а – набор векторов кодовой книги. Набор *Y* называется кодовой книгой или шаблоном.



Размер кодовой книги можно счи­тать равным числу уровней скалярных квантователей. Для создания подоб­ной кодовой книги N-размерное про­странство разделяется на *L* областей или ячеек , и вектор однозначно связывается с ячейкой . Квантователь обозначается вектором кодовой книги , если находит­ся в :



, если .



Процесс создания кодовой книги известен также как «обучение» или «настройка» кодовой книги. В каче­стве примера на рис. 2.9 иллюстриру­ется разделение двумерного простран­ства (N = 2) для целей векторно­го квантования. Область, обведенная жирной линией, — ячейка . При векторном квантовании любой входной вектор , лежащий в ячейке , кван­туется как . Другие векторы кодовой книги, соответствующие другим ячей­кам, показаны точками.



Если размер вектора , век­торное квантование трансформируется в скалярное квантование. Скалярное квантование имеет особое свойство, за­ключающееся в том, что хотя ячейки могут иметь разные размеры (размеры ступеней), все они имеют одинаковую форму. Однако при векторном кванто­вании ячейки в двух измерениях могут иметь разные формы, что дает векторному квантованию преимущество над скалярным квантованием.



Гибридные кодеры

Чтобы избавиться от недостатков кодеров формы и вокодеров, был раз­работан гибридный метод кодирова­ния, объединяющий преимущества обоих методов. По виду анализа гибрид­ные кодеры подразделяются на два класса: с частотным разделением и временным разделением.

## Гибридные кодеры с частотным разбиением

Главная концепция кодирования с частотным разбиением состоит в раз­делении речевого спектра на частот­ные полосы или компоненты. Со­ответственно могут использоваться либо набор фильтров, либо блок-преобразователь. После кодирования и декодирования эти составляющие ис­пользуются для точного воспроизве­дения модели входного сигнала путем суммирования сигналов, полученных на выходе фильтров, или инверсных значений, полученных после преобра­зования. Главное допущение при ко­дировании с частотным разбиением со­стоит в том, что сигнал, подвергаемый кодированию, очень медленно изменя­ется во времени и может быть описан мгновенным спектром. Это связано с тем, что в большинстве систем, а осо­бенно в системах реального времени, в текущий момент доступен только крат­ковременный сегмент входного сигна­ла.

В случае использования набора фильтров частота ω фиксирована, так что , а сигнал частотного доме­на представляет собой сигнал на выходе постоянного во времени ли­нейного фильтра с импульсной харак­теристикой , возбуждаемого моду­лированным сигналом :



(3.1)



где определяет ширину полосы ре­чевого сигнала вокруг централь­ной частоты и является импульс­ной характеристикой анализирующе­го фильтра; знак означает свертку функций.



При использовании блока, реали­зующего преобразование Фурье, вре­менной индекс *h* фиксируется на зна­чении *h = ho*, a представляет собой обычное преобразование Фурье взвешенной последовательности :



(3.2)



где – преобразование Фурье.



Здесь определяет отрезок времени анализа относительно момен­та времени *h = ho* и является «окном анализа» .



Уравнение синтезирующего набо­ра фильтров

(3.3)



может быть представлено как инте­грал (или сумма) компонентов – крат­ковременных спектров с не­сущими частотами .



Для синтеза с помощью блока пре­образования уравнение выглядит сле­дующим образом:

(3.4)



Его можно интерпретировать как сум­му инверсных преобразований Фурье, примененных к временным сигналам .



## CELP (Code Excited Linear Prediction)

Метод кодирования CELP основан на линейной авторегрессионной моде­ли процесса формирования и воспри­ятия речи и входит в группу мето­дов анализа через синтез, реализую­щих современные и эффективные ал­горитмы информационного сжатия ре­чевых сигналов. Алгоритмы данного класса занимают промежуточное поло­жение между кодерами формы сигна­ла, в которых сохраняется форма коле­бания речевого сигнала в процессе его дискретизации и квантования, и пара­метрическими вокодерами, основанны­ми на процедурах оценки и кодирова­ния небольшого числа параметров ре­чи, объединяя преимущества каждого из них.

Линейная авторегрессионная мо­дель процесса формирования речевых сигналов с локально постоянными на интервалах 10. . .30 мс параметрами получила в настоящее время наиболь­шее распространение. Для этой моде­ли

(3.5)



где *М* — порядок модели; – последовательность отсчетов речево­го сигнала; – коэффициенты линейного предсказания, характери­зующие свойства голосового тракта; – порождающая последователь­ность или сигнал возбуждения голосо­вого тракта.



Авторегрессионная модель речево­го сигнала описывает его с достаточ­но высокой степенью точности и по­зволяет применять развитый матема­тический аппарат линейного предска­зания. При этом обеспечивается более высокое качество декодированной ре­чи, устойчивость к входному акустиче­скому шуму и ошибкам в канале свя­зи по сравнению с системами с иными принципами кодирования.

В рамках данной модели наибо­лее перспективными методами кодиро­вания считаются методы «анализа че­рез синтез» с использованием многоимпульсного возбуждения. Новизна мно­гоимпульсного возбуждения заключа­ется в том, что в сигнале остатка ли­нейного предсказания выбираются та­кие его значения, которые наиболее важны для повышения качества синте­зированной речи. При этом использу­емая в процедуре анализа через син­тез схема кодирования, помимо учета ошибок квантования, включает крите­рии субъективной оценки качества ре­чевого сигнала, что обеспечивает есте­ственное звучание синтезированной ре­чи.

При многоимпульсном возбужде­нии сигнал остатка линейного предска­зания представляется в виде последо­вательности импульсов с неравномер­но распределенными интервалами и с различными амплитудами (около 8-10 импульсов за 10 мс). Амплитуды и по­ложение этих импульсов определяют­ся на покадровой основе (кадр за ка­дром). Основным преимуществом мно­гоимпульсного возбуждения является то, что она определяется для любого речевого сегмента и при этом не тре­буется знаний ни о вокализованности данного сегмента, ни о периоде основ­ного тона.

Методы анализа через синтез ис­пользуют синтезатор (декодер) ре­чевого сигнала как составную часть устройства кодирования. При этом задача анализа сводится к процедуре оценки передаваемых в канал связи па­раметров речи, проводимой в соответ­ствии с некоторым критерием рассо­гласования между исходным и декоди­рованным сигналами. Для учета спе­цифики слухового восприятия в каче­стве критерия рассогласования обычно используется взвешенная по частоте квадратическая ошибка

(3.6)



где и – преобразование Фу­рье исходного и синтезированного ре­чевых сигналов; – весовая функ­ция. Принимая во внимание важ­ность для восприятия речи не только формант, но и межформантных обла­стей, для алгоритмов анализа речи че­рез синтез в качестве эталонной была предложена весовая функция следую­щего вида:



(3.7)



где – передаточная харак­теристика синтезирующего фильтра; γ – параметр, регулирующий энер­гию ошибки или шум квантования. Фактически при таком окне взвешива­ния подчеркивается ошибка в межфор­мантных областях и тем самым обеспе­чивается более равномерное по часто­те распределение отношения мощности полезного сигнала к мощности ошибки кодирования.



В алгоритмах кодирования с «анализом через синтез» повышение эффективности информационного уплотнения речевых сигналов производится, преимущественно, за счет сокращения избыточности последовательности *x(h)*, которая осуществляет возбуждение синтезирующего фильтра линейного предсказания, формирующего огибающую сигнала, с коэффициентом передачи



(3.8)



Для этой цели применяется также дополнительный фильтр с характеристикой

(3.9)



с коэффициентом предсказания и задержкой на период основного тона *T*. Фильтр выполняет функции генератора квазипериодических колебаний голосовых связок при произношении вокализованных звуков.



В зависимости от способа описания сигнала *x(h)*, поступающего на вход фильтра (3.9), можно выделить алгоритмы кодирования:

с возбуждением прореженной последовательности импульсов — MPLP (Multi Pulses Linear Prediction);

с самовозбуждением — SELP (Self Excited Linear Prediction);

с кодовым возбуждением — CELP

Экспериментально установлено, что кодовое возбуждение обеспечивает наиболее высокое качество кодирования речевого сигнала, в том числе и при наличии входных акустических помех.

CELP наиболее эффективно применяется при передаче речевого сиг­ала в диапазоне скоростей от 4 до 6 кбит/с.

По существу, в алгоритме CELP производится векторное квантование последовательности , т.е. позиции выборок и их амплитуды в сигнале многоимпульсного возбуждения оптимизируются одновременно. При том отрезок (сегмент) сигнала возбу­ждения выбирается из предварительно формированной постоянной совокупности – кодовой книги, содержащей достаточно большое количество реализаций, например, некоррелированного гауссовского шума. Выбранная реализация усиливается и подается на вход цепочки фильтров (3.9) и (3.8).



Поиск оптимальных значений и *Т* синтезатора основного тона, коэф­фициента усиления и номера элемента кодовой книги осуществляется посредством «анализа через синтез». В канал связи передаются номер (индекс) элемента кодовой книги с соответству­ющим коэффициентом усиления, параметры синтезатора основного тона, а также коэффициенты линейного предсказания, характеризующие состояние голосового тракта.



Являясь одной из самых распро­страненных, схема с линейным пред­сказанием и возбуждением от кода CELP является лучшей схемой AbS-LPC для низких скоростей. В CELP имеется линейный фильтр с изменяю­щимися во времени параметрами для выделения грубой и точной спектраль­ной информации. Возбуждение выпол­няется путем перебора всех векто­ров из возбуждающей кодовой кни­ги. Векторная последовательность, обеспечивающая минимальную взве­шенную ошибку, считается оптималь­ным возбуждением. Процедура AbS в CELP требует больших вычислитель­ных ресурсов, а основная кодовая кни­га является результатом очень боль­шой исследовательской работы. Хо­тя CELP является сложным методом, он способен синтезировать речь с высо­ким качеством даже на низких скоро­стях. Вариант кодирования CELP вы­бран для многих систем голосовой свя­зи.

Хотя CELP, главным образом, ориентирован на низкие скорости, на нем базируются многие стандарты. Испытания показывают его приемле­мость и для высоких скоростей. Стан­дарт для скорости 16 кбит/с с ма­лой задержкой (LD-CELP — Low-Delay CELP) будет рассмотрен ниже.

## Рекомендации G.723.1 и G.729

Рекомендация G.723.1 определяет кодовое представление, которое может использоваться на очень низких скоро­стях для компрессии речевых или дру­гих аудиосигналов в средствах муль­тимедиа. В кодере, реализующем ре­комендации G.723.1, принципиальным приложением является низкоскорост­ная видеотелефония как часть общего семейства стандартов Н.324.

Кодер обеспечивает работу на двух скоростях — 5,3 и 6,3 кбит/с. Более высокая скорость обеспечивает лучшее качество. Тем не менее, и бо­лее низкая скорость обеспечивает хо­рошее качество и предоставляет разра­ботчикам систем связи дополнитель­ные возможности. И кодер и декодер должны обязательно поддерживать обе скорости. Существует возможность пе­реключения скоростей. Возможно так­же изменение рабочей скорости с ис­пользованием прерывистой передачи и заполнение шумом пауз.

