МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА»

Кафедра «Радиоуправления и связи»

МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

«МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ РЦР»

Направление подготовки
11.03.02 Инфокоммуникационные технологии и системы связи

Направленность (профиль) подготовки
«Программно-аппаратная инженерия в телекоммуникациях "интернет вещей"»

Уровень подготовки
Бакалавриат

Квалификация выпускника – бакалавр

Формы обучения – очная

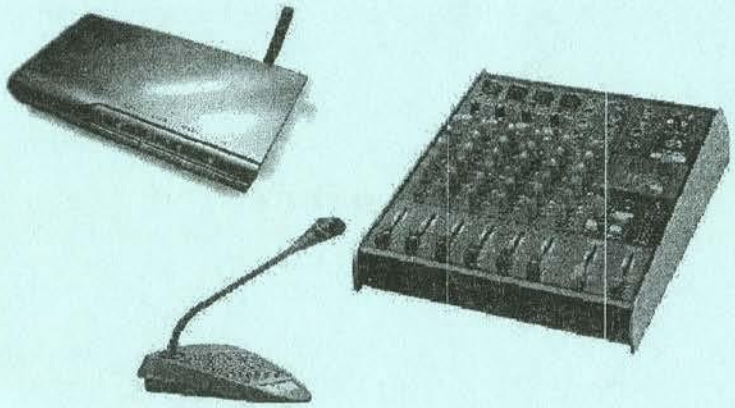
Рязань 2025 г

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

С.Н. КИРИЛЛОВ, В.Т. ДМИТРИЕВ

КОДЕКИ РЕЧЕВЫХ СИГНАЛОВ В МТКС



Рязань 2018

Министерство образования и науки Российской Федерации

Рязанский государственный радиотехнический университет

С.Н. КИРИЛЛОВ, В.Т. ДМИТРИЕВ

КОДЕКИ РЕЧЕВЫХ СИГНАЛОВ В МТКС

Учебное пособие

Рязань 2018

УДК 621.396.43

Кодеки речевых сигналов в МТКС: учеб. пособие/ С.Н. Кириллов, В.Т. Дмитриев; Рязан. гос. радиотехн. ун-т. Рязань, 2018.48 с.

Теоретическая часть содержит основную информацию о кодеках речевого сигнала и включает в себя алгоритмы кодирования. Практическая часть содержит задание для моделирования кодека РС и пример реализации кодека в среде программирования Matlab.

Материал предназначен для курсового проектирования и соответствует программам дисциплин подготовки магистров техники и технологии по направлению 11.04.02 «Инфокоммуникационные технологии и системы связи» и специалистов по специальности 11.05.01 «Радиоэлектронные системы и комплексы».

Табл. 7. Ил. 11. Библиогр.:15 назв.

Кодеки речевых сигналов, алгоритмы кодирования, обработка речевых сигналов, оценка качества речи, моделирование кодека

Печатается по решению редакционно-издательского совета Рязанского государственного радиотехнического университета.

Рецензент: кафедра радиоуправления и связи Рязанского государственного радиотехнического университета (зав. кафедрой д-р техн. наук, проф. С.Н. Кириллов)

Кириллов Сергей Николаевич
Дмитриев Владимир Тимурович

Кодеки речевых сигналов в МТКС

Редактор Р.К. Мангутова
Корректор С.В. Макушина
Подписано в печать 25.07.18. Формат бумаги 60х84 1/16.
Бумага писчая. Печать трафаретная. Усл. печ. л. 3,0.
Тираж 50 экз. Заказ 3512.
Рязанский государственный радиотехнический университет.
390005, Рязань, ул. Гагарина, 59/1.
Редакционно-издательский центр РГРТУ.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Учебное пособие посвящено алгоритмам кодирования формы сигнала, измерению разборчивости речи артикуляционным методом и измерению качества речи методом оценки по селективным признакам. Цель данного учебного пособия – ознакомить магистров с реализацией кодеков РС на заданной скорости передачи в программной среде Matlab или с помощью программы «VOCdemo»,а также провести работу согласно ГОСТ Р 50840-95 и ГОСТ Р 51061-97.

Основной учебной литературой по курсовому проектированию кодеков РС является [1...4]. В задачу курсового проекта входит развитие у студентов навыка научно-исследовательской работы в области кодирования речи.

В рамках проекта студенты должны изучить структуру предложенного алгоритма кодирования, научно-техническую литературу, а также использовать стандарты, справочники и техническую документацию.

В ходе выполнения курсового проекта студент должен:

1. Реализовать выбранный кодек РС на заданной скорости передачи в программной среде Matlab. Для реализации кодеков можно воспользоваться программой «VOCdemo» (инструкция к программе приведена в приложении).
2. Разработать согласно ГОСТ Р 50840-95 и ГОСТ Р 51061-97 методику оценки качества речевых сигналов на выходе кодеков и провести запись РС бригадой дикторов.
3. Произвести эксперимент по подаче на входе кодека сигналов и получению на выходе кодека кодированных РС при различных акустических шумах и шумах и искажениях в канале связи.
4. При воздействии различных шумов выявить корреляцию между объективными и субъективными оценками.

1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ. КОДЕКИ РЕЧЕВЫХ СИГНАЛОВ

1.1 Классификация алгоритмов кодирования РС

Все методы цифрового кодирования речи можно разделить на две категории: кодеры формы сигнала и кодеры источника.

В свою очередь схемы кодирования речи могут быть разделены на три основных класса. Основная задача и функция этих схем кодирования - проанализировать входной сигнал, удалить избыточность и соответствующим образом закодировать информативные части сигнала. Для снижения скорости цифрового потока приходится разрабатывать все более сложные методы устранения избыточности [1, 2].

Наиболее распространенные стратегии кодирования речевых сигналов приведены на рис.1.1.

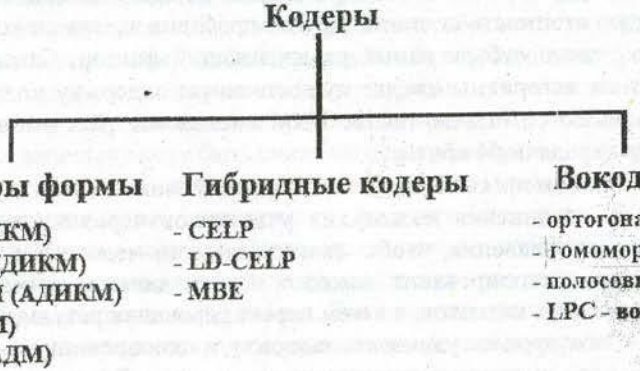


Рисунок 1.1 - Иерархия систем кодирования речевых сигналов

Рассмотрим параметры сравнения методов кодирования речи. К наиболее важным относятся следующие параметры.

Скорость - это диапазон требуемых скоростей передачи. Для систем с более низкой битовой скоростью требуется меньшая полоса частот, по этой причине они обеспечивают более высокую эффективность использования спектра и мощности и в конечном итоге приводят к сотовым системам радиосвязи с повышенной емкостью.

Качество. Критерий, используемый при сравнении, - насколько хорошо звучит речь в идеальных условиях - так называемая чистая речь, без ошибок передачи, только при кодировании. Следует отметить, что качество является субъективным результатом измерения и оценивания.

Пороговая вероятность ошибки на бит. Более высокое значение пороговой ошибки ведет к более робастной структуре системы, а следовательно, к более низким требованиям к отношению сигнал/шум и увеличению емкости сети.

Задержки кодирования системы передачи речи - это фактор, тесно связанный с характеристикой качества системы. Задержка кодирования состоит из алгоритмической задержки (буферизации речи для анализа), вычислительной задержки (времени, необходимого для обработки сохраненных речевых выборок) и задержки в процессе передачи. Для подвижных систем и систем спутниковой связи широко применяется процедура подавления эха, так как присутствует ощутимая задержка распространения. Однако в случае телефонных сетей, где задержка очень мала, при использовании кодера с большими задержками будут требоваться сверхмощные подавители эха, что увеличит общую стоимость системы. Другая проблема задержки кодера (декодера) - чисто субъективный раздражающий фактор. Самые низкоскоростные алгоритмы вводят существенную задержку кодирования по сравнению со стандартными ИКМ системами, рассчитанными на скорость передачи 64 кбит/с.

В приложениях типа телеконференций может возникнуть необходимость соединения нескольких участников через многоточечное устройство управления, чтобы связать каждого человека с другими. Это требует декодирования каждого потока данных, суммирования декодированных сигналов, а затем перекодирования результирующего сигнала. Этот процесс удваивает задержку и одновременно уменьшает качество речи из-за неоднократного кодирования. Такая система телемоста может допустить максимальную задержку в одну сторону 100 мс, так как мост приведет к удвоению задержки системы в одну сторону до 200 мс.

Сложность и потребление энергии. Появление микросхем цифровой обработки сигналов (DSP - Digital Signal Processor) и специализированных БИС (ASIC - Application Specific Integrated Circuits) дало возможность значительно снизить стоимость вычислительных операций. Одной из основных технологий снижения потребления энергии и увеличения эффективности канала является цифровая интерполяция речи (DSI - Digital Speech Interpolation). DSI использует тот факт, что речь в действительности занимает лишь около половины времени переговоров и во время «простоев» канал может быть использован для других целей. Это позволяет снижать активность передатчика, т.е. экономить энергию.

Обработка других сигналов звукового диапазона. Каналы связи, изначально предназначенные для передачи речевых сигналов, нередко используются для передачи другого рода сигналов, таких как сигналы от модема, факсимильного аппарата и др. Статистические параметры формы и частотный спектр таких сигналов совершенно отличны от соответствующих параметров речи, поэтому алгоритм обработки должен оперировать с обоими видами сигналов. При разработке алгоритма возможность преобразования неречевых сигналов звукового диапазона зачастую отходит на второй план и рассматривается на заключительном этапе, что, безусловно, является ошибкой, особенно при разработке кодеров для сетей общего пользования [1, 2].

