УДК 004.934

СОВРЕМЕННЫЕ МЕТОДЫ КОДИРОВАНИЯ РЕЧЕВОГО СИГНАЛА

М.Ф. Бондаренко[1], А.В. Работягов[2], С.В. Щепковский[3]

[1]ХНУРЭ, г. Харьков, Украина,
[2]ХНУРЭ, г. Харьков, Украина, beloswet@kture.kharkov.ua
[3]ХНУРЭ, г. Харьков, Украина, svserg@kture.kharkov.ua

В статье проведен обзорный анализ современного состояния исследований в области кодирования речевого сигнала. Рассмотрены основные применяемые и перспективные методы кодирования.

КОДИРОВАНИЕ ФОРМЫ ВОЛНЫ, ПАРАМЕТРИЧЕСКОЕ КОДИРОВАНИЕ, ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

Введение

В настоящее время в связи с активным развитием техники связи, особенно мобильной, решение задачи кодирования речевой информации имеет большое значение.

Приоритетными областями применения кодирования речевой информации являются:

1) уплотнение канала связи для обеспечения большей пропускной способности;

2) обеспечение конфиденциальности (защиты) передаваемой по техническому каналу связи речевой информации.

Первая область применения обусловлена увеличением объемов передаваемой речевой информации по каналу связи: при возрастании количества передаваемой информации (например, увеличение объемов телефонных переговоров внутри фирмы, увеличение количества входящих звонков) возникает необходимость обеспечения большой пропускной способности используемых каналов связи для поддержания информационного обмена.

Вторая область применения обусловлена тем, что одновременно с развитием радиоэлектронных средств связи развиваются и средства электронного шпионажа. Это обстоятельство ставит задачу защиты речевой информации в ряд наиболее актуальных.

Каналы утечки информации достаточно многочисленны. Они могут быть как естественными, обусловленными, например, техническим несовершенством канала связи, так и искусственными, то есть созданными с помощью технических средств. В табл. 1 рассматриваются некоторые характеристики и особенности применения основных средств радиоэлектронного шпионажа.

Отметим, что речевая информация принципиально отличается от текстовой информации. При

Таблица 1

Контролируемое устройство	Приемник информации	Место установки	Стоимость аппаратуры	Вероятность применения	Качество перехвата
Телефон	Индуктивный или контактный датчик	Телефонная линия от аппарата до АТС	Низкая	Высокая	Хорошее
Телефон	Контактный датчик	Телефонная линия от аппарата до АТС	Низкая	Низкая	Плохое
Телефон, любое устройство с питанием от сети	Радиомикрофон с передачей по телефонной сети или по сети 220В	Тел. аппарат, тел. розетка, любое устройство с питанием	Низкая	Высокая	Хорошее
Любое место в помещении	Автономные радиомикрофоны, направленные микрофоны, в т.ч. лазерные	Любое место в помещении	Высокая	Средняя	Хорошее
Радиотелефон, радиостанция	Панорамный радиоприемник	Прием с эфира	Средняя	Высокая	Хорошее
Сотовый телефон	Устройство прослушивания сотовой сети	Прием с эфира	Высокая	Высокая	Хорошее
Монитор ПК	Широкополосная антенна	Прием с эфира	Очень высокая	Низкая	Посредственное
Монитор ПК	Широкополосный контактный датчик	Питающая электросеть	Очень высокая	Низкая	Посредственное
Магистраль компьютерной сети	Индуктивный или контактный датчик	Кабель магистрали	Высокая	Высокая	Хорошее

шифровании текста используют определенный ограниченный набор символов, поэтому при работе с текстом можно использовать такие шифры, как шифры перестановки, шифры замены, шифры взбивания и так далее. Речь же нельзя (во всяком случае, на современном уровне развития технологии распознавания речи) представить набором каких-либо символов. Поэтому применяются иные принципы кодирования.

При кодировании речевой информации ставят задачу максимального сжатия речевого сигнала [1]. Существует предел сжатия речевой информации. Теорема Грея-Бергера определяет предел сжатия последовательностей по минимально достижимой среднеквадратической ошибке. В соответствии с данной теоремой этот предел составляет 100 бит/с. Эффективность (качество) кодирования оценивается по субъективному восприятию речи.

Качество сигнала измеряется часто по пятибалльной шкале MOS (*mean opinion score* – средняя субъективная оценка) (см. табл. 2). Оценка по шкале MOS определяется путем обработки усредненных оценок, даваемых группами слушателей нескольким речевым сигналам, воспроизводимым различными громкоговорителями.

Каждый слушатель дает оценку для каждого сигнала.