Кодер G.723.1 оптимизирован для сжатия речи с высоким качеством на установленной скорости при ограни­ченной полосе. Музыка и другие ау­диосигналы также могут быть под­вергнуты компрессии с использовани­ем этого кодера, однако, не с таким же высоким качеством, как речь.

Кодер G .723.1 преобразует речь или другие аудиосигналы во фрей­мы длительностью 30 мс. Кроме то­го, существует возможность просмотра фреймов на скорости 7,5 мс, что приво­дит к общей алгоритмической задерж­ке 37,5 мс. Дополнительные задержки возникают из-за:

времени, затрачиваемого на обра­ботку данных в кодере и декодере;

времени передачи по линии связи;

дополнительной буферной задержки протокола мультиплексирования.

Кодер G.723.1 предназначен для работы с цифровыми сигналами после предварительной фильтрации полосы аналогового телефонного канала (рекомендации G.712), дискретизации с частотой 8 кГц и преобразования в 16-битную линейную ИКМ последовательность для передачи на вход кодера. Выходной сигнал декодера преобразуется обратно в аналоговый сигнал аналогичным образом. Другие характеристики входа/выхода такие же, как и определенные рекомендациями G.711 для 64-битной ИКМ. Перед кодированием данные должны быть преобразованы в 16-битную ИКМ последовательность или в соответствующий формат после декодирования из 16-битной ИКМ.

Кодер, основанный на принципах кодирования методом «анализ через синтез» с линейным предсказанием, минимизирует взвешенный сигнал ошибки, работает с блоками (фреймами) по 240 выборок каждый, что в частоте дискретизации 8 кГц эквивалентно длительности 30 мс. Каждый фрейм проходит через фильтр верхних частот для удаления постоянной составляющей, а затем разделяется на четыре субфрейма по 60 выборок в каждом. Для каждого субфрейма используется фильтр десятого порядка кодера с линейным предсказанием. Для последнего субфрейма коэффициенты LPC-фильтра квантуются с использованием прогнозирующего квантизатора вектора разбиения (PSVQ). Квантованные LPC-коэффициенты используются для создания кратковременного взвешивающего фильтра, который применяется для фильтрации всего фрейма и для получения взве­шенной оценки речевого сигнала. На основе этой оценки для каждых двух субфреймов (120 выборок) вычисляется период основного тона . Оценка тона представляется блоками по 120 выборок. Период основного тона лежит в диапазоне от 18 до 142 выборок.



С помощью вычисленной заранее оценки периода тона создается фильтр формы гармонического шума. Комбинация из фильтра синтеза LPC, филь­тра взвешивания формант, фильтра формы гармонического шума используется для синтеза импульсной харак­теристики, необходимой для дальнейших вычислений.

Оценки периода основного тона и импульсного отклика используются при работе предсказателя тона пятого порядка. Период тона вычи­сляется как приращение относительной оценки периода основного тона. На декодер передаются тоновый период и разностные величины. На следующем этапе аппроксимируются непериодиче­ские составляющие возбуждения. Для высокой скорости используется много­импульсное возбуждение с квантовани­ем и алгоритмом максимального прав­доподобия (MP-MLQ), а для низких скоростей – алгебраическое кодовое возбуждение.



Рекомендации ITU-T G.729 со­держат описание алгоритма кодирова­ния речевых сигналов на скорости 8 кбит/с с использованием алгебраиче­ского линейного предсказания с ко­довым возбуждением с сопряженной структурой (CS-ACELP).

Подобный кодер создан для ра­боты с цифровыми сигналами, полу­ченными после предварительной об­работки аналогового входного сигна­ла фильтром низкой частоты, дискре­тизации с частотой 8 кГц и даль­нейшем преобразованием в линейную ИКМ для подачи на вход кодера. Вы­ходной сигнал декодера конвертирует­ся обратно в аналоговый сигнал по­добным же образом. Другие харак­теристики входа/выхода определяют­ся аналогично рекомендациями G.711 для ИКМ последовательностей со ско­ростью 64 кбит/с. После декодирова­ния данные должны быть преобразо­ваны из 16-битовой линейной ИКМ в требуемый формат.

Кодер CS-ACELP основан на моде­ли с линейным предсказанием с кодо­вым возбуждением (CELP) и работает с фреймами речи по 10 мс, соответству­ющих 80 выборкам. Каждый фрейм речевого сигнала длительностью 10 мс анализируется для выделения пара­метров CELP-модели (коэффициенты фильтра линейного предсказания, ин­дексы адаптивной и фиксированной кодовых книг и коэффициенты усиле­ния). Эти параметры кодируются и пе­редаются на приемную сторону. Рас­пределение бит параметров кодера по­казано в табл. 3.1.

Таблица 3.1

Распределение бит для алгоритма CS-ACELP на скорости 8 кбит/с (фреймы по 10 мс)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Параметр | Кодовое слово | Субфрейм 1 | Субфрейм 2 | В целом на фрейм |
| Пары линейного спектра | LU, L1, L2, L3 |  |  |  |
| Задержка адаптивной кодовой книги | P1, Р2 | 8 | 5 | 13 |
| Проверка задержки тона | Р0 | 1 |  | 1 |
| Индекс фиксированной кодовой книги | CT, С2 | 13 | 13 | 26 |
| Запись фиксированной кодовой книги | S1, S2 | 4 | 4 | 8 |
| Усиления кодовой книги (этап 1) | GA1, GA2 | 3 | 3 | 6 |
| Усиления кодовой книги (этап 2) | GBl, GB2 | 4 | 4 | 8 |
| Всего |  |  |  | 80 |

На стороне деко­дера эти параметры используются для восстановления параметров возбужде­ния и фильтра синтеза. Как показа­но на рис. 3.3, речь восстанавливает­ся при фильтрации этого возбуждения фильтром кратковременного синтеза, который основан на фильтре линей­ного предсказания десятого порядка. Долговременный фильтр (или фильтр синтеза тона) выполняется с использо­ванием адаптивной кодовой книги. По­сле синтеза речи происходит дополни­тельное сглаживание в постфильтре.

Входной сигнал посту­пает на фильтр высоких частот и мас­штабируется в блоке предварительной обработки, после чего подвергается по­следующему анализу. Анализ с линей­ным предсказанием (LP-анализ) вы­полняется один раз для фрейма дли­тельностью 10 мс с целью вычисле­ния коэффициентов фильтра линейно­го предсказания, которые затем пре­образуются в пары линейного спектра (Line Spectrum Pairs, LSP) и квантуют­ся (18 бит) с использованием двухэтапного векторного квантования с предсказанием.

Сигнал возбуждения выбирается с использованием поисковой процедуры «анализ через синтез», при которой ошибка между исходной и восстанавливаемой речью минимизируется в соответствии с измерением взвешенных искажений. Это выполняется путем фильтрации сигнала ошибки фильтром взвешивания, коэффициенты которого извлечены из неквантованного LP-фильтра.

Параметры возбуждения (параметры фиксированной и адаптивной кодовых книг) определены для субфрейма длительностью 5 мс (40 выборок). Коэффициенты квантованного и неквантованного фильтра с линейным предсказанием используются для второго субфрейма, в то время как в первом субфрейме используются интерполированные коэффициенты LP-фильтра.

Задержка основного тона оцени­вается один раз для фрейма длиной 10 мс на основе взвешенного речево­го сигнала. Затем для каждого суб­фрейма повторяются следующие опе­рации. Искомый сигнал вычисля­ется при фильтрации остаточного ли­нейного предсказания во взвешиваю­щем фильтре синтеза . При фильтрации ошибки начальные состо­яния этих фильтров обновляются. Это эквивалентно результату выделения нулевого входного отклика взвешива­ющего фильтра синтеза из взвешенно­го речевого сигнала. Вычисляется им­пульсная характеристика взвеши­вающего фильтра синтеза, после че­го выполняется анализ тона для нахо­ждения задержки адаптивной кодовой книги путем анализа значения задерж­ки вблизи основного тона с использова­нием искомого сигнала и импульс­ной характеристики . Задержка тона кодируется восемью битами в пер­вом субфрейме и пятью битами во вто­ром субфрейме. Искомый сигнал используется при поиске фиксирован­ной кодовой книги для нахождения оптимального возбуждения. Семна­дцатибитовая алгебраическая кодовая книга используется для возбуждения фиксированной кодовой книги. Коэф­фициенты усиления вкладов адаптив­ной и фиксированной кодовых книг — это векторы, квантованные семью би­тами.



Индексы параметров кодовых книг вы­деляются из принятого потока бит и декодируются для получения следую­щих параметров кодера, соответствую­щих речевому фрейму длиной 10 мс: LP-коэффициенты (коэффициенты ли­нейного предсказания), две частичные задержки тона, два вектора фиксиро­ванной кодовой книги и два набора ко­эффициентов адаптивной и фиксиро­ванной кодовых книг. Коэффициенты LSP интерполируются и преобразуют­ся в коэффициенты LP-фильтра для каждого субфрейма. Для каждого суб­фрейма выполняются следующие ша­ги:

восстанавливается возбуждение путем добавления векторов адаптив­ной и фиксированной кодовых книг с соответствующими им коэффициента­ми усиления;

восстанавливается речь путем пропускания через фильтр LP-синтеза;

восстанавливаемый речевой сиг­нал пропускается через ступень пост­обработки, которая включает адаптив­ный постфильтр, состоящий из долго­временного и кратковременного пост­фильтров синтеза, фильтр высоких ча­стот и операцию масштабирования.

Кодер кодирует речь и другие ау­диосигналы по фреймам длительно­стью 10 мс. В результате осуществля­ется задержка 5 мс, что приводит в ре­зультате к общей алгоритмической за­держке 15 мс. Все дополнительные за­держки при практическом исполнении такого кодера обусловлены следующи­ми причинами:

временем обработки, необходимым для операции кодирования и декодиро­вания;

временем передачи по линиям свя­зи;

задержкой мультиплексирования, когда аудиоданные объединяются с другими данными.

Таким образом, рекомендация G.729 предусматривает фреймы возбу­ждения по 5 мс и формирует четыре импульса. Фрейм из 40 выборок разде­ляется на четыре части. Первые три имеют восемь возможных позиций для импульсов, четвертая — шестнадцать. Из каждой части выбирается по одно­му импульсу. В результате образуется четырехимпульсный ACELP возбужде­ния кодовой страницы (табл. 3.2).