Качество речи и скорость передачи - два конфликтующих фактора. Чем более низкая скорость речевого кодера, т.е. более высокая степень компрессии сигнала, тем больше страдает качество - в данном случае степень достоверности. Для систем, которые работают в телефонных сетях общего пользования (ТфСОП) и подобных им, основным требованием является качество кодирования (кодеры формы). Однако для закрытых сетей, таких как частные коммерческие или военные системы связи, требования к качеству могут быть снижены с целью повышения скорости. Хотя обычно требуется абсолютное качество, этим пренебрегают для другой группы стандартов, цель которых - добиться более высокой степени сжатия (гибридные кодеры). Например, в системах подвижной радиосвязи высокая степень сжатия при среднем качестве - это решающий фактор при выборе стандарта кодирования. Это качество обычно принимается в расчет как при хороших, так и при плохих условиях передачи. Оказывается, что первый фактор относительно просто измерить по уменьшению скорости цифрового потока. Определить и оценить искажения труднее. Причина состоит в том, что речевой сигнал воспринимает не устройство, а человек. К сожалению, отсутствие моделей восприятия не позволяет разработать объективные методы оценки искажений речи.

1.2 Кодеки формы

Кодек G.711 - рекомендация, утверждённая МККТТ в 1984 г., описывает кодек, использующий ИКМ преобразование аналогового сигнала с точностью 8 бит, тактовой частотой 8 кГц и простейшей компрессией амплитуды сигнала. Скорость потока данных на выходе преобразователя составляет 64 кбит/с. Для снижения шума квантования и улучшения преобразования сигналов с небольшой амплитудой при кодировании используется нелинейное квантование по уровню согласно специальному псевдологарифмическому закону.

Кодек G.721 - это стандартный кодек ITU-T, который использует адаптивную дифференциальную импульсно-кодовую модуляцию (ADPCM). Форма импульсно-кодовой модуляции (PCM) для получения цифрового сигнала с более низкой скоростью передачи по битам, чем стандартный PCM. Этот стандарт ITU для речевых кодеков использует ADPCM на канале со скоростью 32 кбит/с. G.721 был впервые представлен в 1984 году. В 1990 году этот стандарт был сложен в G.726 вместе с G.723.

Кодек G.722 — широкополосный голосовой кодек стандарта ITU-T, работающий со скоростью 48, 56 и 64 кбит/с. Технология кодека основана на АДИКМ. Этот стандарт был принят в 1988 г. и в настоящее время сильно устарел. G.722.1 — это более новая версия кодека G.722 от 1999 г. Он предназначен для сжатия широкополосного аудиосигнала и базируется на третьем поколении технологии сжатия Siren® от компании Polycom. Этот стандарт обеспечивает широкополосный аудиосигнал, более близкий по качеству к FM радио, чем к обычному телефону. G.722.1 определяет работу кодека на скоростях 24 и 32 кбит/с при ширине полосы пропускания 50 Гц – 7 кГц. Кодек G.722.1 Annex C базируется на патентованной технологии Siren 14® от компании Polycom. Качество аудиосигнала приближено к CD. Этот алгоритм сжатия обеспечивает сверхширокополосный аудиосигнал 14 кГц при скоростях передачи 24, 32 и 48 кбит/с.

Кодек G.723.1. Рекомендация G.723.1 утверждена ITU-T в ноябре 1995 года. Форум IMTC выбрал кодек G.723.1 как базовый для приложений IP- телефонии.

Кодек G.723.1 производит кадры длительностью 30 мс с продолжительностью предварительного анализа 7.5 мс. Предусмотрено два режима работы: 6.3 кбит/с (кадр имеет размер 189 битов, дополненных до 24 байтов) и 5.3 кбит/с (кадр имеет размер 158 битов, дополненных до 20 байтов). Режим работы может меняться динамически от кадра к кадру. Оба режима обязательны для реализации. Оценка MOS составляет 3.9 в режиме 6.3 кбит/с и 3.7 в режиме 5.3 кбит/с.

Кодек специфицирован на основе операций как с плавающей точкой, так и с фиксированной точкой в виде кода на языке С. Реализация кодека на процессоре с фиксированной точкой требует производительности около 16 MIPS.

Кодек G.723.1 имеет детектор речевой активности и обеспечивает генерацию комфортного шума на удаленном конце в период молчания. Эти функции специфицированы в приложении А (Annex A) к рекомендации G.723.1. Параметры фонового шума кодируются очень маленькими кадрами размером 4 байта. Если параметры шума не меняются существенно, передача полностью прекращается.

Кодек G.726. Алгоритм кодирования АДИКМ (рекомендация ITU-TG.726, принятая в 1990 г.). Он обеспечивает кодирование цифрового потока G.711 со скоростью 40, 32, 24 или 16 кбит/с, гарантируя оценки MOS на уровне 4.3 (32 кбит/с), что часто принимается за эталон уровня качества телефонной связи (toll quality). В приложениях IP-телефонии этот кодек практически не используется, так как он не обеспечивает достаточной устойчивости к потерям информации.

Кодек G.728 использует оригинальную технологию с малой задержкой LD-CELP (low delay code excited linear prediction) и гарантирует оценки MOS, аналогичные АДИКМ G.726 при скорости передачи 16 кбит/с. Данный кодек специально разрабатывался как более совершенная замена АДИКМ для оборудования уплотнения телефонных каналов, при этом было необходимо обеспечить очень малую величину задержки (менее 5 мс), чтобы исключить необходимость применения эхокомпенсаторов. Это требование было успешно выполнено учеными Bell Labs в 1992 году: кодер имеет длительность кадра только 0.625 мс. Реально задержка может достигать 2.5 мс, так как декодер должен поддерживать синхронизацию в рамках структуры из четырех кадров.

Недостатком алгоритма является высокая сложность - около 20 MIPS для кодера и 13 MIPS для декодера - и относительно высокая чувствительность к потерям кадров.

Кодек G.729 очень популярен в приложениях передачи речи по сетям Frame Relay. Он использует технологию CS-aCeLP (Conjugate Structure, Algebraic Code Excited Linear Prediction). Кодек использует кадр длительностью 10 мс и обеспечивает скорость передачи 8 кбит/с. Для кодера необходим предварительный анализ сигнала продолжительностью 5 мс.

Существуют два варианта кодека:

- G.729 (одобрен ITU-T в декабре 1996), требующий около 20 MIPS для кодера и 3 MIPS для декодера.
- Упрощенный вариант G.729A (одобрен ITU-T в ноябре 1995), требующий около 10.5 MIPS для реализации кодера и около 2 MIPS для декодера.

В спецификациях G.729 определены алгоритмы VAD, CNG и DTX. В периоды молчания кодер передает 15-битовые кадры с информацией о фоновом шуме, если только шумовая обстановка изменяется.

В таблице 1.1 приведены основные характеристики для разных кодеков.

Таблица 1.1 - Характеристики основных кодеков речевых сигналов

Стандарт	Вокодер	Скорость, кБит/с	Год	Применение	Наименование	
G.711	PCM	64	1977	ТЛФ	Pulse-Code Modulation	Импульсно-кодовая модуляция
G.722	SB-ADPCM	64 56 48	1988	ТЛФ	Sub-band ADPCM	---
G.726	ADPCM	16 24 32 40	1984	ТЛФ	Adaptive Differential Pulse-Code Modulation	Адаптивная дифференциальная импульсно-кодовая модуляция
G.728	LD-CELP	16	1992	ТЛФ	Low-Delay Code Excited Linear Prediction	Линейное прогнозирование, генерируемое кодом с низкой задержкой
G.729, G.729a	CS-ACELP	8 6,4 11,8	1997	ТЛФ	Conjugate Structure, Algebraic Code Excited Linear Prediction	Линейное прогнозирование, генерируемое алгебраическим кодом сопряженной структуры
G.723.1	MP-MLQ	6,3	1996	ТЛФ	MultiPulse Maximum Likelihood Quantization	Многоимпульсное квантование с максимальным правдоподобием
G.723	ACELP	5,3	1996	ТЛФ	Algebraic Code Excited Linear Prediction	Линейное прогнозирование, генерируемое алгебраическим кодом
ETSI GSM	RPE-LTP	13	1992	GSM	Regular Pulse Excition - Long Term Prediction	---
ETSI TETRA	ACELP	4,8	1996	GSM	Algebraic Code Excited Linear Prediction	Линейное прогнозирование, генерируемое алгебраическим кодом
США	MELP	2,4	1998	--	---	---

1.3 Алгоритмы кодирования формы сигнала

Импульсно-кодовая модуляция ИКМ (PCM - Pulse Code Modulation). Цифровое преобразование непрерывного речевого сообщения в соответствии с рекомендацией G.711 (рис.1.2) используется наиболее часто.

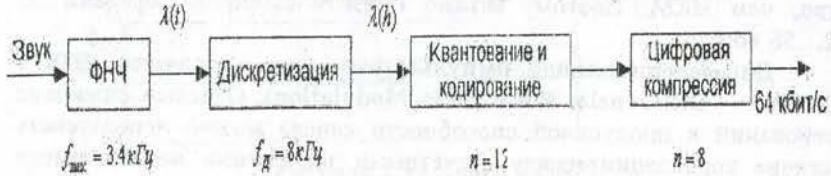


Рисунок 1.2 - Цифровое преобразование непрерывного речевого сообщения в код

При этом максимальная частота сигнала f_{max} = 3,4 кГц, частота дискретизации $f_{диск}$ = 3,4 кГц. После равномерного квантования при числе уровней L=12 и предварительного кодирования производится цифровая компрессия, в результате чего длина кодовой комбинации уменьшается до n=8 разрядов. Результатом преобразования является двоичная последовательность, передаваемая со скоростью 64 кбит/с.

Блочная ИКМ (БИКМ). Из различных систем адаптивной ИКМ (АИКМ) наибольшее распространение получила система блочной ИКМ (БИКМ), которую 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 разряд (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 $R = f_A + \left(k + \frac{m}{N}\right)$. На практике, как правило, используют следующие параметры: $f_{дискр} = 8$ кГц; $n = 10 \ldots 13$; $K = 6 \ldots 8$; $N = 8 \ldots 16$; $m = 6 \ldots 8$.

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). С целью снижения требований к пропускной способности канала можно использовать наличие корреляции между отсчетными значениями передаваемого сообщения. Такой метод называется передачей с предсказанием [5…7]. При этом последовательность значений $\lambda(h)$ поступает на один вход вычитающего устройства (рис.1.3), в то время как на другой вход поступает предсказанное значение $\tilde{\lambda}(h)$, 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 (рис.1.3)[7].

Рисунок 1.3 - Метод предсказания при передаче

На приемном конце значения сообщения $\lambda(h)$ восстанавливаются путем добавления принятого сигнала ошибки предсказания $\varepsilon[h]$ к предсказываемому значению $\tilde{\lambda}(h)$ (рис.1.4).