Таблица 3.2

Параметры кодеров

|  |  |  |  |
| --- | --- | --- | --- |
| Параметры кодера | Кодер | | |
| G.729 | G.729A | G.723.1 |
| Скорость бит, кбит/с | 8 | 8 | 5,3…6,3 |
| Размер фрейма, мс | 10 | 10 | 30 |
| Размер подфрейма, мс | 5 | 5 | 7,5 |
| Алгебраическая задержка, мс | 15 | 15 | 37,5 |
| Быстродействие, млн. оп./с | 20 | 10 | 14…20 |
| Объем ПЗУ, байт | 5,2 К | 4 К | 4,4 К |
| Качество | Хорошее | Хорошее | Хорошее |

Для режима 5,3 кбит/с рекомен­дация G.723.1 предусматривает фрей­мы возбуждения длительностью 7,5 мс и также использует четырехимпульсное ACELP-возбуждение кодовой стра­ницы. Для скорости 6,3 кбит/с ис­пользуется технология многоимпульс­ного возбуждения с квантованием и ал­горитмом максимального правдоподо­бия (MP-MLQ). В этом случае пози­ции фреймов группируются в подгруп­пы с четными и нечетными номерами. Для определенного номера импульса из четной последовательности (пятый или шестой в зависимости от того, является ли сам фрейм четным или не­четным) используется последователь­ный многоимпульсный поиск. Похо­жий поиск повторяется для подфрей­мов с нечетными номерами. Для возбуждения выбирается группа с мини­мальными общими искажениями.

На стороне декодера информа­ция кодера с линейным предсказа­нием (LPC) и информация адаптив­ной и фиксированной кодовой книг демультиплексируется и использует­ся для реконструкции выходного сиг­нала. Для этих целей используется адаптивный постфильтр. В случае ко­дера G.723.1 сигнал возбуждения пе­ред прохождением через фильтр син­теза LPC пропускается через LT (long-term — долговременный) постфильтр и ST (short-term — кратковременный) постфильтр.

## LD-CELP (Long-Delay CELP). Рекомендация G.728

В Рекомендации содержится опи­сание алгоритма кодирования речевых сигналов на скорости 16 кбит/с с помо­щью линейного предсказания с кодиро­ванием сигнала возбуждения с малой задержкой. Алгоритм LD-CELP опи­сывает работу кодера и декодера.

В алгоритме LD-CELP сохране­на суть метода CELP, представляю­щего собой метод «анализа через син­тез» путем поиска сигналов в кодовой книге. Для получения алгоритмиче­ской задержки порядка 0,625 мс ис­пользуется адаптация предсказателей и уровней сигнала возбуждения по вы­ходу. Передается только индекс сиг­нала возбуждения, найденный в кодо­вой книге. Обновление коэффициен­тов предсказания производится с помо­щью LPC-анализа ранее квантованной речи. Уровень возбуждения обновля­ется с помощью информации, содержа­щейся в ранее квантованном сигнале возбуждения. Размер блока для ада­птации вектора сигнала возбуждения и уровня составляет всего лишь пять отсчетов. Обновление взвешивающего фильтра, учитывающего восприятие, производится с помощью LPC-анализа неквантованной речи.

После выполнения преобразова­ния сигнала ИКМ по закону *А* или μ в линейный ИКМ-сигнал входной сигнал делится на блоки по пять последова­тельных отсчетов. Для каждого вход­ного блока кодер пропускает каждый из 1024 векторов кодовой книги (хра­нящихся в кодовой книге сигнала воз­буждения) через устройство масштаби­рования уровня сигнала возбуждения и синтезирующий фильтр. Из полу­ченных в результате пропускания всех 1024 векторов-кандидатов квантован­ного сигнала кодер определяет один, минимизирующий величину взвешен­ной по частоте среднеквадратической ошибки относительно вектора входно­го сигнала. 10-битовый индекс, соот­ветствующий наилучшему вектору в кодовой книге, который соответствует наилучшему вектору-кандидату кван­тованного сигнала, передается в деко­дер. На следующем этапе для обно­вления памяти фильтра и подготов­ки к кодированию следующего векто­ра сигнала наилучший кодовый вектор проходит через устройство масштабирования уровня сигнала возбуждения и синтезирующий фильтр. Коэффициенты синтезирующего фильтра и уровень сигнала возбуждения периодически обновляются путем адаптации по выходу, базирующейся на квантованном сигнале, масштабированном по уровню, и сигнале возбуждения.

Индекс в книге векторного квантования (VQ) возбуждения представляет собой единственную информа­цию, которая в явной форме переда­ется из кодера в декодер. Три других типа параметров: уровень сигналa возбуждения, коэффициенты синтезирующего фильтра и коэффициенты взвешивающего фильтра, учитыва­ющего восприятие, обновляются пери­одически. Эти параметры получаются путем адаптации по выходу из сигна­лов, которые появляются до текуще­го вектора сигнала. Уровень сигнала возбуждения обновляется для каждого вектора, а коэффициенты взвешиваю­щего фильтра, учитывающего воспри­ятие, и коэффициенты синтезирующе­го фильтра обновляются для каждых четырех векторов (т.е. для каждых 20 отсчетов или для периода обновления длительностью 2,5 мс). Следует отме­тить, что хотя последовательность об­работки в алгоритме имеет цикл ада­птации, равный четырем векторам (20 отсчетов), емкость основного буфера составляет только один вектор (пять отсчетов). Такая малая емкость буфе­ра позволяет получить задержку при передаче в одном направлении менее 2 мс.

## Многополосное кодирование и кодирование с адаптивным преобразованием

Среди методов кодирования с ча­стотным разбиением известны две тех­нологии: многополосное кодирова­ние — SBC (Sub-Band Coding) и ко­дирование с адаптивным преобразо­ванием — АТС (Adaptive Transform Coding). Основной принцип обеих схем — разделение спектра входного на несколько частотных поддиапазо­нов (полос), которые затем кодируют­ся отдельно. В SBC набор фильтров выполнен так, что разбивает входной речевой сигнал обычно на 4-16 широ­ких частотных поддиапазонов (широ­кополосный анализ). В АТС для обес­печения более точных частотных пока­зателей число поддиапазонов увеличе­но до 128-256 (узкополосный анализ).

Многополосное кодирование обыч­но рассматривается как метод коди­рования формы сигнала, который ис­пользует широкополосный кратковре­менный анализ и синтез. После раз­деления речевого спектра на несколь­ко поддиапазонов низшая частота ка­ждого из них приводится к нулю, затем поддиапазон дискретизируется в соот­ветствии с частотой Найквиста (минимальной частотой дискретизации), квантуется, кодируется, мультиплек­сируется и передается. В приемнике поддиапазоны демультиплексируются, декодируются и переводятся обратно в их частотные позиции. Результи­рующие сигналы поддиапазонов затем складываются для получения аппрок­симированного исходного речевого сиг­нала.

# IP-телефония

Основные стандарты кодирования речи, применяемые в 1Р-телефонии, приведены в табл. 4.1.

Таблица 4.1

Стандарты ITU-T по кодированию речи, применяемые в IP-телефонии

|  |  |
| --- | --- |
| Стандарт | Описание |
| G.711 | Импульсно-кодовая модуляция 64 кбит/с (ИКМ) (*А*-закон и μ-закон) |
| G.722 | Широкополосные кодеры, работающие на скорости 64, 56 или 48 кбит/с |
| G.726 | Рекомендации по кодерам АДИКМ, которые охватывают G721 и G723 |
| G.727 | АДИКМ, работающие на скоростях 40, 32, 24 или 16 кбит/с |
| G.728 | Вокодеры с линейным предсказанием, с кодовым возбуждением, с низкой задержкой, скорость 16 кбит/с (LD-CELP) |
| G.729 | Вокодеры с линейным предсказанием, с алгебраическим кодовым возбуждени­ем, с сопряженной структурой, скорость 8 кбит/с (CS-ACELP) |
| G.723.1 | Низкоскоростные вокодеры для связей мультимедиа, работающие на скорости 6,3 и 5,3 кбит/с |

Каждая из при­веденных в таблице рекомендаций ITU может служить основой для передачи речи по Интернету и другим сетям, так как все они обеспечивают низкие ско­рости передачи и достаточно просты в реализации персональным компьюте­ром или в микропроцессорном исполне­нии.

Основной целью проектирования кодеров является уменьшение скоро­сти передачи речи при безусловном сохранении требуемого уровня каче­ства речи для конкретного приложе­ния. Приложения по передаче речи в Интернет или Интранет могут быть либо самостоятельными, либо в фор­ме мультимедиа. Так как мультиме­диа подразумевают наличие несколь­ких средств кодирования речи, для та­ких приложений подразумевается, что поток речевых данных передается по линии связи совместно с другими сиг­налами. Некоторые из таких приложе­ний могут включать:

одновременную передачу речи и видео;

приложения с одновременной ци­фровой передачей речи и данных (DSVD);

одновременную передачу речи и факса.

Особенности функционирования каналов для передачи речевых данных и прежде всего сети Интернет, а также возможные варианты построения си­стем телефонной связи на базе Интер­нет предъявляют ряд специфических требований к речевым кодерам (воко­дерам). Благодаря пакетному прин­ципу передачи и коммутации речевых данных отпадает необходимость коди­рования и синхронной передачи оди­наковых по длительности фрагментов речи.

Наиболее целесообразным и есте­ственным для систем IP-телефонии является применение кодеров с пере­менной скоростью кодирования рече­вого сигнала. В основе кодера речи с переменной скоростью лежит класси­фикатор входного сигнала, определя­ющий степень его информативности и, таким образом, задающий метод коди­рования и скорость передачи речевых данных. Наиболее простым классифи­катором речевого сигнала является де­тектор активности речи (VAD — Voice Activity Detector), который выделяет во входном речевом сигнале активную речь и паузы. При этом фрагменты сиг­нала, классифицируемые как актив­ная речь, кодируются каким-либо из известных алгоритмов (как правило, методом CELP) с типичной скоростью 4…8 кбит/с. Фрагменты, классифи­цированные как паузы, кодируются и передаются с очень низкой скоростью (порядка 0,1.. .0,2 кбит/с) или не пере­даются вообще. Передача минималь­ной информации о паузных фрагмен­тах предпочтительна.

С помощью более эффективных классификаторов входного сигнала мо­жет более детально осуществляться классификация фрагментов, соответ­ствующих активной речи. Это позво­ляет оптимизировать выбор стратегии кодирования (скорости передачи дан­ных), выделяя для особо ответствен­ных за качество речи участков рече­вого сигнала большее число бит (соответственно большую скорость), для менее ответственных – меньше бит (меньшую скорость). В результате мо­гут быть достигнуты еще более низкие средние скорости (2...4 кбит/с) при высоком качестве синтезируемой речи.

Передатчик состоит из кодера речи, VAD, усреднителя фоновых шумов и переключателя на канал, который управляется выходом VAD. Когда на вход есть речь, передатчик постоянно включен. Во время пауз передатчик выключается, но после определенного времени, которое должно быть достаточно коротким, передатчик снова включается на один фрейм, чтобы передать информацию о среднем фоне для точного генерирования в приемнике комфортного шума. На приемной стороне, если определено наличие речи, происходит нормальный синтез. Если определено наличие паузы, выполняется одно из двух действий. Если не передается новой информации о фоне, используются существующие параметры шума, генерируется комфортный шум и используется для текущего фрейма. Если передаются новые параметры фонового шума, то старые параметры заменяются на вновь декодированные, а за тем генерируется новый комфортный шум. Обычно, на стороне декодера также используется индикатор «хоро­ший/плохой» фрейм, чтобы показать верны или нет декодированные пара­метры, и если нет, используется заме­на фрейма. Эффективность DTX зависит от точности VAD.

## Кодеры стандарта D-AMPS

Цифровой стандарт мобильной ра­диосвязи D-AMPS (Digital Advanced Mobile Phone Service), принятый в США в 1990 г., по своим функциональ­ным возможностям и предоставляе­мым услугам приближается к стандар­ту GSM. Стандарт D-AMPS не принят в европейских странах, за исключени­ем России, где он в основном ориенти­рован на региональное использование.

Блок предварительной обработки выполняет следующие функции:

предварительную цифровую фильтрацию входного сигнала с целью подъема верхних частот, на долю кото­рых в спектре речевого сигнала прихо­дится меньшая мощность;

«нарезание» сигнала на сегменты по 160 выборок (20 мс).

Для каждого 20-мс сегмента оце­ниваются параметры фильтра кратко­временного линейного предсказания – 10 коэффициентов частичной корреля­ции , (порядок предсказа­ния *М* = 10), которые непосредствен­но кодируются для передачи в канал связи без каких-либо дополнительных преобразований, и оценивается ампли­тудный множитель *р*, определяющий энергию сегмента речи.



Сигнал с выхода предваритель­ной обработки фильтруется фильтром-анализатором кратковременного ли­нейного предсказания *A(z)*, имеюще­го форму трансверсального линейного фильтра, для чего коэффициенты ча­стичной корреляции преобразуются в коэффициенты линейного предсказа­ния .



Выходной сигнал фильтра крат­ковременного предсказания (остаток предсказания ) используется для оценки параметров фильтра дол­говременного предсказания – задерж­ки τ и коэффициента предсказания , причем параметры долговременного предсказания оцениваются в отдель­ности для каждого из четырех подсегментов по 40 выборок, на которые раз­деляется сегмент из 160 выборок.



Для каждого из подсегментов определяются параметры сигнала воз­буждения. Для этого в составе коде­ра используется схема, аналогичная входящей в состав декодера, которая включает фильтры-синтезаторы крат­ковременного и долговременного предсказания и две кодовые кни­ги и реализует метод «анализа-через-синтез». Каждая из кодовых книг сиг­нала возбуждения содержит 128 кодо­вых векторов, по 40 элементов в ка­ждом.



Все кодовые векторы одной книги являются элементами 7-мерного линейного подпространства в 40-мерном пространстве. Каждая кодовая книга, содержащая 128 векторов, задается семью базисными векторами и 128 кодовыми словами (7-элементными векторами коэффициентов линейных комбинаций) с однобитовыми элементами.

Сигнал возбуждения фильтр синтезатора кратковременного предсказания, в соответствии со схемой декодера рис. 5.4, является суммой векторов возбуждения из двух кодовых книг и вектора с выхода фильтра синтезатора долговременного предсказания. Векторы возбуждения из кодовых книг до подачи на сумматор умножаются на соответствующие коэффициенты усиления и , а входным сигналом фильтра-синтезатора долговременного предсказания является, в зависимости от участка сегмента, выходной сигнал того же фильтр или суммарный сигнал возбуждения фильтра-синтезатора кратковременно­го предсказания. Параметры сигна­ла возбуждения – номера векторов возбуждения и из первой и вто­рой кодовых книг и соответствующие коэффициенты усиления и – определяются по критерию миниму­ма среднеквадратичной ошибки на вы­ходе фильтра-синтезатора кратковре­менного предсказания, входящего в со­став кодера. Предварительно базис­ные векторы обеих кодовых книг ортогонализируются: для первой книги – по отношению к выходному вектору фильтра-синтезатора долговременно­го предсказания, для второй книги – по отношению к тому же выходному вектору и к базисным векторам первой книги.



В результате выходная информа­ция кодера речи для 20-мс сегмента включает:

• параметры фильтра кратковре­менного линейного предсказания – 10 коэффициентов частичной корреляции , , и амплитудный множи­тель *р* – один набор на весь сегмент;



• параметры фильтра долговре­менного линейного предсказания – ко­эффициент предсказания и задержку τ – для каждого из четырех подсегментов;



• параметры сигнала возбуждения – номера и векторов возбуждения из двух кодовых книг и соответствую­щие коэффициенты усиления и – для каждого из четырех подсегментов.



В табл. 5.2 приведено содержание выходной информации кодера с указа­нием числа бит, используемых для ко­дирования.

Таблица 5.2

Кодирование выходной информации кодера речи стандарта D-AMPS

|  |  |  |
| --- | --- | --- |
| Передаваемые параметры | Число бит | Примечание |
| Параметры кратковременного предсказания (коэффициенты частичной корреляции , ) | 38 | – 6 бит;  – по 5 бит;  – по 4 бита;  – по 3 бита;  – 2 бита |
| Амплитудный множитель (энергия сегмента) *р* | 5 |  |
| Задержка фильтра долговременного предска­зания τ (для каждого из четырех подсегментов) | 28 | 7 бит на каждый подсегмент |
| Номера векторов возбуждения и из двух кодовых книг (для каждого из четырех подсег­ментов | 56 | h и i2 по 7 бит |
| Коэффициенты усиления , и (для каждо­го из четырех подсегментов) | 32 | 8 бит на каждый подсегмент; векторно­му квантованию и кодированию подвер­гаются некоторые функции от , и |
| Всего на 20-мс сегмент | 159 |  |

Общий объем информации, выдаваемой для 20-мс сегмента речи, составляет 159 бит. Поскольку исход­ный объем информации на входе коде­ра составляет 1280 бит (160 выборок по 8 бит), кодер осуществляет сжатие информации более чем в 8 раз. Пе­ред передачей в канал связи выходная информация кодера речи подвергается дополнительному канальному кодиро­ванию, причем разные параметры в за­висимости от их важности для обеспе­чения качества речи кодируются с раз­личной степенью избыточности.

Функционирование декодера осу­ществляется по следующему алгорит­му. Сигнал возбуждения фильтра-син­тезатора кратковременного предсказа­ния формируется таким же образом, как и в синтезирующей схеме кодера:

по номерам и из кодовых книг выбираются векторы возбуждения, ко­торые умножаются соответственно на коэффициенты усиления и и складываются с выходным вектором фильтра-синтезатора долговременного предсказания, определяемого параме­трами и τ.



Окончательно сигнал возбужде­ния фильтруется фильтром-синтеза­тором кратковременного предсказа­ния, выполненного в форме трансверсального фильтра, т.е. параметры фильтра преобразуются из коэффициентов частотной корреляции в коэффициенты предсказания . Для улучшения субъективного качества синтезированной речи выходной сигнал фильтра-синтезатора подвергается цифровой адаптивной постфильтрации и с выхода постфильтра получается восстановленный цифровой речевой сигнал.



## Кодеры TETRA

TETRA (Trans-European Trunked Radio) представляет собой стандарт цифровой транкинговой радиосвязи, состоящий из ряда спецификаций, разработанных Европейским институ­том телекоммуникационных стандар­тов ETSI.

TETRA — открытый стандарт, т.е. доступ к спецификациям TETRA сво­боден для всех заинтересованных сто­рон. В связи с этим оборудование раз­личных производителей должно быть совместимо.

Стандарт TETRA создавался как единый общеевропейский цифровой стандарт. Стандарт разработай на основе технических решений и реко­мендаций стандарта GSM и ориенти­рован на создание систем связи, эф­фективно и экономично поддержива­ющих совместное использование сетей различными группами пользователей с обеспечением секретности и защищен­ности информации.

Речевой кодер TETRA основан на модели кодирования CELP – с ли­нейным предсказанием с кодовым воз­буждением. В этой модели блок из N речевых выборок синтезируется пу­тем фильтрации соответствующей об­новленной последовательности из ко­довой книги, масштабированной коэф­фициентом усиления , с помощью двух изменяющихся во времени филь­тров.



Первый фильтр является филь­тром долгосрочного предсказания (фильтром основного тона), цель ко­торого – моделирование псевдоперио­дического речевого сигнала, а второй – фильтр краткосрочного предсказа­ния – моделирует огибающую речево­го спектра.

Передаточная характеристика долгосрочного фильтра (или фильтра синтеза основного тона) определяется формулой



где Т – задержка основного тона; – коэффициент усиления основного тона. Фильтр синтеза основного тона вы­полнен как адаптивная кодовая книга, где для задержек, меньших чем дли­на подфрейма, повторяется последнее возбуждение.



Краткосрочный фильтр синтеза определяется формулой



где , , – параметры ли­нейного предсказания; *р* – порядок предсказателя. В кодере TETRA по­рядок *р* = 10.



При способе анализа-через-синтез синтезированная речь вычисляется для всех кандидатов – последователь­ностей, составляя особую последова­тельность, которая и формирует вы­ходной сигнал, наиболее близкий к исходному, в соответствии с взвешен­ной величиной измеренных искажений. Фильтр взвешивания, корректирую­щий ошибку предыскажений в области форманты спектра речи, определяется формулой

(5.1)



где – обратный (инверсный) фильтр линейного предсказания; (используется значение ). Для взвешивающего филь­тра и фильтра синтеза формант используются квантованные па­раметры линейного предсказания.



В алгебраическом CELP (ACELP) используется специальная кодовая книга, имеющая алгебраическую структуру. Эта алгебраическая струк­тура имеет некоторые преимущества в отношении сохранения, сложности поиска и устойчивости (робастности). Кодер TETRA использует специаль­ную динамическую алгебраическую кодовую книгу возбуждения, посред­ством которой, а также динамической матрицы формы образуются фиксиро­ванные векторы возбуждения. Матри­ца формы – это функция модели A(z) линейного предсказания. Главная ее роль – формировать векторы возбу­ждения в частотной области так, что­бы их энергии были сконцентрированы в наиболее важных частотных полосах. Используемая матрица формы являет­ся триангулярной Теплицевой матри­цей низшего порядка, сформированной из импульсного отклика фильтра:

(5.2)



где *A(z)* — инверсный фильтр линей­ного предсказания (в конкретных реа­лизациях и ).



В кодере TETRA используются фреймы речи по 30 мс. Это требует­ся для того, чтобы параметры кратко­срочного предсказания вычислялись и передавались в каждом речевом фрей­ме. Речевой фрейм разделен на четы­ре подфрейма по 7,5 мс (60 выборок). Основной тон и параметры алгебраиче­ской кодовой книги также передаются в каждом подфрейме. В табл. 5.3 пред­ставлено распределение бит для коде­ра TETRA. Должно быть сформиро­вано 137 бит для каждого фрейма по 30 мс, что в результате дает скорость 4567 бит/с.