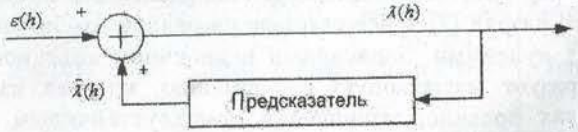


Рисунок 1.4 - Метод предсказания при приеме

В системе с дифференциальной импульсно-кодовой модуляцией (ДИКМ) отсчетные значения $\varepsilon[h]$ ошибки предсказания подвергаются квантованию с переходом к значениям $\varepsilon_q[h]$ 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.1.5 и рис.1.6.

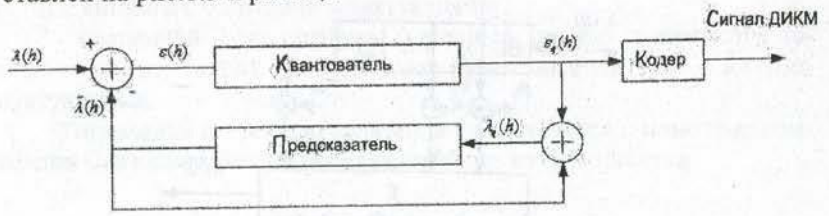


Рисунок 1.5 - Кодер ДИКМ

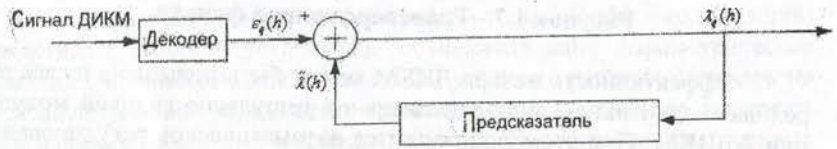


Рисунок 1.6 - Декодер ДИКМ

При этом имеют место соотношения:

$$\varepsilon[h] = \lambda[h] - \tilde{\lambda}(h); \quad (1.1)$$

$$\lambda_q[h] = \tilde{\lambda}(h) + \varepsilon_q[h]. \quad (1.2)$$

В качестве предсказываемого значения сообщения $\tilde{\lambda}(h)$ в простейшем случае может быть использовано предыдущее отсчетное значение, хотя в общем случае используется выражение

$$\tilde{\lambda}(h) = \sum_{i=1}^{M} c_i \lambda_q[h-i], \quad (1.3)$$

так что предсказатель может быть реализован в виде трансверсального фильтра на основе M - отводной линии задержки (регистра сдвига) с временем задержки между отводами, равным интервалу временной дискретизации Δt (рис.1.7).

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.

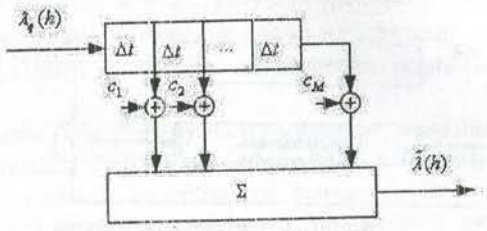


Рисунок 1.7 - Трансверсальный фильтр

Эффективность метода ДИКМ может быть повышена путем перехода к адаптивной дифференциальной импульсно-кодовой модуляции АДИКМ. При этом производится автоматическое регулирование величины шага квантования сигнала ошибки предсказания, а также автоматическая подстройка коэффициентов c_i трансверсального фильтра устройства предсказания (рис.1.7) в соответствии с изменением текущего спектра передаваемого сообщения. Для этого как в передающее, так и в приемные устройства вводятся дополнительные цепи автоматической регулировки усиления, подстройки параметров предсказателя на основе статистического оценивания параметров передаваемого сообщения. За счет использования АДИКМ достигается уменьшение скорости результирующего цифрового потока практически без снижения качества звука.

ADPCM - один из наиболее общепринятых и давно используемых алгоритмов сжатия речи, который регламентируется стандартом G.726, был принят в 1984 г. [1,3…5,7]. Он замещает собой другие стандарты —G.721, который описывает ADPCM передачу голоса полосой в 32 кбит/с, и G.723, который описывает ADPCM передачу в 24 и 40 кбит/с. Четыре полосы кодека G.726 соотносят обычно с размерами выборок (отсчетов) в битах, это 2-, 3-, 4- и 5-битовый соответственно.

Этот алгоритм дает практически такое же качество воспроизведения речи, как и PCM, 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 [5…7].

Основной принцип, реализуемый при масштабировании, заключается в бимодальной адаптации [2… 8]:

- 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.

1.4 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.

По способу анализа и синтеза речи вокодеры можно разделить на два класса: речеэлементные и параметрические [2, 6…8].

В речеэлементных вокодерах при кодировании распознаются произносимые элементы речи (например, фонема) и на выход кодера подаются только их номера. В декодере эти элементы создаются по правилам речеобразования или берутся из памяти декодера. Фонемные вокодеры предназначены для получения предельной компрессии речевых сигналов. Область применения фонемных вокодеров - линии командной связи, управление и говорящие автоматы информационно-справочной службы. В таких вокодерах происходит автоматическое распознавание

слуховых образов, а не определение параметров речи и соответственно теряются все индивидуальные особенности диктора.

Параметрический вокодер представляет собой устройство, которое совершает так называемое параметрическое компандирование речевых сигналов. Компрессия речевых сигналов в кодере осуществляется в анализаторе, который выделяет с речевого сигнала медленно меняющиеся параметры. В декодере при помощи местных источников сигналов, которые управляются принятыми параметрами, синтезируется речевой сигнал. В параметрических вокодерах с речевого сигнала выделяют два типа параметров и по этим параметрам в декодере синтезируют речь:

- параметры, которые характеризуют источник речевых колебаний (генераторную функцию) - частота основного тона, ее изменение во времени, моменты появления и исчезновения основного тона (согласованные или гортанные звуки), шумового сигнала (шипящие и свистящие звуки);
- параметры, которые характеризуют огибающую спектра речевого сигнала.

В декодере соответственно по заданным параметрам генерируются основной тон, шум, а затем пропускаются через гребенку полосовых фильтров для восстановления огибающей спектра речевого сигнала.

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 [5…9]:

- 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 $F_{ср}$. Таким образом, сигналы на выходе

каждого канала изменяются с частотой менее $F_{ср}$. Их передача возможна в аналоговом или цифровом виде [5… 10].

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.

Характеристики вокодеров

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 [2]. 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.

Производительность алгоритма. Вокодеры частот выполняются на основе цифровых сигнальных процессоров (ЦСП). В соответствии с компьютерной терминологией их производительность может быть измерена в миллионах 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.

1.5 Помехоустойчивость

Рассмотрим методы цифрового представления речи, к которым относятся:

- прямое аналого-цифровое преобразование (или импульсно-кодовая модуляция, ИКМ);
- эффективное кодирование речи, ЭКР (здесь можно выделить кодеры формы, вокодеры и кодеры, реализующие алгоритмы анализа через синтез).

Кроме указанных существуют кодеры с многополосным кодированием, с ортогональным преобразованием и с выявлением избыточности предсказания.

Прямое аналого-цифровое преобразование является низкоэффективным (т. е. имеющим малую скорость кодирования при заданном качестве) высококачественным методом кодирования. Кодеки, построенные на базе данного метода, работают на скоростях не ниже 32 кбит/с. При этом полоса входного аналогового сигнала ограничена диапазоном 0,3 - 3,4 кГц. Для повышения качества преобразования полоса может быть расширена до 6 кГц, что соответствует скорости передачи 88 кбит/с при частоте дискретизации 12 кГц (при дальнейшем расширении полосы качество представления речи не повышается).

На рис.1.8, где приведены обобщенные кривые, характеризующие помехоустойчивость различных методов цифрового представления речи, кривая 1 соответствует ИКМ-представлению. Здесь Рош - вероятность ошибки на символ, а SNR - отношение сигнал/шум, рассчитанное через среднеквадратическую ошибку восстановления. ИКМ-кодеки имеют наихудшие показатели помехоустойчивости.

На рисунке можно увидеть, что из всех кривых (характеристик разных способов цифрового представления речи) самый короткий относительно других типов ЦПР участок А (нечувствительность к ошибкам в канале) имеет кривая 1. Кодеки могут потерять работоспособность,

даже если вероятность ошибки равна 10^{-5}, что соответствует параметрам канала среднего класса.

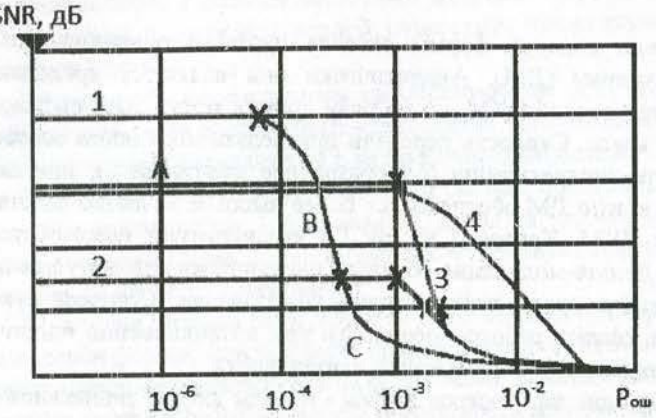


Рисунок 1.8 - Обобщенные кривые, характеризующие помехоустойчивость различных методов цифрового представления речи:
А - область нечувствительности к ошибкам; В - слабая чувствительность; С - потеря работоспособности

Системы с ИКМ работают только в области нечувствительности к ошибкам в канале, но даже в этом случае вводятся специальные меры для устранения последствий возникновения одиночных ошибок. При использовании алгоритма ИКМ со скоростью передачи 64 кбит/с кодек имеет максимальную область нечувствительности к ошибкам в канале при высоком качестве восстановления. Поэтому данный алгоритм рекомендован для большинства систем цифровой передачи речи в качестве метода предварительного аналого-цифрового преобразования.

Другое направление цифрового представления речи, **эффективное кодирование,** иногда, называют сжатием речи.

Раньше, чем все остальные способы, для эффективного цифрового преобразования речи были разработаны вокодеры. Основываясь на выбранной модели речеобразования, вокодер с помощью алгоритма передачи анализирует параметры речевого сигнала, который поступает по каналу связи в приемник; приемный алгоритм позволяет проводить синтез сигнала. Осциллограммы исходного и синтезированного сигнала не совпадают, и речь может носить "искусственный" характер.