Таблица 5.3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Параметр | Номер сегмента | | | | Всего в кадре |
| 1 | 2 | 3 | 4 |
| Коэффициенты линей­ного предсказания |  |  |  |  | 26 |
| Период основного тона | 8 | 5 | 5 | 5 | 23 |
| Индекс алгебраической кодовой книги | 16 | 16 | 16 | 16 | 64 |
| Коэффициенты уси­ления | 6 | 6 | 6 | 6 | 24 |
| Всего |  |  |  |  | 137 |

## Кодеры стандарта АРСО 25

АРСО 25 – стандарт транкинговой радиосвязи, описывающий струк­туру цифровой транкинговой систе­мы и некоторые ее интерфейсы. Для цифровой передачи речи стандарт АРСО 25 предусматривает использова­ние кодера IMBE (Improved MultiBand Excitation, модифицированный метод многополосного возбуждения). Кодер формирует цифровой поток со скоро­стью 4,4 кбит/с. Для исправления ошибок в цифровом речевом сигнале используется избыточное кодирование, порождающее дополнительный цифро­вой поток со скоростью 2,8 кбит/с.

Цифровой речевой сигнал переда­ется кадрами длительностью 180 мс. Два речевых кадра образует суперкадр длительностью 360 мс. Перед переда­чей речи следует преамбула длитель­ностью 82,5 мс, которая содержит синхропакет (48 бит), идентификатор сети (64 бита), служащий для предотвраще­ния конфликтов между радиостанци­ями, работающими на одной частоте; информацию для алгоритма шифрова­ния, идентификатор ключа алгоритма шифрования и другие служебные иден­тификаторы (всего 126 бит). Кадры ре­чи, кроме собственно речевой инфор­мации, содержат дополнительную ин­формацию (управления связью, кана­ла сигнализации и т.д.)

Речевой IMBE-кодер основан на модели речи, которая относится к моделям с многополосным возбужде­нием (МВЕ). Основная идея работы кодера состоит в разделении цифрово­го речевого входного сигнала на пере­крывающиеся речевые сегменты (или фреймы) с использованием окна Кай­зера. Затем для определенного фрей­ма оценивается набор параметров.

Речевой MBE-кодер является во­кодером, т.е. он не кодирует входной речевой сигнал выборка за выборкой, а синтезирует сигнал, который содержит ту же информацию для восприятия че­ловеком, что и исходный речевой сиг­нал. Заметим, что когда речь не явля­ется вокализованнной, исходный и син­тезированный сегменты речи могут не иметь никакого сходства во временной области.

Речевой MBE-кодер имеет два основных преимущества перед ра­нее используемыми вокодерами: во-первых, он основан на МВЕ рече­вой модели, которая является более устойчивой, чем традиционные речевые модели в рассмотренных вокодерах; во-вторых, данный метод использует более сложный алгоритм оценки параметров модели речевого синтеза речевого сигнала из параметров модели.

Главное отличие речевых традиционных вокодеров от модели МВЕ состоит в сигнале возбуждения. В обычных речевых моделях для каждого речевого сегмента используется единственное решение вокал/невокал. В отличие от этого речевая модель МВЕ разделяет сигнал возбуждения на несколько неперекрывающихся частотных полос и принимает решение вокал/невокал для каждой частотной полосы. Это позволяет представить сигнал возбуждения для определенного речевого сегмента в виде смеси периодической (вокализованной) энергии и шумоподобной (невокализованной) энергии. Из-за этих множественных определений вокал/невокал эта модель называется моделью с многополосным возбуждением. Такая речевая модель позволяет синтезировать речь с более качеством, чем традиционные модели. Кроме того, речевая модель МВЕ более устойчива к фоновому шуму.

В речевой модели MBE сигнал возбуждения формируется из сигнала основного тона (или основной частоты) и решений вокал/невокал. Для вокализованной речи сигнал возбуждения является периодической импульсной последовательностью, в которой расстояние между импульсами определяется периодом основного тона . Для невокализованной речи сигнал возбуждения представляет собой белый шум. Периодический спектр создается из взвешенной периодической последовательности импульсов, которая полностью определяется окном взвешивания и периодом основного тона. Его спектр формируется из взвешенной последовательности случайного шума.



Обычно алгоритмы для оценки параметров возбуждения и алгоритмы для оценки параметров огибающей спектра работают независимо. Эти параметры оцениваются на основе нескольких критериев без ясных оснований, насколько синтезированная речь должна быть близка к исходной. Это может проявиться в том, что синтезированный спектр будет слегка отличаться от исходного.

В речевом IMBE-кодере параметры возбуждения и огибающей спектра оцениваются одновременно так, что синтезированный спектр является самым близким к исходному речевому спектру.

Блок-схема алгоритма анализа показана на рис. 5.7.



Рисунок 5.7

Параметры МВЕ модели речи, которые должны быть оценены для каждого речевого фрейма следующие:

период основного тона (или основная частота);

решение вокал/невокал;

спектральные амплитуды, харак­теризующие огибающую спектра.

В декодере вокализированная и невокализированная компоненты син­тезируются отдельно и на заключи­тельной стадии объединяются для по­лучения полного речевого сигнала. Алгоритмы, которые используются для синтеза вокализированных и невокализированных частей речи, осно­ваны на двух различных способах.

Невокализованная часть речи ге­нерируется из гармоник, которые объ­явлены невокализованными. Для ка­ждого фрейма речи блок случайно­го шума взвешивается и преобразует­ся с помощью быстрого преобразова­ния Фурье. Области спектра, которые соответствуют вокализованным гармо­никам, принимаются равными нулю.

Так как вокализованная речь мо­делируется ее индивидуальными гар­мониками в частотной области, на сто­роне декодера она восстанавливается как совокупный сигнал регулируемых генераторов. Каждой гармонике вокализованной области фрейма поставлен в соответствие генератор, который характеризуется частотой и фазой. Однако из-за того, что вокализованная часть речи не является периодической на интервалах, состоящих нескольких фреймов анализа, отклонения от ожидаемых параметров соседних фреймов могут вызвать скачки по концам фреймов, что приведет к значительному ухудшению качества речи. Для разрешения этой проблемы во время синтеза проверяются параметры текущего и предыдущего фреймов для уверенности, что на границе фреймов происходит плавный переход. Это делается для того, чтобы на границах фреймов вокализированная речь была непрерывной. Для обеспечения непрерывности в начале и конце фрейма речи функция амплитуды линейно интерполируется между значениями оценок для текущего и предыдущего фреймов.

Синтез речи в IMBE-декодере требует информации об основной частоте, решении вокал/невокал, величине спектральных составляющих и фазе вокализованных гармоник. Так как фазы вокализованных гармоник можно предсказать, информация о фазе не передается между кодером и декодером. Основная частота (основой тон) квантуется с половинной точностью выборки во временной области, причем возможный диапазон тона перекрывается восемью битами. Peшение вокал/невокал является двоичным числом и не требует квантования. Об­щее распределение бит для каждого фрейма приведено в табл. 5.4.

Таблица 5.4.

Распределение бит IMBE-кодера в системе АРСО 25

|  |  |
| --- | --- |
| Параметр | Число бит |
| Основная частота | 8 |
| Информация вокал/невокал | b |
| Спектральные амплитуды | 79 – b |
| Синхронизация | 1 |

Число полос, на которые разбива­ется речевой фрейм в частотной обла­сти, зависит от основного тона фрейма, но не превышает 12.

Таким образом, в кодере IMBE фрейм речи имеет длительность 20 мс, содержит 144 бита, из которых 56 ис­пользуются для канального кодирова­ния, 88 – для кодирования параме­тров речевой модели. Кодер работает на скорости 4,4 кбит/с. Скорость пере­дачи в канале – 7,2 кбит/с.

## Кодирование речи в системе INMARSAT-M

Для системы мобильной спутни­ковой связи INMARSAT была выбра­на улучшенная версия речевого коде­ра МВЕ, описанного в предыдущем параграфе как вер­сия стандарта на скорости 6,4 кбит/с для наземных систем подвижной свя­зи. Размер фрейма кодера 20 мс. При этом обеспечивается кодирование ис­точника и канала с помощью 128 бит на каждый фрейм. Из этих 128 бит 45 (2,25 кбит/с) зарезервированы для коррекции ошибок, а оставшиеся 83 би­та (4,15 кбит/с) разделены между раз­личными параметрами речи, как пока­зано в табл. 5.5.

Таблица 5.5

Распределение бит в системе INMARSAT-M

|  |  |
| --- | --- |
| Параметр | Число бит |
| Основная частота | 8 |
| Информация вокал/невокал | b |
| Спектральные амплитуды | 75 – b |

Синтез речи в МВЕ декодере тре­бует информации об основной частоте, решении вокал/невокал, спектраль­ных величинах и фазах вокализиро­ванных гармоник. Так как фазы вока­лизированных гармоник можно пред­сказать, информация о фазе не пере­дается от кодера к декодеру. Основная частота (основной тон) обычно кван­туется с половинной точностью вы­борки во временной области, причем возможный диапазон тона перекры­вается восемью битами. Решение во­кал/невокал является двоичным чи­слом и не требует квантования. Набор спектральных величин требует большей точности и эффективности кван­тования. Общее распределение бит для каждого фрейма следующее:

1) восемь бит для точного кванто­вания тона;

2) *b* бит для решения вокал/не­вокал, где *b* – число полос принятия решения вокал/невокал;

3) оставшиеся биты для квантова­ния гармоник.

Так как число гармоник зависит от основного тона, который изменя­ется, число бит, присваиваемых для решения вокал/невокал, может так­же изменяться. Типичное число вокализированных/невокализированных полос равно 12 (4), что означает, что максимальное число бит для описания вокал/невокал равно 12. Общее чи­сло бит для квантования гармоник определяется формулой



где максимум *b* равен 12; – общее число бит для фрейма.



Рассмотрим способ квантования, ис­пользуемый в системе INMARSAT-M.

Квантование основной часто­ты. Основная частота квантуется при преобразовании ее в эквивалентный период основного тона . Значения периода основного тона обычно ограни­чены диапазоном . В си­стеме МВЕ, предназначенной для ра­боты на скорости 6,4 кбит/с, этот пара­метр равномерно квантуется с исполь­зованием 8 бит при размере шага 0,5, что дает возможность обеспечить поло­винную точность выборки.



Кодирование решений во­кал/невокал. *b* решений вокал/не­вокал являются двоичными числами, поэтому могут быть закодированы с использованием одного бита для ка­ждой полосы решения. Система на скорости передачи 2,4 кбит/с исполь­зует максимум 12 полос, каждая с ши­риной Гармоники вне *b* полос во­кал/невокал приняты невокализированными.



Квантование спектральных величин. Перед квантованием спек­тральных величин для уменьшения их изменения используется этап предска­зания.

Остатки предсказания группиру­ются в шесть последовательных бло­ков, содержащих остаточных выборок каждый. Каждый блок преобразуется по частоте с использованием дискретного косинусоидального преобразования (ДКП) размером .



Постоянные составляющие *L* шести ДКП-блоков (первые коэффициенты) группируются как вектор и квантуются с использованием кодовой книги шестибитового скалярного усиления и десятибитового вектора формы. Коэффициенты более высокого порядка М квантуются с использованием скалярных квантователей, где в процессе присвоения бит распределяются оставшиеся биты в соответствии с важностью для субъективного восприятия каждой величины. Блок-схема общей процесса квантования величин показана на рис. 5.8.



Рисунок 5.8

После правильного восстановления спектральных величин МВЕ-кодер пытается улучшить качество восприятия синтезированной речи с использованием усилителя. Усиление спектральных величин выполняется генерированием набора спектральных ве­сов из принятых параметров текущего фрейма.

В типичном МВЕ-кодере большая часть бит выделена для квантования спектральных величин. В случае си­стемы INMARSAT-M, где кодер источ­ника работает на скорости 4,15 кбит/с при скорости фреймов 50 Гц, только 20 (8+ 12) бит используются для форми­рования информации об основном пе­риоде и информации «вокал/невокал». Остальные биты используются для квантования спектральных величин. Таким образом, для успешного синте­за речи необходимо точно знать основ­ной период. Для покрытия речевого спектра 4 кГц достаточно 12 полос во­кал/невокал. В результате общая ско­рость кодера может быть уменьшена за счет более эффективного квантова­ния величин. В системе INMARSAT-M все спектральные величины, кроме ше­сти, проквантованы с использованием скалярных квантователей. Сокраще­ние скорости можно добиться также за счет векторного квантования всех величин. Однако, так как число спек­тральных величин может изменяться от 9 до более чем 60 в зависимости от основной частоты, векторную кодо­вую книгу, учитывающую эти измене­ниями, создать очень сложно. Поэто­му целесообразно векторное квантова­ние использовать только для основной формы спектра, которая может быть принята независимой от основной ча­стоты.

# Перспективы кодирования речи.

В данной главе будут коротко рассмотрены перспективы использования различных речевых кодеков в сетях связи общего пользования.

Рассмотрим некоторые параметры наиболее распространенных кодеков сетей общего пользования.

Однако, кроме кодеков ИКМ по G.711, на ССОП в настоящее время при­меняются аналого-цифровые преобра­зователи других типов, использующие те или иные способы компрессии речи и поэтому имеющие меньшую скорость передачи цифрового сигнала по срав­нению со стандартным кодеком ИКМ. Основные области применения низкоскоростных кодеков: • системы подвижной связи (в част­ности, цифровые сотовые и транкинговые системы); • аппаратура DCME (в небольшом количестве используется на сети ОАО «Ростелеком»); • абонентские компьютерные и мультимедийные терминалы, аппара­ты IP-телефонии;

• цифровые беспроводные телефоны.

Таблица 6.1. Наиболее распространенные кодеки.

|  |  |  |  |
| --- | --- | --- | --- |
| Кодек | Наименование | Скорость кбит/с | Стандарт |
| Standard PCM | Стандартный ИKM кодек для сетей с коммута­цией каналов | 64 | ITU-T G.711 |
| GSM-FR PRE-LTP | Кодек 1-го поколения GSM с возбуждением регулярной последовательностью импульсов и долговременным предсказанием | 13 | ETSIGSM 06.16 |
| GSM-HR VCELP | Кодек GSM (с «половинной» скоростью) с ли­нейным предсказанием и возбуждением вектор­ной суммой | 5,6 | ETSIGSM 06.20 |
| GSM-EFR ACELP | Кодек 2-го поколения GSM с алгебраическим кодовым возбуждением и линейным предсказа­нием | 12,2 | ETSIGSM 06.60 |
| Videophone ACELP | Речевой кодек для мультимедийной связи с ал­гебраическим кодовым возбуждением и линей­ным предсказанием | 5,3 | ITU-T G.723.1 |
| Videophone MP-MLQ | Речевой кодек для мультимедийной связи с мно­гоимпульсным квантованием по критерию прав­доподобия | 6,3 | ITU-T G.723.1 |
| ADPCM | Кодек АДИКМ (адаптивной дифференциальной ИКМ) | 40, 32, 24, 16 | ITU-T G.726 |
| LD-CELP | Кодек с линейным предсказанием, с кодовым возбуждением и малой задержкой | 16, 12, 8, 9, 6 | ITU-T  G.728 |
| CS-ACELP | Кодек с линейным предсказанием, алгебраи­ческим кодовым возбуждением и сопряженной структурой | 8 | ITU-T  G.729 |

Низкоскоростным кодекам свойственны определенные ухудшения параметров, влияющие на качество передачи речи, по сравнению со стандартным кодеком ИКМ. Важно, что эти ухудшения накапливаются при тандемном включении как однородных, так и разнородных низкоскоростных кодеков.

Следует отметить следующие основные факторы, влияющие на каче­ство передачи речи при использовании кодеков:

• искажения квантования;

• временная задержка;

• амплитудно-частотные искаже­ния;

• битовые ошибки;

• проскальзывания;

• потеря кадров;

• потеря пакетов.

Планирование речевых соедине­ний требует обязательного учета ухуд­шений, вносимых каждым переходом А-Ц и Ц-А, и определения на этой ос­нове допустимого количества таких пе­реходов. Для этой цели используется так называемая Е-модель, разработанная ETSI и рекомендуемая МСЭ-Т при планировании речевых соединений «из конца в конец». Эта модель позволяет в комплексе учесть практически все ухуд­шающие факторы.

Наиболее важным параметром Е-модели является коэффициент ухудшения за счет аппа­ратуры, обозначаемый Ic. Чем больше этот коэффициент, тем большую долю деградации вносит данная аппаратура (конкретно – кодек). В табл. 6.2 для срав­нения показаны значения для различ­ных кодеков.

Таблица 6.2

|  |  |  |  |
| --- | --- | --- | --- |
| Кодек | Скорость кбит/с | Стандарт | Ic (G.113) |
| Standard PCM | 64 | ITU-T G.711 | 0 |
| GSM-FR PRE-LTP | 13 | ETSIGSM 06.16 | 20  (25-42)\* |
| GSM-HR VCELP | 5,6 | ETSIGSM 06.20 | 23  (32-45)\* |
| GSM-EFR ACELP | 12,2 | ETSIGSM 06.60 | 5  (15-35)\* |
| Videophone ACELP | 5,3 | ITU-T G.723.1 | 19 |
| Videophone MP-MLQ | 6,3 | ITU-T G.723.1 | 15 |
| ADPCM | 40, 32, 24, 16 | ITU-T G.726 | 2, 7, 25, 50 |
| LD-CELP | 16, 12, 8, 9, 6 | ITU-T  G.728 | 7, 20 |
| CS-ACELP | 8 | ITU-T  G.729 | 10 |
| \* – при наличии битовых ошибок | | | |

Проблема задержки сигнала. Среди многих факторов, влияющих на каче­ство передачи речи, можно отметить задержку сигнала в терминалах и узлах сети.

Желательной является задержка, не превышающая 150 мс, поскольку кроме задержки следует учитывать и другие ухудшающие факторы. Как уже указы­валось, задержка, вносимая стандарт­ными кодеками ИКМ, незначительна и составляет меньше 0,4 мс.

Существенное увеличение задержки по сравнению со стандартными кодека­ми ИКМ дают низкоскоростные кодеки. В частности, только сами кодеки в тер­миналах GSM вносят задержку в 60 мс, что эквивалентно времени прохождения через волоконно-оптическую линию связи (ВОЛС) длиной 12000 км. Проце­дура «фрейминга» (формирования кад­ров) на радиоинтерфейсе добавляет еще 35 мс задержки.

В табл. 6.3 представлены задержки, вносимые речевыми кодеками (МСЭ-Т G.114) различных типов, и соответст­вующие эквивалентные длины ВОЛС. Из таблицы следует, что задержки в низкоскоростных кодеках весьма велики, что делает дополнительные перекоди­ровки практически недопустимыми. При этом нельзя упускать из виду по­вышенные задержки в таких сетевых элементах как шлюзы, маршрутизато­ры и т. д.

Таблица 6.3

|  |  |  |  |
| --- | --- | --- | --- |
| Кодек | Стандарт | Средняя задержка, мс | Эквивалентная длина ВОЛС, км |
| Standard PCM | ITU-T G.711 | 0,375 | 75 |
| GSM-FR RPE-LTR | ETSI GSM 06.10 | 95 | 19000 |
| GSM-HR VCELP | ETSI GSM 06.20 | 95 | 19000 |
| GSM-EFRACELP | ETSI GSM 06.60 | 95 | 19000 |
| Videophone ACE LP | ITU-TG.723.1 | 97,5 | 19500 |
| Videophone MP-MLQ | ITU-T G.723.1 | 97,5 | 19500 |
| ADPCM | ITU-T G.726 | 0,375 | 75 |
| LD-CELP | ITU-T G.728 | 1.875 | 375 |
| CS-ACELP | ITU-T G.729 | 35 | 7000 |

Таким образом, приходится делать выбор между качеством связи и шириной канала поэтому задача проектирования наиболее качественных кодеков для низкоскоростных каналов с высоким уровнем помех весьма актуальна.

Рассмотрим наиболее эффективные, сегодня методы практической реализации низкоскоростных (1,2—2,4 кбит/с) MELP-вокодеров. предназначенных для работы в канале с высоким процентом канальных ошибок, и основные направления совершенствования данных методов.

В качестве базового алгоритма вокодера был выбран алгоритм MELP–2400, разработанный фирмой Texas Instruments и выигравший открытый конкурс по замене кодека федерального стандарт США FS-1016. Данный алгоритм основан на традиционной параметрической модели кодирования с линейным предсказанием и, кроме того, содержит ряд дополнительных особенностей:

• вся рабочая область частот делится на пять полос; в каждой из которых принимается решение о классе сигнала возбуждения — «шумовой» или «голосовой». Таким образом суммарный сигнал возбуждения является смешанным;

• форма «голосового» сигнала возбуждения реконструи­руется в декодере с помощью амплитуд коэффициентов Фу­рье, вычисленных в анализирующей части вокодера:

• для реализации одиночных импульсов возбуждения применяются «апериодические» импульсы;

• с целью улучшения «натуральности» звучания синте­зированной речи применяются дисперсионный и адаптивный фильтры.

Суммарный список параметров, передаваемых от коле­ра к декодеру за один речевой фрейм длительностью 22.5 мс, представлен в табл. 6.4.

Таблица 6.4.

|  |  |  |
| --- | --- | --- |
| Параметр | «Голосовой»  фрейм (бит) | «Шумовой»  фрейм (бит) |
| Линейные спектральные пары | 25 | 25 |
| Амплитуды преобразования Фурье сигнала возбуждения | 8 | – |
| Коэффициенты усиления (2 за фрейм) | 8 | 8 |
| Период основного тона, общая озвученность фрейма | 7 | 7 |
| Озвученность по полосам | 4 | – |
| Флаг периодичности | 1 | – |
| Защита от ошибок | – | 13 |
| Синхробит | 1 | 1 |
| Всего за кадр, 22.5 мс | 54 | 54 |

Выбранный вокодер обеспечивает хорошее звучание синтезированной речи при скорости битового потока 2,4 кбит/с и величине битовых ошибок не более.

Задача практической реализации низкоскоростного воко­дера, имевшего приемлемое качество синтезированной речи при ошибках в канале до 5%, решалась в два этапа. Первоначально была снижена скорость битового речевого потока до 1,2 кбит/с без существенной деградации качества выходного речевого сигнала. Далее были выбраны оптимальные в смысле качества синтезированной речи коды, исправляющие ошибки, и разработан метод замены выбитых речевых параметров на интерполированные значения параметров предшествующих фреймов. После этого суммарная скорость потока информационных и проверочных бит составила 2,4 кбит/с.