Значительные результаты в области эффективного кодирования речи достигнуты на базе общего подхода "кодирования с предсказани-

ем". Большая часть стандартизированных Международным союзом электросвязи алгоритмов кодирования относится именно к этому направлению.

Среди кодеров формы сигнала первыми появились методы дельта-модуляции (ДМ). Аналитически они являются предельными случаями разностной ИКМ, но по ряду причин могут быть выделены в отдельный класс. Скорость передачи при дельта-модуляции соответствует частоте дискретизации (одноразрядное квантование); при скоростях 40-30 кбит/с ДМ обеспечивает более высокое качество восстановления, чем ИКМ. Кривая 2 на рис.1.8 характеризует помехоустойчивость ДМ. Дельта-модуляция обладает наилучшими параметрами помехоустойчивости среди всех методов кодирования. Соответствующие системы не теряют работоспособности при возникновении одиночных ошибок и их пакетов (серий) малой длительности.

Еще один вид кодеров формы - методы дифференциальной (разностной) ИКМ (ДИКМ). Их классификационными признаками 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. Однако, в отличие от систем с ДМ, они теряют работоспособность при вероятности одиночной ошибки, составляющей около $5 \cdot 10^{-3}$, и передаче пакетов ошибок малой длительности.

К достижениям в области ЭКР можно отнести кодеры, реализующие алгоритмы анализа через синтез. Они сохраняют форму речевого сигнала (во всяком случае, к ним применима среднеквадратическая мера оценки восстановления, СКО). В этих кодерах используются алгоритмы сжатия, основанные на оценке параметров модели речеобразования, которые прежде применялись исключительно в вокодерах.

Все описанные методы предполагают передачу большого количества параметров речевого сигнала и эквивалента сигнала-остатка (используемого разностной ИКМ), которые квантуются с разной точностью. Прежде оценка признака тон/шум считалась отличительной чертой вокодера, теперь же она осуществляется и в кодерах анализа через синтез, что стирает границы между кодерами формы и вокодерами (поэтому их иногда называют полувокодерами).

Работа кодеков с многополосным кодированием (МПК, SubBand Coder) основана на различной чувствительности слуха к звукам, принадлежащим к разным частотным полосам. Это позволяет кодировать сигналы в полосах с разной точностью. Число полос может колебаться от 3 до 16.

В кодерах с ортогональным преобразованием (ОПА) скорость передачи снижается за счет грубого квантования спектральных составляющих, полученных разложением в ряд в каком-либо базисе.

Особенностью помехоустойчивости систем, основанных на последних двух методах (рис.1.8, кривая 3), является то, что благодаря различной точности кодирования в полосах отсутствует пороговый переход к области неработоспособности.

Появление методов MPE, RELP и CELP связано с совершенствованием кодеров формы, которое было предпринято для сохранения качества восстановленного речевого сигнала при менее высоких скоростях. В этих методах выявляется избыточность погрешности предсказания.

В кодерах с линейным предсказанием и усеченным возбуждением (RELP - Residual Excited Linear Prediction, ЛПУВ) сигнал погрешности ограничивается по частоте и прореживается. Кодеры с многоимпульсным возбуждением (MPE - MultiPulse Excitation, ЛПМВ) используют вместо сигнала-остатка искусственную последовательность возбуждения речевого сигнала на некотором временном интервале, параметры которой передаются в декодер. Выбор фазы такой последовательности осуществляется с помощью интерактивной процедуры по критерию близости формы исходного и синтезированного сигналов. На основе этого метода разработан алгоритм кодека стандарта GSM для подвижной связи, реализующий скорость передачи 13,8 кбит/с.

В последнее время большую популярность приобрели кодеры CELP (Code Excited Linear Prediction), разновидностями которых являются SELP, LD-CELP, V-CELP и A-CELP. Эти высокоэффективные кодеры обеспечивают отличное качество звука при низких скоростях (2,4-8 кбит/с). Для кодирования погрешности предсказания в них используются кодовые книги, состоящие из блоков с конечным числом символов. Перечисленные разновидности кодеров различаются способами формирования и хранения этих последовательностей. Чаще всего последовательность хранится в сжатом виде. Дополнительные буквы в названии кодера (LD, V и др.) указывают на способ реализации предсказателя, синтеза квантователя или кодовой книги. Кривая 4 на рис.1.8 характеризует помехоустойчивость таких кодеров. Здесь явно видны

две основные области: А соответствует помехоустойчивой работе (вероятность ошибки 10^{-3}), В - резкому уменьшению помехоустойчивости.

Проблема создания помехоустойчивых высокоэффективных кодеров является, по сути, проблемой согласования сигнала с каналом связи. Анализируя традиционные подходы к решению задачи согласования, можно отметить, что им присущи весьма существенные недостатки (например, при разделении операций эффективного и помехоустойчивого кодирования для обеспечения помехоустойчивости необходима высокая избыточность, что приводит к ужесточению требований к алгоритмам сжатия).

Еще одна проблема связана с выбором модели канала связи. Наиболее "неприятными" считаются каналы подвижной телефонии, характеристики которых зависят от нахождения подвижного объекта; причем 90 % пользователей утверждают, что качество канала постоянно меняется. Такой канал может быть описан с помощью некоторой переменной модели, например основанной на переменной вероятности ошибки. Это влияет на разработку метода помехозащиты.

Система передачи аналогового сигнала по цифровому каналу является оптимальной, когда кодер источника обеспечивает максимальное сжатие без потерь в качестве, а цифровой канал - максимальную скорость передачи при заданной вероятности ошибки. В этом случае уменьшение избыточности входного сообщения осуществляется кодером источника, а уменьшение вероятности ошибки - кодером канала. Оба алгоритма разрабатываются независимо друг от друга. Современные методы ЦПР убирают не всю избыточность речевого сигнала, поэтому такой подход применяется для существующих методов ЦПР, когда разработчик не хочет или не может использовать особенности речевого сигнала и способа его преобразования для повышения помехоустойчивости.

В рамках этого подхода используются специализированные помехоустойчивые коды, которые наиболее эффективны для алгоритмов CELP.

1.6 Гибридные методы кодирования РС

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 [11].

В случае использования набора фильтров частота ω фиксирована, так что $\omega = \omega_0$, а сигнал частотного домена $S_h(e^{j\omega_0})$ представляет собой сигнал на выходе постоянного во времени линейного фильтра с импульсной характеристикой $v(h)$, возбуждаемого модулированным сигналом $\lambda(h)e^{-j\omega_0 h}$:

$$S_h(e^{j\omega_0}) = v(h) \otimes [\lambda(h)e^{-j\omega_0 h}], \qquad (1.4)$$

где $v(h)$ определяет ширину полосы речевого сигнала $\lambda(h)$ вокруг центральной частоты ω_0 и является импульсной характеристикой анализирующего фильтра; знак ⊗ означает свертку функций Θ.

При использовании блока, реализующего преобразование Фурье, временной индекс h фиксируется на значении $h = ho$, а $S_{h_0}(e^{j\omega})$ представляет собой обычное преобразование Фурье взвешенной последовательности $v(h_0 - m)\lambda(m)$:

$$S_{h_0}(e^{j\omega}) = F[v(h_0 - m)\lambda(m)], \qquad (1.5)$$

где $F[\cdot]$ – преобразование Фурье.

Здесь $v(h_0 - m)$ определяет отрезок времени анализа относительно момента времени $h = ho$ и является «окном анализа» $\omega(h_0 - - m)$.

Уравнение синтезирующего набора фильтров

$$\hat{\lambda}(h) = \frac{1}{2\pi v(0)} \int_{-\pi}^{\pi} S_h(e^{j\omega}) e^{j\omega m} d\omega \qquad (1.6)$$

может быть представлено как интеграл (или сумма) компонентов – кратковременных спектров $S_h(e^{j\omega_0 h})$ с несущими частотами ω_0.

Для синтеза с помощью блока преобразования уравнение выглядит следующим образом:

$$\lambda(h) = \frac{1}{V(e^{j\omega_0})^r} = \sum_{r=-\infty}^{\infty} F^{-1}[S_r(e^{j\omega})]. \qquad (1.7)$$

Его можно интерпретировать как сумму инверсных преобразований Фурье, примененных к временным сигналам $v(r-h)\lambda(h)$.

Вокодеры с многополосным возбуждением MBE (Multi-Band Excitation) [14]. При сравнении с традиционными вокодерами LPC для скоростей ниже 4,8 кбит/с и кодерами с адаптивным предсказанием для скоростей около 16 кбит/с кодеры MPLPC и CELP (рис.1.8) дают существенное улучшение качества речи для диапазона скоростей 4,8… ..16 кбит/с. Эти улучшения касаются, в основном, характеристики квантования сигнала возбуждения после удаления структуры основного тона. При этих улучшенных способах (MPLPC и CELP) в качестве критерия ошибки используется взвешенная разница между исходной и синтезированной версиями речи. Поэтому эти способы можно рассматривать как гибридные кодеры формы сигнала, в которых для определения качества речи используется схожесть или близость исходной и синтезированной речи. Однако на скорости 4,8 кбит/с и ниже ограничения в точности определения возбуждения становятся определяющими, что приводит к быстрому ухудшению качества речи. Снижение качества речи при использовании MPLPC и CELP связано с неточностью представления гармоник в вокализированных частях речевого спектра, что вызывает общее зашумление речи.

В кодерах с многополосным возбуждением (MBE) для представления сигнала возбуждения используются разные способы (рис.1.9). Кодеры речи MBE заменяют единственную классификацию вокал/невокал классических вокодеров на несколько таких определений по гармоническим интервалам в частотной области. Это дает возможность представить каждый сегмент как смесь вокала и невокала. Процесс определения совпадения в кодере MBE более ориентирован на восприятие, для чего схожесть формы сигнала не важна. Вообще, когда речь является невокалом, исходный и синтезированный сегменты речи не должны иметь никаких сходств во времени области. Огибающая речи может представляться традиционными методами.

Анализ взвешенного сегмента речи по частоте делается путем выполнения преобразования Фурье. На коротком интервале преобразование Фурье взвешенного сегмента речи можно имитировать как продукт преобразования огибающей спектра $H(\omega)$ спектра возбуждения $X(\omega)$. Огибающая спектра описывает форму огибающей спектра речи, которая может бы рассмотрена как сглаженный вариант спектра исходной речи. Поэтому он может быть представлен различными способами. Хотя точное представление огибающей речи столь же важно в MBE, как

и в других кодерах, форма представления огибающей спектра не является главным фактором в MBE кодерах.

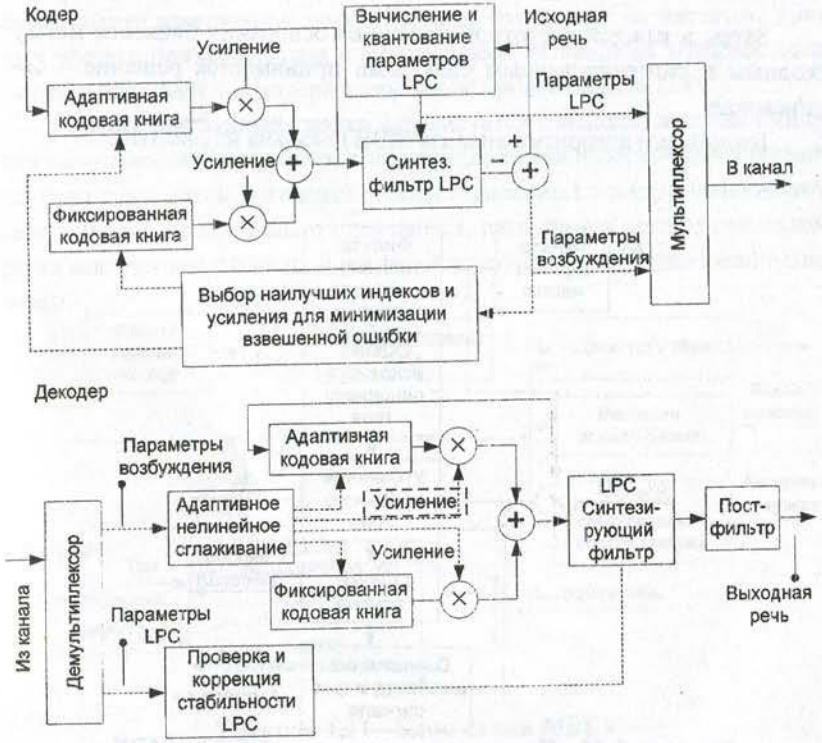


Рисунок 1.9 - Блок-схема многополосного кодера/декодера

Спектр возбуждения в модели речи MBE имеет одно главное отличие от традиционных простых моделей, в традиционных вокодерах, т.е. в LPE и канальных вокодерах, спектр возбуждения определен основной частотой ω_0 и единственным решением вокал/невокал для остального спектра. В кодерах речи MBE спектр возбуждения определяется основной частотой ω_0 и более точными определениями вокал/невокал относительно частоты. Из-за множественных определений вокал/невокал эта модель называется моделью с многополосным возбуждением.

Анализ речи делится на два основные этапа. На первом — оцениваются период основного тона и параметры огибающей спектра для минимизации ошибки между исходным $S_W(\omega)$ и синтезированным спектрами $\hat{S}_W(\omega, \omega_0)$ по критерию минимума среднеквадратичной ошибки:

$$\varepsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|S_W(\omega) - \hat{S}(\omega,\omega_0)\right|^2 d\omega. \qquad (1.8)$$

Затем в каждой частотной полосе на основании близости между исходным и синтезированным спектрами принимается решение – вокал/невокал.

Блок-схема алгоритма анализа MBE показана на рис.1.10.

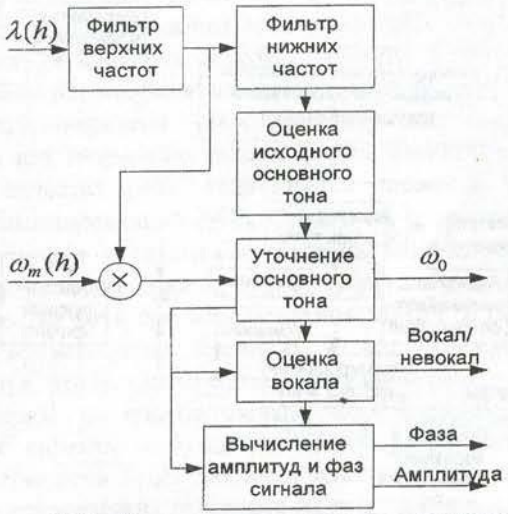


Рисунок 1.10 - Блок-схема алгоритма анализа MBE

Параметры речевого кодера MBE, которые нужно оценить для каждого фрейма речи, являются периодами основного тона, характеристиками — вокал/невокал и амплитудами гармоник, которые характеризуют огибающую спектра.

MBEV — Multi-Band Excited Vocoder. В MBEV вводится высокоэффективное тональное моделирование, которое позволяет кодеру синтезировать речь с хорошим качеством даже на скорости 4,8 кбит/с.

Блок-схема MBEV, представленная на рис.1.11, работает следующим образом. На первом этапе оценивается тональный период речевого фрейма с использованием автокорреляционного метода для НЧ-фильтрованной речи (частота среза фильтра приблизительно 1,2 кГц) [11…15]. Затем речь взвешивается и подвергается быстрому преобразо-

ванию Фурье. Целью взвешивания является подавление краевых эффектов при преобразовании. При допущении, что речь является периодической, спектр взвешенной речи содержит гармоники на частотах, кратных частоте основного тона с амплитудами, огибающая которых идентична огибающей для кодера с линейным предсказанием (LPC).

Синтезированный спектр формируется с использованием амплитуд гармонических составляющих исходной речи и оценочных значений частоты тона. Затем исходный и синтезированный спектры сравниваются и частоты, которые мало отличаются, признаются вокализованными, тогда как частоты с большой разницей признаются невокализованными.

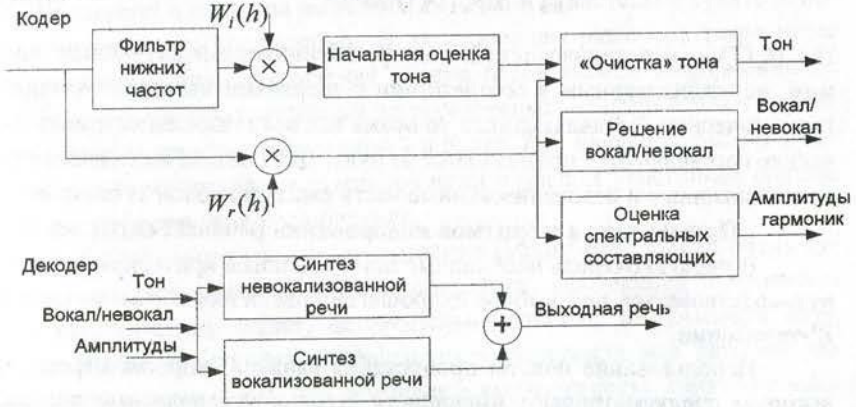


Рисунок 1.11 - Блок-схема MBEV

Кодер передает следующие параметры для описания сегмента входного речевого сигнала:

1) тональный период фрейма. После оценки тона с использованием автокорреляционной функции тон во фрейме может быть «очищен»;

2) амплитуды гармоник. Значения невокализованных составляющих представляются среднеквадратическим значением шума в окрестности каждой гармоники в поддиапазоне. Вокализованные составляющие представлены амплитудами гармоник преобразованной речи;

3) флаг вокал/невокал.

Для каждого фрейма решение принимается с использованием одной или нескольких гармоник в каждом частотном поддиапазоне. Если присутствуют гармоники на частотах более 3,4 кГц, они принимаются невокализованными.

В декодере вокализованные составляющие спектра восстанавливаются с использованием соотношения

$$\lambda_g(h) = \sum_{l=1}^{L} A_l \cos(h\omega_l + \varphi_l), h = 0,1,\dots,H, \quad (1.9)$$

где A_l, ω_l и φ_l – амплитуда, частота и фаза гармоники номер l соответственно; H – размер фрейма. Информация о фазе на низких скоростях предсказывается из текущей и предыдущих тональных частот (у гармоник фаза должна быть такой, чтобы все гармоники суммировались). Невокализованные составляющие спектра находятся с использованием соотношения

$$\lambda_{HB} = \sum_{f=1}^{W} U_u(f) e^{j2\pi ft}, \quad (1.10)$$

где $U_u(f)$ – невокализованный спектр, образованный случайным шумом, нормализованным в соответствии с энергиями невокализованных гармонических составляющих, в то время как вокализованные гармонические составляющие принимаются за нуль. Для синтеза выходной речи вокализованные и невокализованные части слагаются блок за блоком.

Применение алгоритмов кодирования речевых сигналов

В первую очередь необходимо понять, какими критериями нужно руководствоваться при выборе «хорошего» кодека для использования в IP-телефонии.

Использование полосы пропускания канала. Скорость передачи, которую предусматривают имеющиеся сегодня узкополосные кодеки, лежит в пределах 1.2-64 кбит/с. Естественно, что от этого параметра прямо зависит качество воспроизводимой речи. Существует множество подходов к проблеме определения качества. Наиболее широко используемый подход оперирует оценкой MOS (Mean Opinion Score), которая определяется для конкретного кодека как средняя оценка качества большой группой слушателей по пятибалльной шкале. Для прослушивания экспертам предъявляются разные звуковые фрагменты - речь, музыка, речь на фоне различного шума и т.д. Оценки интерпретируют следующим образом:

• 4 - 5 - высокое качество; аналогично качеству передачи речи в ISDN или еще выше;

• 3.5 – 4 - качество ТфОП (toll quality); аналогично качеству речи, передаваемой с помощью кодека АДИКМ при скорости 32 кбит/с. Такое качество обычно обеспечивается в большинстве телефонных разговоров. Мобильные сети обеспечивают качество чуть ниже toll quality;

• 3 - 3.5 - качество речи по-прежнему удовлетворительно, однако его ухудшение явно заметно на слух;

• 2.5 - 3 - речь разборчива, однако требует концентрации внимания для понимания. Такое качество обычно обеспечивается в системах связи специального применения (например, в вооруженных силах).