Вокодер с битовой скоростью 1,2 кбит/с. При понижении скорости битового потока основное внимание уделялось сохранению разборчивости речи. Анализ алгоритма MELP–2400 показал, что амплитуды преобразования Фурье сигнала возбуждения отвечают в основном за узнаваемость диктора и передачу интонации. Поэтому в спроектированном вокодере MELP–1200 они не используются. Передача коэффициента усиления один раз за фрейм также не приводит к существенному ухудшению разборчивости. Для дальнейшего понижения битовой скорости была разработана техника передачи линейных спектральных пар (ЛСП) только по четным фреймам с помощью адаптивного выбора фреймов на основе решетчатой структуры с критичными изменениями. Для нечетных фреймов вектор ЛСП находился с помощью интерполяционных методов. Также для уменьшения числа бит была использована частота границы голосовой активности с двухбитовой кодировкой.

Суммарный список параметров алгоритма MELP-1200, передаваемых от кодера к декодеру за один речевой фрейм длительностью 45 мс, представлен в табл. 6.5.

Таблица 6.5

|  |  |  |
| --- | --- | --- |
| Параметр | «Голосовой»  фрейм (бит) | «Шумовой»  фрейм (бит) |
| Линейные спектральные пары | 25 | 25 |
| Амплитуды преобразования Фурье сигнала возбуждения | – | – |
| Коэффициенты усиления (2 за фрейм) | 5+5 | 5+5 |
| Период основного тона, общая озвученность фрейма | 7+7 | 7+7 |
| Озвученность по полосам | 2+2 | – |
| Флаг периодичности | 1 | 1 |
| Защита от ошибок | – | – |
| Синхробит | – | – |
| Всего за кадр, 22.5 мс | 54 | 50 |

Вокодер с битовой скоростью 2,4 кбит/с для работы в каналах с величиной ошибок до 5% (MELP–C–2400). Для исправления ошибок битового потока было исследовано несколько классов кодов. Наиболее оптимальными в смысле качества восстановленной речи оказались коды Хэмминга (8,4), позволяющие исправлять одиночную ошибку в четырех информационных битах и определять наличие ошибки в 2 бита, а также коды Голея (23,12), исправляющие три ошибки в двенадцати информационных битах. При проектировании был сделан выбор в пользу кодов Хэмминга. В результате к 54 информационным битам добавлялись 54 проверочных, и общая скорость битового потока составила 2,4 кбит/с.

Предлагаемый алгоритм по­строения низкоскоростного вокодера MELP—С-2400 был про­моделирован на персональном компьютере с использованием языка программирования «СИ» при представлении данных в формате с фиксированной запятой. Для определения качества синтезированной речи использовалась диагностическая мера приемлемости (Diagnostic Acceptability Measure — DAM). Для этой цели были использованы шесть wav-файлов с муж­скими голосами и шесть wav-файлов с женскими голосами, которые затем прослушивались на выходе тракта 30 слуша­телями, после чего определялась средняя оценка мнений — MOS. Такой метод называется методом субъективной оцен­ки качества.

Таким образом, на основании полученных ре­зультатов можно сделать следующие выводы:

1. Вокодер MELP-I200 имеет разборчивость синтезиро­ванной речи близкую к вокодеру MELP-2400 и может быть ис­пользован в каналах с пропускной способностью 1,2 кбит/с,

2. При битовых ошибках в канале более 1% вокодер MELP—С-2400 существенно улучшает качество синтезиро­ванной речи.

Что касается улучшения качества корректирующих ко­дов, то наиболее целесообразным представляется применение сверточных кодов со скоростью 1/3. Как показывают исследо­вания, в этом случае при использовании схемы декодера Виттерби с мягким решением приемлемое качество речи можно обеспечивать при вероятности канальной ошибки до 10%. В каналах с памятью эффективным может оказаться приме­нение каскадных кодов (например, последовательное коди­рование кодами Рида-Соломона или сверточными кодами) или турбокодов.

Дальнейшее понижение скорости битового потока во­кодера возможно при использовании модифицированных кодовых книг, отражающих возможные сочетания ЛСП, и суммарного векторного кодирования речевых параметров. При этом ожидаемая скорость речепреобразующих устройств может не превысить 800 бит/с при словесной разборчивости речи не менее 90%.

Далее рассмотрим кодеки и протоколы, непосредственно используемые в IP-телефонии. Эта тема очень важна, так как есть концепции по переводу в режим IP-телефонии всех сотовых и стационарных сетей.

Задача управления установлением соединений. Общий прин­цип действия Интернет-телефонии состоит в передаче аналогового речевого сигнала от телефонного аппарата или учреж­денческой АТС в маршрутизатор для сжатия и преобразова­ния речевого сигнала в пакеты данных. Эти пакеты передаются по сети Интернет в удаленный маршрутизатор. Последний преобразует пакеты данных обратно в речевой сигнал, кото­рый и передается в телефонный аппарат или на УАТС. В соответствии с такой схемой, описанной, в частности, в рекомендациях МСЭ Н.323/Н.248, требуется выполнение до­статочно сложных преобразований между техникой Интернет-телефонии и коммутируемой телефонной сетью общего по­льзование. Одним из путей преодоления возникших труднос­тей, предпринятых группой инженерной поддержки Интернета IETF (Internet Engineering Task Force), стада разработка протокола запуска соединения SIP (Session Initiation Protocol). Он применим для интегрированной среды Интернета и коммутируемой телефонной сети общего пользования.

SIP относится к протоколам прикладного уровня семиу­ровневой эталонной модели ВОС МОС как структура про­токола HTTP типа "клиент-сервер". При обработке пакетов этого протокола команды и состояния могут передаваться в виде чистого текста посредством считывания данных пакетов HTTP. Поэтому протокол SIP очень подходит для архитектуры передачи по широкомасштабной информационно-вычи­слительной сети. В его структуре должен быть сформирован, по крайней мере, один сервер соединений SIP в дополнение к агентам пользователей. Сервер соединений SIP может рабо­тать как сервер-представитель (proxy-server), сервер изменения направления (redirect server), сервер регистрации (registry server), сервер речевой почты (voice mail server), и др. Сервер соединений функционально является интегрированным про­граммным обеспечением и может быть соединен с существую­щей коммутируемой сетью общего пользования, сетью Ин­тернет-телефонии и т.п.

Протоколы Н.323 и SIP во многом схожи. Оба они поддерживают как двухстороннюю, так и многостороннюю связь. Протоколы обеспечивают возможность передачи муль­тимедийных данных по протоколу реального времени RTP (Real Time Protocol) и родственному ему управляющему транспортному протоколу реального времени RTCP (Real­time Transport Control Protocol). Последний выполняет функ­ции поддержки обратной связи, синхронизации, обеспечения пользовательского интерфейса, но не управляет передачей данных. Вместе с тем протоколы Н.323 и SIP резко разли­чаются концепциями и основополагающими принципами. Н.323 довольно тяжеловесен. Его описание занимает 1400 страниц и содержит целый стек протоколов, точно регламен­тирующих все процедуры. Это упрощает взаимодействие существующих сетей, но вызывает затруднения при адаптации новых применений. Что же касается протокола SIP. то это типичный Интернет-протокол, работа которого основана на обмене короткими тестовыми строками. Его описание зани­мает гораздо меньший объем (250 страниц), и он хорошо взаимодействует с другими протоколами.

Строго говоря, протокол SIP тоже определяет многоуров­невый стек, который включает сетевой уровень, транспортный уровень, уровень транзакций, необязательный уровень диало­га и собственно прикладной уровень, или уровень услуги. Однако эти уровни достаточно просты по сравнению с иерархией протокола Н.323. Так, сетевой уровень протокола SIP обеспечивает связь и взаимодействие с соответствующим протоколом сети Интернет (IP), чтобы каждый элемент протокола SIP связывался по Интернету. Транспортный уровень протокола SIP определяет, как клиент посылает запросы и принимает ответы и как сервер принимает запросы и посылает ответы по сети. Сервер протокола SIP предста­вляет собой сетевой элемент, принимающий запросы, чтобы обслужить их и посылает обратно ответы на эти запросы. Транспортный уровень ответствен также за формирование сообщений SIP и их передачу по сети. Следующим уровнем протокола S1P является уровень транзакций. Транзакцией называется запрос, посылаемый уровнем транзакций клиента (с помощью транспортного уровня) уровню транзакций сервера вместе СО всеми ответами на запросы, посланные уровнем транзакций сервера, обратно клиенту. Уровень транзакций ответствен за согласование последовательности сообщений и за повторную передачу и фильтрацию дубли­рующих сообщений протокола SIP при ненадежности транс­портного уровня. В любой задаче, выполняемой клиентом агента пользователя, применяется последовательность тран­закций. Уровень, находящийся над уровнем транзакций, – это пользователь транзакций или уровень услуг. Иначе говоря, это приложение, запускаемое на самом верхнем уровне стека протокола SIP, который обеспечивает конкретную функцию элемента.

Между уровнем услуг я уровнем транзакций может существовать необязательный уровень диалога. Диалог по протоколу SIP идентифицирует набор соответствующих тран­закций. Например, при стандартном телефонном соединении двумя транзакциями, относящимися к одному диалогу по протоколу SIP, являются установка соединения и разъедине­ние. Уровень диалога ответствен за согласование последова­тельностей транзакций и управление при их неполноте. Элементы протокола SIP посылают запросы и ответы другим элементам SIP в форме сообщений. Эти сообщения содержат обширную информацию, касающуюся таких деталей, как адреса источников, адреса назначения, подробности маршру­тов, указатели соединений, последовательные номера, и другие сведения относительно работы протокола. Формат сообщений SIP обладает значительной гибкостью в том смысле, что информация в заголовке может быть составлена и упорядочена внутри сообщения, и возможна ситуация, когда сообщения логически эквивалентны, тогда как синтаксически они различны. Например, протокол SIP не указывает для многих заголовков порядок, в котором они должны появля­ться в сообщении. Кроме того, заголовки SIP обычно нейтральны, и стеки протокола от разных поставщиков могут строить сообщения различными способами. Однако важно, что все стеки протокола SIP совместимы друг с другом.

Декодирование сообщений SIP выполняется грамматиче­ским анализатором, который является интегральной соста­вляющей стека протокола SIP. Он изучает сообщение и извлекает информацию, относящуюся к конкретному уровню. Вследствие различия путей, по которым может быть получено сообщение SIP, передающее одну и ту же информацию, грамматический анализатор имеет возможность выделять информацию заголовка, данные о параметрах, и т.п., незави­симо от формирования сообщения. Например, грамматиче­ский анализатор может копировать печатные знаки верхнего и нижнего регистра, варианты выделения заголовка, печатные знаки в строке, пробелы, знаки препинания, знаки в таблицах, и т.п. Таким образом, грамматический анализатор обеспечи­вает функции комплексного анализа. Стеки протокола SIP обычно представляются как общие готовые компоненты и реализуют полную спецификацию протокола SIP, делая их пригодными для использования с любым типом элемента SIP.