В рамках существующих технологий качество ТфОП (toll quality) невозможно обеспечить при скоростях менее 5 кбит/с. Подавление периодов молчания (VAD, CNG, DTX). При диалоге один его участник говорит в среднем только 35 % времени. Таким образом, если применить алгоритмы, которые позволяют уменьшить объем информации, передаваемой в периоды молчания, то можно значительно сузить необходимую полосу пропускания. В двустороннем разговоре такие меры позволяют достичь сокращения объема передаваемой информации до 50 %, а в децентрализованных многоадресных конференциях (за счет большего количества говорящих) - и более. Нет никакого смысла организовывать многоадресные конференции с числом участников больше 5-6, не подавляя периоды молчания.

Нужно отметить, что определение границ пауз в речи очень существенно для эффективной синхронизации передающей и приемной сторон: приемник может, незначительно изменяя длительности пауз, производить подстройку скорости воспроизведения для каждого отдельного сеанса связи, что исключает необходимость синхронизации тактовых генераторов всех элементов сети, как это имеет место в ТфОП.

1.7 Методика испытаний

Запись тестовых речевых сигналов

В качестве речевого материала используются слоговые артикуляционные таблицы и тестовые фразы, приведенные в ГОСТ Р 50840-95. Речевой материал начитывается 10 различными дикторами.

Запись РС осуществляется в специальном помещении кабинетного типа (с размерами 5,7*2,9*3 м и временем реверберации около 350 мс) при наличии естественного фонового шума слабого уровня.

Для записи используется профессиональный диктофон Olimpus LS-10 (Linear PCM recorder), обеспечивающий возможность записи РС в формате WAV с параметрами: частота дискретизации 44,1 кГц, разрядность квантования 16 бит, тип кодирования ИКМ, диктофон устанавливается на расстоянии 0,5 м перед диктором на уровне лица.

Описание схемы получения объективной оценки качества речи

Записанный на диктофон РС поступает на полосовой фильтр ПФ с полосами пропускания 0,1-3,4 кГц; 0,1-5,0 кГц; 0,1-7,0 кГц и 0,1-8,0 кГц и далее конвертируется программным конвертором ПК с частотами дискретизации 8000, 11025, 16000 и 22050 Гц соответственно, а также разрядностями квантования 8 и 16 бит.

Сформированные в блоке формирования помехи БФП широкополосные и низкочастотные, квазистационарные и нестационарные помехи программно суммируются с РС (с выхода БАО), и полученная аддитивная смесь поступает на вход блока КОДЕР РС. В блоках КОДЕР РС и ДЕКОДЕР РС происходит кодирование и декодирование РС в соответствии с принятыми по международным стандартам кодеками.

Декодированный РС после блока восстановления пауз БВП поступает на акустические колонки, а также в блок объективной оценки качества БООК, где в зависимости от положения ключа К: «1» – выходными данными является объективная оценка качества РС, включающая оценки слоговой разборчивости и узнаваемости декодированного РС, при наличии тестового РС; «2» – выходными данными является объективная оценка слоговой разборчивости декодированного РС при отсутствии тестового РС. В блоке БООК при положении ключа К «1» осуществляется предварительная обработка сигнала: выравнивание по уровню и времени исходного и декодированного РС, слуховое преобразование, имитирующее особенности человеческого уха. При положении ключа К «2» осуществляется только слуховое преобразование.

Блок управления БУ осуществляет выбор режима работы блоков ПФ, ПК, БСЧД, К, БАО, БПФ, КОДЕР РС и ДЕКОДЕР РС.

Формирование искажений речевого сигнала и аддитивных помех

В качестве широкополосной стационарной аддитивной помехи используются реализации белого гауссовского шума (БГШ), генерируемые программно в пакете Matlab. Интегральные отношения сигнал-шум (ОСШ) в аддитивной смеси РСи помехи составляют 20, 10 и 0 дБ. В качестве низкочастотной квазистационарной помехи используется гауссовский шум, ограниченный в полосе (0,3-1) кГц и (0,1-1,0) кГц. Помеха получена на выходе КИХ-фильтра, АЧХ которого имеет наклон 9 дБ/октаву в сторону высоких частот. В качестве нестационарной широкополосной помехи используется акустическая запись произвольного фрагмента музыки с длительностью, соответствующей длительности тестового РС, и с полосой частот (0,3-3,4) кГц или (0,1-8,0) кГц. В качестве нестационарной низкочастотной помехи используется акустическая запись транспортного потока с длительностью, соответствующей длительности тестового РС, и с полосой частот (0,3-1) кГц и (0,1-1,0) кГц.

Кодирование речевого сигнала

Кодирование осуществляется с помощью стандартизованных звуковых и речевых кодеков (параметрические, психоакустические и с кодированием формы волны сигнала) со скоростью потока от 0,8 до 64 кбит/с.

Субъективная оценка качества речи

Получение средних субъективных оценок качества звучания РС по показателям слоговой разборчивости и узнаваемости диктора проводится согласно ГОСТ Р50840-95. В НИР участвовали 10 дикторов и 10 аудиторов при тестировании каждого кодека.

Результаты заносятся в протокол, приведенный в таблице 1.2.

Таблица 1.2 – Протокол, содержащий результаты измерений

<table>
<tr><td colspan="7">Таблица № Дата:
Диктор: Аудиторы:
Условия проведения испытаний:</td></tr>
<tr><td rowspan="4"></td><td colspan="6">Кодер:</td></tr>
<tr><td colspan="6">Искажения: АО_____ дБ; СЧД_____ кГц
Помеха: _____;
Диапазон/значение помехи:_____;</td></tr>
<tr><td colspan="6">Номер записи</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Разборчивость</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Узнаваемость</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

2. ПРАКТИЧЕСКАЯ ЧАСТЬ

Инструкция к программе «VOCdemo».

1. Открыть программу «*VOCdemo.exe*».
2. В поле «*Vocoder*» выбрать исследуемый кодек.
3. В поле «*Input*» нажать на значок папки и в открывшемся окне выбрать входной файл:
папка **"В кодек"**, далее открыть папку **«сигнал»** (или **«производная»**), затем папку **«исследуемого шума»**, затем папку **«исследуемого уровня»** и, наконец, **исследуемый файл**.
4. В поле «*Output*» нажать на значок папки и в открывшемся окне выбрать папку **"От кодека"**, далее папку **«сигнал»** , затем папку **с названием исследуемого кодека**, затем папку **«исследуемого шума»**, затем папку **«исследуемого уровня»**. В поле «Имя файла» необходимо ввести тот же номер, что и у входного файла.
5. Далее необходимо нажать на кнопку с красным кругом в поле «*Output*» и тем самым записать выходной файл.

Подобные операции необходимо проделать со всеми файлами из папки «**В кодек**» через все кодеки.

Примечание

Необходимо, чтобы пути в поле «*Input*» и «*Output*» строго соответствовали друг другу.

Например:

Путь в поле « *Input*»:

Вкодек\сигнал\сигнал+двигатель\0дБ\1-s-InK.wav

Тогда путь в поле «*Output*» должен быть:

От кодека\сигнал\ исследуемый кодек (например: itu G 726_24)\сигнал+двигатель\0дБ\1

2.1 Измерение разборчивости речи артикуляционным методом

1. Измерения проводит бригада операторов в составе которой должно быть не менее трех дикторов (двух мужчин и одной женщины) и трех аудиторов. Состав бригады аудиторов произвольный. Аудиторы могут быть дикторами, прошедшими специальное обучение

(тренировку) путем прослушивания на головных телефонах слоговых артикуляционных таблиц.

2. Бригаду операторов рекомендуется обучать в два этапа.
3. На первом этапе обучения операторы знакомятся со структурой речевого материала, осваивают технику его произношения, а также адаптируются к восприятию речи, искаженной в испытуемом тракте (аппаратуре) в соответствующих акустических условиях.
4. Чтение слогов осуществляется диктором ровным голосом, четко, но без подчеркивания отдельных звуков с постоянным уровнем речи, который контролируется шумомером на испытательной фразе «Не видали мы такого невода».

Слоги следует читать в следующем ритме: 1 слог в (3 ± 0,3) с.

5. Диктор должен выдерживать постоянный ритм речи на протяжении чтения всей таблицы.
6. Аудитор записывает принятые слоги в бланк, форма которого приведена на рис.2.1.

Таблица № _____ Дата _____
Диктор: _____
Тип Тракт _____
Аудитор _____
Уровень шума, дБ _____

Чтение таблицы слогов: (в столбик, в строчку)

1		11		21		31		41	
2		12		22		32		42	
3		13		23		33		43	
4		14		24		34		44	
5		15		25		35		45	
6		16		26		36		46	
7		17		27		37		47	
8		18		28		38		48	
9		19		29		39		49	
10		20		30		40		50	

Рисунок **2.1 - Форма бланка для записи слогов**

Если аудитор не понял переданного слога, он прочеркивает соответствующую пронумерованную строку в бланке принятых слогов.

На втором этапе тренировки проводится цикл измерений при использовании испытуемого тракта.

Цикл измерений включает в себя прием всеми аудиторами от всех дикторов по $5 * k$ таблиц, где $k = 1,2,3..$ Пятерки таблиц должны иметь номера 1-5,6-10,11-15 и т.д. Использование при измерении неполной пятерки не допускается. Слоговые таблицы с одинаковыми номерами должны использоваться не чаще одного раза в неделю.

Для каждого измерения вычисляется среднее значение разборчивости (S) по формуле:

$$S = \frac{1}{N}\sum_{i=1}^{N} S_i, \qquad (2.1)$$

где S_i - результат измерения, % (диктор - таблица - аудитор); N - число измерений. Вычисляют среднеквадратичное отклонение (СКО) по формуле:

$$R = \frac{\sum_{i=1}^{N}[S_i - S]}{N}. \qquad (2.2)$$

Результаты измерений, для которых $[S_i - S] > 2\sigma$ или $[S_i - S] >> 2R$, исключаются и производится вычисление нового среднего значения по формуле:

$$S = \frac{1}{N-k}\sum_{i=1}^{N-k} S_i, \qquad (2.3)$$

где S_i - результат измерения, % (диктор - таблица - аудитор); N - число измерений; k - число исключенных измерений. Тренировку считают законченной при достижении бригадой стабильных результатов измерения разборчивости (повторяемость значений средней разборчивости по бригаде в течение 2-3 дней).

Аудиторы, результаты которых имеют систематическое отклонение от средних значений по бригаде более чем на величины, указанные в таблице 2.1, подлежат замене или исключению из бригады.

Таблица 2.1 - Максимально допустимое отклонение от среднего значения по бригаде

Среднее значение разборчивости по бригаде	Отклонение от среднего значения
91 и более	5
86-90	6
81-85	7
71-80	8
70 и менее	9

Время работы бригады должно быть не более 4 часов за один день. После приема пяти таблиц делается перерыв 5-10 минут. Общее число таблиц за одно измерение до 40.

При работе в акустических шумах бригада приступает к измерениям спустя 5-10 минут после пребывания в условиях шума. Общее число таблиц - 30 (при уровне шума 80-100 дБ) и 20 (при уровне шума более 100 дБ). Существуют следующие классы качества речи, приведенные в таблице 2.2.

Опыты, проведенные по оценке предельной разборчивости [7], показали, что полная потеря связи соответствует D < 60 %, с другой стороны, удовлетворительная разборчивость фраз имеет место при D > >75 %. Учитывая это, примем за величину предельной разборчивости Jmin = 70 %, что соответствует D = 77 % и S = 46 %. В особо тяжелых условиях работы (очень высокий уровень шумов) можно допустить уменьшение D до 70 % (среднее значение разборчивости IV класса), при этом речь разбирается с трудом, передача идет с переспросами и повторениями.

Таблица 2.2 - Классы качества речи

Класс качества	Характеристика класса качества	Артикуляция, %		
		Звуковая	Слоговая	Словная
1	Понимание передаваемой речи без малейшего напряжения внимания	Более 90	Более 90	Более 98
2	Понимание передаваемой речи: без затруднений	Св. 85 - 90	Св. 80 - 90	Св. 94 - 98
3	Понимание передаваемой речи с напряжением внимания без переспросов и повторений	Св. 78 - 85	Св. 68 - 80	Св. 89 - 94
4	Понимание передаваемой речи с большим напряжением внимания с переспросами и повторениями	Св. 60 - 78	Св. 20-68	Св. 70 - 89
5	Неразборчивость связного текста, срыв связи	60 и менее	20 и менее	70 и менее

Интересно отметить, что при низких значениях слоговой артикуляции артикуляция фраз меньше артикуляции слов. Это вполне закономерно, так как при плохой слышимости фраза становится понятной лишь при правильном приеме определенного минимума ключевых слов.

2.2 Измерение качества речи методом оценки по селективным признакам

1. Измерения проводит бригада в составе, указанном в 2.1, п.1, путем прослушивания аудиторами фраз, прошедших через контрольный и испытуемый тракты (аппаратуру).

Фразы для прослушивания передают с интервалом 2—3 с. Число прослушиваний каждой фразы неограниченно.

2. Аудиторы проводят сравнение звучания фразы, прошедшей через контрольный тракт, и фразы, прошедшей через испытуемый тракт (аппаратуру), и определяют наличие следующих селективных признаков искажения в звучании речи относительно контрольного тракта:

— картавость;
— плаксивость ;
— гнусавость;
— механический голос;
— дребезжание, хрип;
— помеха в паузах речи.

Таблица 2.3 - Оценка степени искажения

Степень искажения признака	Норма, балл
Отсутствует	0
Присутствует (редко встречается)	1
Выражен сильно (присутствует постоянно)	2

Оценку степени искажения признаков (в баллах) в голосе каждого диктора осуществляют в соответствии с таблицей 2.3.

Оценку проставляют в бланке, форма которого приведена на рис.2.2.

Дата_____ Тип тракта_____
Аудитор_____ Уровень шума, дБ _____

Наименование признака	Диктор № 1	Диктор № 2	Диктор № 3	Диктор № 4	Диктор № 5
Картавость					
Гнусавость					
Плаксивость					
Механический голос					
Дребезжание, Хрип					
Помеха					

Рисунок 2.2 – Бланк для проставления оценок

3. По данным измерений вычисляют среднее значение X степени искажения каждого из шести селективных признаков по формуле

$$X = \frac{1}{N}\sum X_i \quad (2.4)$$

где X_i — результат единичного измерения (аудитор — диктор);

N — число единичных измерений.

4. При использовании дополнительного оборудования на экране дисплея ПЭВМ высвечивается перечень селективных признаков.

Аудитор должен набрать на клавиатуре порядковый номер признака искажения и дать оценку в баллах степени его выраженности в соответствии с таблицей 2.3. Методика работы на клавиатуре содержится в пакете программ «КРЕС».

5. Метод рекомендуется использовать при углубленном анализе факторов искажения речи, а также для установления класса качества речи.

2.3 Текстовые фразы для оценки качества речи и узнаваемости голоса диктора

1 Основные фразы:

1.1 Если хочешь быть **здоров**, советует Татьяна Илье, чисти **зубы** пастой «Жемчуг»!

1.2 Так ты считаешь, что техникой мы обеспечены на **весь** сезон?

Примечание: полужирным шрифтом выделены слова, на которые делается ударение. Вставочные конструкции во фразе 1 произносятся в убыстренном темпе. Фраза 1.2 произносится без пауз между словами.

2. Дополнительные фразы:

2.1 Раз. Это жирные фазаны ушли под палубу.

2.2 Алло, слушаю! Кто у телефона? Ах, это Вы! Я был вчера у Вас.

Получение средних субъективных оценок качества звучания РС по показателям слоговой разборчивости и узнаваемости диктора проводится согласно ГОСТ Р50840-95. При выполнении работы использовать 10 дикторов и 10 аудиторов при тестировании каждого кодека.

Результаты измерений заносятся в протокол, приведенный в таблице 2.4.

Таблица 2.4 – Запись результата

Дата: Диктор: Аудиторы: Условия проведения испытаний:						
	Кодер:					
	Помеха: _____; Диапазон/значение помехи : _____;					
	Номер записи					
Разборчивость						
Качество						

2.4 Задание для моделирования кодека РС

1. Изучить структуру предложенного алгоритма кодирования, взятого из таблицы 2.5 согласно вашему варианту задания, привести описание стандарта.
2. Реализовать выбранный кодек РС на заданной скорости передачи в программной среде Matlab. Для реализации кодеков можно воспользоваться программой «VOCdemo».
3. Разработать согласно ГОСТ Р 50840-95 и ГОСТ Р 51061-97 методику оценки качества речевых сигналов на выходе кодеков.
4. Согласно методике, ГОСТ Р 50840-95 и ГОСТ Р 51061-97 провести запись РС бригадой дикторов.
5. Подать на вход кодека сигналы, записанные группой дикторов. Получить на выходе кодека кодированные РС.
6. Эксперимент провести при различных акустических шумах и шумах и искажениях в канале связи.
7. Провести согласно полученной методике оценку субъективного качества полученных сигналов. Получить усредненную оценку качества.

8. Провести согласно алгоритмам объективной оценки анализ полученных сигналов.
9. Сравнить полученные объективную и субъективную оценки.
10. При воздействии различных шумов выявить корреляцию между объективными и субъективными оценками.

В отчете должны быть: структурная схема кодека, его описание, область практического применения.

Таблица 2.5 – Вариант задания

Номер варианта	Варианты заданий
1	АДИКМ – 32 кбит/с
2	АДИКМ – 24 кбит/с
3	АДИКМ – 16 кбит/с
4	LBRAMR – 1 кбит/с
5	LBRAMR – 2 кбит/с
6	MMBE – 2,4 кбит/с
7	G723.1 – 6,3 кбит/с
8	ICELP – 4,8 кбит/с
9	G729a – 8 кбит/с
10	G728i – 16 кбит/с
11	AMR – 6,6 кбит/с
12	AMR – 12,65 кбит/с
13	AMR – 15,85 кбит/с
14	AMR – 19,85 кбит/с
15	AMR – 23,85 кбит/с
16	GSM – 13 кбит/с
17	GSM – 17 кбит/с
18	GSM – 35 кбит/с
19	G726 – 16 кбит/с
20	G726 – 24 кбит/с
21	G726 – 32 кбит/с
22	G726 – 40 кбит/с
23	Vorbis OGG – 64 кбит/с
24	MPEG 1.2.8 – 32 кбит/с
25	MPEG 1.2.8 – 48 кбит/с
26	MPEG 1.2.8 – 56 кбит/с
27	MPEG 1.2.8 – 64 кбит/с