Каждый тип элементов протокола SIP представляет раз­ные степени функциональной сложности, и количество сооб­щений, обрабатываемых каждым типом элемента, варьирует­ся. Например, компоненты агента пользователя SIP и серверы агента пользователя могут выполнять сложные задачи обра­ботки и связи, например, при установке соединения, хотя число сообщений, обрабатываемых клиентом агента пользо­вателя, обычно сравнительно невелико. Так, сервер агента пользователя может только обрабатывать сообщения в связи с запросами на установление соединений, поступающих от клиента. Если соединение установлено и никаких изменений в его параметрах не произошло, клиент или сервер агента пользователя не будут обрабатывать дальнейшие сообщения до тех пор, пока соединение не завершится. С другой стороны, иные элементы (такие, как серверы изменения направления протокола SIP) выполняют более простые задачи обработки, но принимают значительно большее количество сообщений. Например, сервер изменения направления протокола SIP выполняет сравнительно простую задачу регистрации отобра­жения между адресом источника SIP и адресом IP, которого адрес источника SIP может достичь. Это фундаментальная задача в сети SIP, так как серверы изменения направления используются представителями SIP, чтобы получить IP-адрес аппарата пользователя для маршрутизации соединения. Это важно, в частности, в случаях, когда доступ к сети получают мобильные клиенты, так как обычно адреса IP назначаются динамически и могут часто меняться. Кроме того, для гарантии, что отображаемая информация не устарела, каждый мобильный клиент обычно посылает с частыми интервалами сообщение REGISTER. Из изложенного ясно, что число сообщений, которые должны обрабатываться сервером изме­нения направления, существенно больше, чем число сообще­ний, обрабатываемых клиентом агента пользователя. Напри­мер, серверы текущей регистрации могут обрабатывать от 200 до 1000 операций регистрации в секунду.

Задачи кодирования речевых сигналов. Важным фактором ощущаемого качества кодера является полоса звуковых частот, в которой передается кодируемый сигнал. До настоящего времени большинство кодеков, испо­льзуемых в современной телефонной связи, занимают полосу 300...3400 Гц (так называемая узкополосная речь). Это огра­ничение существует почти 100 лет, причем сами сети полосу частот не ограничивают (ограничение обусловлено характери­стиками применяемых преобразователей). Именно частота 3,4 кГц была принята в качестве верхней граничной для комму­тируемой телефонной сети общего пользования в стандарте цифровой передачи G.711. Хотя большая часть энергии чаше всего содержится в гласных звуках, которые занимают полосу частот ниже 3 кГц, согласные, несущие критическую информацию, часто требуют полосы частот выше 3 кГц. Поэтому узкополосные системы могут ухудшить разборчи­вость, например, звуки "с" и "ф" различаются только за счет частот выше 3 кГц. С другой стороны, увеличение полосы частот сигнала до 50...7000 Гц (так называемой широкой полосы), улучшает разборчивость, что требует от слушателя меньшей концентрации внимания, а следовательно, значите­льно снижает усталость.

Сегодня в эксплуатации много узкополосных речевых кодеров - начиная с G.711, который применяется в КТСОП со скоростью передачи 64 кбит/с, до G.729 (8 кбит/с) и G.723.1 (6,4 и 5,3 кбит/с), которые используются в услугах мультимедиа. Одним из кодеров, разработанных для мобиль­ных сетей и работы с разными скоростями передачи вплоть до 43 кбит/с, является адаптивный многоскоростной кодек AMR (Adaptive Multi-Rate). В настоящее время мало работ по новым узкополосным кодерам - может быть, потому, что уже существуют кодеры для большинства применений. Един­ственная область, в которой появляются новые узкополосные кодеры - это специальные разработки для пакетных систем Интернет-телефонии. В них созданы более робастные к потерям пакетов кодеры, чем предыдущие кодеры, разрабо­танные для сетей с коммутацией каналов. Это кодеры iLBC (предложенные комиссией IETF) и Enhanced G.711J про­изводства Global IP Sound, Steex.

Рассмотрим их основные характеристики и применение.

iLBC (internet Low Bitrate Codec) – это свободный от лицензионных отчислений кодек для голосовой связи через интернет. Кодек предназначен для узкополосных интернет каналов, со скоростью передачи аудио сигнала (человеческой речи) 13.33 кбит/с при длине кадра в 30 мс или 15.20 кбит/с при 20 мс. Кодек iLBC позволяет добиться хорошего качества передачи аудио сигнала даже при некоторых искажениях, которые происходят в связи с потерей или задержкой пакетов.

iLBC описан в стандарте в RFC 3951. Это один из кодеков, который используется в Gizmo Project, Ekiga, OpenWengo, Google Talk, Skype и Yahoo! Messenger.

* Частота дискретизации 8 кГц/16 бит (160 отсчетов для 20-мс кадров, 240 отсчетов для 30-мс кадров)
* Управляемая реакция на потерю пакетов, задержки и джиттер
* Фиксированный битрейт (15.2 кбит/с для 20-мс кадров, 13.33 кбит/с для 30-мс кадров)
* Фиксированный размер кадра (304 бита в кадре для 20-мс кадров, 400 бит в кадре для 30-мс кадров)
* Обеспечивается устойчивость к потерям пакетов на уровне ИКМ со скрытием потерь пакетов, как в ITU-T G.711
* Загрузка процессора на уровне G.729a при более высоком качестве и лучшей реакции на потерю пакетов
* Лицензионная чистота и свобода от лицензионных отчислений
* Коммерческое использование исходного кода, предлагаемого GIPS, требует лицензирования
* Тестирование PSQM при идеальных условиях приводит к усредненной субъективной оценке (MOS) в 4.14 для iLBC (15.2 кбит/с), сравнимой с оценкой 4.45 для G.711 (Мю-закон)

Speex – это свободный кодек для сжатия речевого сигнала, который может использоваться в VoIP приложениях и подкастах. Он не имеет никаких патентных ограничений и лицензирован под последней версией лицензии BSD (без третьей статьи). Speex может быть использован совместно с медиа-контейнером Ogg или передаваться напрямую через UDP/RTP.

Разработчики позиционируют их проект как дополнение к Vorbis, формату сжатия звука общего назначения.

В отличие от многих других кодеков речи, Speex в основном предназначается не для сотовых телефонов, а для использования в Voice over IP (VoIP) и создания файлов со сжатым звуком. Speex оптимизирован для получения высококачественного речевого сигнала при низких битрейтах. Для достижения этой цели кодек использует переменный битрейт и поддерживает разные диапазоны частот: сверхширокий (англ. ultra-wideband, частота дискретизации 32 КГц), широкий (англ. wideband, 16 КГц) и узкий (англ. narrowband, качество телефонной линии, 8 КГц). Направленность на Voice over IP (VoIP) вместо сотовой связи означает, что Speex должен быть устойчив к потерям пакетов данных, но не к повреждению их, так как UDP (протокол неподтверждаемой доставки сообщений) предоставляет информацию лишь двух видов — данные прибыли неповрежденными или же потеряны. Эта особенность определяет выбор для Speex техники кодирования Code Excited Linear Prediction (CELP).

Основные характеристики кодека:

* Свободное и открытое программное обеспечение, не имеет патентных ограничений
* Интеграция широко- и узкополосного канала в одном потоке данных
* Динамическое переключение битрейта и переменный битрейт (англ. Variable bit-rate, VBR)
* Детектор речевой активности (англ. Voice Activity Detection, VAD, интегрирован с VBR)
* Variable complexity
* Опция декодера — интенсивное стерео (англ. Intensity stereo)

Однако наряду с разработками узкополосных речевых кодеков большое внимание специалистов привлекают широ­кополосные речевые кодеки, рассчитанные на полосу частот речевого сигнала 50...7000 Гц. Первые широкополосные кодеры G.722 (48. 56 и 64 кбит, с) были стандартизированы МСЭ в 1988 г. Первоначально предполагалось, что они заменят кодеры G.711, когда ЦСИО получат более широкое распространение. Это кодеры со сложной формой сигнала, которые работают с двумя поддиапазонами и имеют хорошие характеристики, но невысокий коэффициент сжатия. Затем последовала рекомендация G.722.1 (24 и 32 кбит/с) на кодер, широко используемый в настоящее время в терминалах конференц-связи.

В 2000 г. проектом 3GPP был стандартизирован кодер AMR-WB (Adaptive Multi-Rate Wide Band) для применения в мобильных системах третьего поколения. В 2001 г. он был стандартизирован МСЭ в качестве последнего широкополос­ного кодера G.722.2. Кодер работает с разными скоростями передачи от 6,6 до 23,85 кбит/с, но его вычислительная сложность может быть ограничена возможностями реализа­ции. Очень важно, что обе организации (МСЭ и 3GPP) приняли один и тот же кодер, так как это устраняет необходимость перекодирования при работе между провод­ными и беспроводными сервисами. В конечном счете это снизит стоимость и улучшит характеристики передачи "из конца в конец". По проекту 3GPP также реализована версия AMR-WB+ (в 2004 г.). Она позволяет улучшить характери­стики при передаче неречевых сигналов и является обратно совместимым расширением стандарта AMR-WB. Цель версии возможность применения в системах передачи с коммута­цией пакетов также услуг мультимедиа. В Интернете можно найти сведения и о многих других разработках широкополос­ных кодеков. Расширен для широкополосного применения и кодек G.729 путем создания возможности многоскоростной передачи (рекомендация G.729EV 2006 г.).

Обзор развития достижений в области техники кодирования речевых сигналов дает основания считать, что в настоящее время существуют все предпосылки для постепенного повсе­местного перевода телефонной связи на полосу 50-7000 Гц. Этот вывод относится не только к сети Интернет и пакетной передаче, но и к традиционным цифровым телефонным сетям общего пользования. Скорость передачи 64 кбит/с в таких сетях вполне достаточна для существенного повышения качества передачи и разборчивости речи за счет изменения системы кодирования в оконечных пунктах.

Список литературы.

1. Шелухин О.И., Лукьянцев Н.Ф. «Цифровая обработка и передача речи». М., «Радио и связь», 2000.
2. А.М. Меккель. «Влияние переходов «аналог-цифра и «цифра-аналог» при построении сетей на основе перспективных технологий». «Электросвязь».–2008. –№6. –с. 41-48.
3. М.И. Максимов, Н.А. Сидорова, О.В. Чернояров. «Проектирование низкоскоростных речепреобразующих устройств для каналов с высоким процентом ошибок». «Электросвязь». –2008. – №7. –с. 48-50.
4. А.А. Иванов, О.И. Фаерберг, К.Ю. Никашев. «Концепция модернизация сети общего пользования». «Электросвязь». –2008. –№8. –с. 18-23.
5. В.И. Нейман, Д.А. Селезнев. «Интернет-телефония и перспективы ее развития». «Электросвязь». –2008. –№1. –с. 6-9.
6. http://ru.wikipedia.org/wiki/Speex
7. http://ru.wikipedia.org/wiki/ILBC