2.5 Пример реализации кодека в среде программирования Matlab

Кодек G723:

```
function G7231Coder
(FNameI, FNameO)
% ITU-T G.723.1 speech
coder
% FNameI: Input audio
file (.wav, .au, ... )
% FNameO: Output bit
stream file (.bit) or
output data file
(.dat)

% P. Kabal,
www.TSP.ECE.McGill.CA

% $Id: G7231Coder.m,v
1.15 2010/02/10
20:14:55 pkabal Exp $

G7231Dir = fileparts
(which
('G7231Coder.m'));
addpath (fullfile
(G7231Dir, 'ACB'),
fullfile (G7231Dir,
'Audio'), ...
        fullfile
(G7231Dir, 'Bit-
Stream'), ...
        fullfile
(G7231Dir, 'Filter'),
fullfile (G7231Dir,
'LP'), ...
        fullfile
(G7231Dir, 'Misc'),
...
        fullfile
(G7231Dir, 'Mul-
tiPulse'), ...
        fullfile
(G7231Dir, 'Tables'),
fullfile (G7231Dir,
'Target'));

% Set up the output
file name
if (nargin <= 1)
 [pathstr name] =
fileparts (FNameI);
 FNameO = fullfile
('', [name, '.bit']);
end
[pathstr name ext] =
fileparts (FNameO);
isBitStream = strcmpi
(ext, '.bit');

% Coder parameters
Np = 10;
LSubframe = [60 60 60
60];
CoderPar = SetCoderPar
(Np, LSubframe);

HPFilt = Coder-
Par.HPFilt;
LPpar = Coder-
Par.LPpar;
LSFpar = Coder-
Par.LSFpar;
Pitchpar = Coder-
Par.Pitchpar;
TVpar = Coder-
Par.TVpar;
SineDetpar = Coder-
Par.SineDetpar;
MPpar = Coder-
Par.MPpar;

% Initialize the coder
memory
CoderMem = InitCoder-
Mem (CoderPar);

% Initialize the
weighted synthesis
filter memories
WSynMem0 = WSynInit
(length
(TVpar.HNWpar.xp));
WSynMem = WSynMem0;

LFrame = sum (LSub-
frame);

% Open the input file
AFPar = OpenAudioFile
(FNameI);
if (AFPar.Sfreq ~=
8000)
 error ('G7231Coder:
Sampling rate must be
8 kHz');
end

% Initialize the pre-
vious frame data
% The frame consists
of 3 parts, [lookback
Frame lookahead]
% The lookback and
lookahead are each
half a window long.
The total memory
% of the system is
lookback+lookahead.
New data is read into
the end of the
% array.
offs = 0;
xpMem = zeros
(LPpar.LMem, 1);

FSt = LPpar.FStart+1;
FEn = FSt + LFrame -
1;
FMode = 1;          %
Multipulse
```

```
TickTock (LFrame / AF-
Par.Sfreq, 1, '< Data
Time: %d s >\n');
IFr = 0;
while (1)

 % Read a frame of au-
dio data
 [x, Nv, AFPar] =
ReadAudioData (AFPar,
offs, LFrame);
 if (Nv < LFrame)
   break
 end

 TickTock;
 IFr = IFr + 1;

 % Highpass filter
 [xf, HPFilt.Mem] =
filter (HPFilt.b,
HPFilt.a, x,
HPFilt.Mem);

 % Form the extended
data buffer of high-
passed signal
 % Concatenate the new
values (1 frame) onto
the saved values
 xe = [xpMem; xf];

 % LP analysis
 a = LPanalFrame (xe,
LPpar);

 %  % Find the open
loop pitch estimate
 % Perceptual weight-
ing filter coeffi-
cients
  xc = xe(FSt:FEn);
% Current frame sam-
ples
  [xt, LOL, WSynCof,
TVpar] = GenTarget
(xc, a, TVpar);

 % Set up the data
memory for the next
frame (save LMem val-
ues)
 xpMem = xe(end-
LPpar.LMem+1:end);

 % Sine detector (from
LP parameters)
 [SineDet, SineDet-
par.rc] = SineDetector
(a, SineDetpar);

 % Quantize the LP pa-
rameters (as LSFs)
 % LSFC:  Codes to in-
dex the LSF codebooks
(one set per frame)
```



```
 LSFC = LPxLSFQ
(a(:,LPpar.SFRef),
CoderMem.lsfQ,
LSFpar);

 % Inverse quantiza-
tion of the LSFs
 % Interpolation of
the LSFs
 % Convert LSFs to LP
coefficients
 % aQI:   Quantized
interpolated LP pa-
rameters (one per sub-
frame)
 % lsfQ:  Quantized
LSF's (saved for dif-
ferential coding)
 [aQI, CoderMem.lsfQ]
= LSFCxLPI (LSFC,
CoderMem.lsfQ,
LSFpar);

 NSubframe = length
(LSubframe);
 LPrev = NaN;
 j = 0;
 for (i = 1:NSubframe)

   N = LSubframe(i);

   % Set up the
 weighted synthesis
 filter coefficients
   % The weighted syn-
thesis filter has
three parts
   % 1. Quantized all-
pole filter (aQ).
These are added here.
   % 2. Formant
weighting filter (set
up earlier), derived
from
   %     the unquan-
tized LP parameters
   % 3. Harmonic
weighting filter (set
up earlier)
   % Calculate the
weighted synthesis
filter impulse re-
sponse

   WSynCof(i).aQ =
aQI(:,i);
   wIR = WSyn
(eye(N,1), WSynCof(i),
WSynMem0);

   % Subtract the
zero-input response
from the target
   % This uses the
state from the previ-
ous subframe
```

```
   zIR = WSyn
(zeros(N,1),
WSynCof(i), WSynMem);
   xtz = xt(j+1:j+N) -
zIR;

   % Adaptive codebook
contribution
   % The ACB searches
around the open loop
lag for the first and
   % third subframes
and around the previ-
ous closed loop lag
for
   % the second and
fourth subframes
   % ACBLC: Lag code
for the pitch lag (lag
relative to the mini-
mum lag
   %       or relative
to the previous lag)
   % ACBbIB: Gain in-
dices, index and code-
book index.
   PMode = Pitch-
par.PMode(i);
   if (PMode == 1)
     Lx = LOL(i);
   else
     Lx = LPrev;
   end
```



```
   [ACBLC(i), ACB-
bIB(:,i), Pitch-
par.Tamepar.E] = ...

GetACB (xtz, Coder-
Mem.eMem, wIR, PMode,
...
Lx, SineDet, Pitch-
par);

   % Update the target
vector with the pitch
contribution

   [L(i), b(:,i)] =
DecodeACBSF (ACBLC(i),
ACBbIB(:,i), PMode,
LPrev, ...

Pitchpar); % Excita-
tion contribution
   ep = PitchContrib
(N, L(i), b(:,i),
CoderMem.eMem, Pitch-
par);
   xp = filter (wIR,
1, ep);
% Weighted signal con-
tribution

   xtpz = xtz - xp;

   % Fixed codebook
pulse positions
```

```
   if (PMode == 1)
     Lx = L(i); % The
lags used are: L(1),
L(1), L(3), L(3)
   else
     Lx = LPrev;
   end
   Pulseval(i) = Mul-
tiPulse (xtpz, wIR,
Lx, i, MPpar);
   LPrev = L(i);

   % Fixed codebook
contribution
   em(:,i) = MPContrib
(N, Pulseval(i));
   es = ep + em(:,i);

   % Update the filter
memories, ignoring the
output
   [temp, WSynMem] =
WSyn (es, WSynCof(i),
WSynMem);

   % Update the exci-
tation memory
   CoderMem.eMem =
ShiftVector (Coder-
Mem.eMem, ...

ClipSignal (es,
CoderPar.Clippar));
   j = j + N;
 end
 % Create the coded
data
 if (isBitStream)
   QC(IFr) =
CodeStream (FMode,
LSFC, ACBLC, ACBbIB,
Pulseval, ...

Pitchpar, MPpar);
 else
   VC(IFr) = CodeValue
(FMode, aQI, L, b,
em);
 end

 % Move ahead in the
speech
 offs = offs + LFrame;

end

% Write the bitstream
or data file
if (isBitStream)
 WriteG7231Stream
(FNameO, QC);
else
 save (FNameO, 'VC',
'-mat');
 fprintf ('G723.1 data
file: %s\n', FullName
(FNameO));
end
```

```
return

%--------------------
function QC =
CodeStream (FMode,
LSFC, ACBLC, ACBbIB,
Pulseval, ...

Pitchpar, MPpar)
% Create the coded bit
stream data

% Code the gains for
the pitch and the mul-
tipulse contributions
CGC = CodeGain
(ACBbIB, Pulseval,
Pitchpar);

% Multipulse grid
codes, sign codes,
positions codes
[MPGridC, MPPosC,
MPSignC] = MPCode
(Pulseval, MPpar);

% Save the codes for
transmission
QC.FType = FMode - 1;
QC.LSFC = LSFC;
QC.ACBLC = ACBLC;
QC.CGC = CGC;
QC.MPGridC = MPGridC;
QC.MPPosC = MPPosC;
QC.MPSignC = MPSignC;
return
%--------------------
function VC = Code-
Value (FMode, a, L, b,
em)
% Save the coded val-
ues
% FType: Type (MP /
ACELP)
% a:   LP coefficients
for each subframe
% L:   Pitch lag for
each subframe
% b:   Pitch coeffi-
cient vectors for each
subframe
```

Содержание отчета

Отчет о курсовом проекте должен содержать:

1. Теоретические сведения о предложенных в задании алгоритмах первичного кодирования речевых сигналов, структурные схе-

мы алгоритмов кодирования, структуру кадра передаваемой информации в канал связи.

2. Программную реализацию предложенного кодека РС в программной среде Matlab.

3. Методику оценки качества речевых сигналов, использовавшуюся при эксперименте согласно образцовой методике оценки качества РС.

4. Таблицы объективного и субъективного качества, полученные при экспериментальных испытаниях.

5. Графики и таблицы зависимостей объективной и субъективной оценок при действии различных акустических шумов от уровня шума.

6. Графики зависимости корреляции объективной и субъективной оценок при воздействии различных акустических шумов от уровня действующего шума.

7. Выводы по каждому из пунктов проделанной работы.

Контрольные вопросы

1. Наиболее распространенные стратегии кодирования речевых сигналов.
2. Основные параметры сравнения методов кодирования речи.
3. Что такое задержки кодирования системы передачи речи?
4. Что представляет собой вокодер?
5. Характеристики вокодеров.
6. Речеэлементные и параметрические вокодеры.
7. Какие типы параметров из речевого сигнала выделяют в параметрических вокодерах?
8. Классы гибридных кодеров.
9. Работа кодеков с многополосным кодированием.
10. Какие параметры передает кодер для описания сегмента входного речевого сигнала?

Библиографический список

1. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов.- М.: Радио и связь, 1985. - 176 с.
2. Шелухин О.И. Цифровая обработка и передача речи / Н. Ф. Лукьянцев; под ред. О.И. Шелухина. – М. : Радио и связь, 2000. – 456 с.
3. Кириллов С.Н., Малинин Д.Ю. Теоретические основы асинхронного маскирования речевых сигналов: учеб. пособие.- Рязань: РГРТА, 2000. – 80 с.
4. Кириллов С.Н., Дмитриев В.Т. Алгоритмы защиты речевой информации в телекоммуникационных системах: учеб. пособие.- Рязань: РГРТА, 2005.
5. Беллами Дж. Цифровая телефония. – М., 1986. - 426 с.
6. Вокодерная телефония. Методы и проблемы /под ред. А.А. Пирогова. – М.: Связь, 1974.
7. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. - 634 с.
8. Сапожков М.А., Михайлов В.Г. Вокодерная связь. – М., 1983.
9. Alan Me. Cree. A 2.4 Kbit/s Melp Coder Candidate for the new U.S. Federal Standart. Proc. ICASSP, 1996.
10. Andermo P. G. CODIT. ICUPC. Ottawa.
11. Thiede Th., Kabot E. A New Perceptual Quality Measure for Bit Rate Reduced Audio//Contribution to the 100th Convention of the Audio Engineering Society, Copenhagen, May 1996.
12. Иоффе М.Г. Автоматический контроль трактов звукового вещания. - М.: Связь, 1980.
13. Руководство но настройке и паспортизации каналов вещания. - М.: Связь, 1970.
14. Zwicker E. Psychoakustik - Springer-Verlag, Berlin Heidelberg, New York, 1982.
15. Beerends J.G., Stemerdink J.A.. A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation//J. Audio Eng. Soc/ Vol 42, No.3. P. 15-123.1994.