Таблица 2

1	2	3	4	5
Плохо	Слабо	Разборчиво	Хорошо	Отлично

Основные усилия в разработке методов и систем цифровой связи осуществляют высокоразвитые страны – США, Япония, Франция, Германия и Англия. Конечные результаты исследований формулируются в виде национальных стандартов и рекомендаций Международного института телекоммуникаций ITU (International Telecommunication Union).

1. Методы кодирования речевого сигнала

Современные методы кодирования речевого сигнала можно классифицировать по принципам кодирования, которые положены в основу метода. Различают следующие принципы кодирования речевого сигнала:

– принцип кодирования формы волны речевого сигнала;

– принцип кодирования параметров речевого тракта человека и источника возбуждения;

– принцип кодирования символьной информации (фонем);

– принцип кодирования лингвистической информации (слов, фраз и тому подобное).

Данные принципы различаются в зависимости от используемых основных свойств речевого сигнала, образования и восприятия речи:

1) основные свойства речевого сигнала: амплитуда, частота (время), площадь (как производная амплитуды и времени);

2) свойства образования речи:

– изменения амплитуды,

– деление речи на звуки, паузы и шумы,

– особенности языковой и фонетической структуры,

– кратковременная корреляция (особенности формантной структуры),

– особенности структуры тона звука (звонкие звуки),

– особенности структуры шума (глухие звуки).

3) свойства восприятия речи:

4) локальный спектральный динамический диапазон речевого сигнала,

5) слуховое маскирование.

На современном этапе используются, в основном, методы, которые базируются на первых двух принципах кодирования.

Принцип кодирования формы волны речевого сигнала. Первым шагом кодирования является измерение значения амплитуды сигнала. Для этого 12-14-ти разрядный динамический диапазон амплитуды разбивают на 8 логарифмических поддиапазонов, в каждом из которых значение амплитуды кодируют 5 разрядами и таким образом достигают сокращения информации до 64000 бит/с (кодирование по μ- и A- законам в соответствии со стандартом ITU G.711) [2]. Следующим шагом является адаптивная дифференциальная импульсно-кодовая модуляция (*adaptive differential pulse code modulation* – ADPCM), (например, в соответствии со стандартами G.721 или G.726), с помощью которой осуществляют кодирование (аппроксимацию) степени приращения амплитуды сигнала во времени. Таким путем удается достичь степени сжатия речевого сигнала порядка 32000-16000 бит/с, причем приемлемое (коммерческое) качество речи (по критерию отношения: полезный сигнал/шум) обеспечивается до 24000 бит/с. При более низких скоростях кодирования сохраняется разборчивость речи, но характерны сильные нелинейные и частотные искажения сигнала и ухудшение отношения сигнал/шум. Дальнейшее уменьшение информационной емкости сигнала с помощью данного подхода считается неэффективным.

Принципы параметрического кодирования. Низкоскоростное кодирование складывается из двух основных процессов:

– параметрическое представление речевого сигнала минимальным набором параметров, характеризующих источник возбуждения и акустический артикуляторный фильтр;

– дискретизация речевых параметров для их передачи по каналу связи при использовании минимальной емкости канала.

Для параметрического описания речи обычно используется подход, основанный на вычислении параметров, описывающих передаточную функцию речевого тракта человека и функцию возбуждения. Такими параметрами, например, являются: коэффициенты линейного предсказания (модель

авторегрессии) и связанные с ними коэффициенты отражения или отношение площадей поперечного сечения смежных акустических резонаторов в соответствии с моделью речевого тракта человека, представленной системой акустических резонаторов. В последнее время наибольшее распространение получил метод, позволяющий вычислять непосредственно полюса передаточной функции речевого тракта в частотной области, упорядоченные по возрастанию частоты (*liner spectral frequency* – LSF). Обычно для кодирования речи используются 8-10 параметров (один из вышеперечисленных наборов), вычисляемых на интервалах порядка 5-40 мс. Кроме того, вычисляется параметр, характеризующий изменение амплитуды либо мощности сигнала, период основного тона речи, а также признак типа тон/шум/пауза, характеризующий способ возбуждения речевого сигнала.

В качестве функции возбуждения речевого сигнала используется дельта функция.

Полученный набор параметров, оптимизированный по критерию точности и минимальной разрядности представления, передается в цифровом виде по каналу связи в реальном времени; на приемном конце осуществляется синтез речевого сигнала по перечисленным параметрам. Таким путем удается снизить информационную емкость речевого сигнала до уровня 16000-1200 бит/с, причем с сохранением разборчивости и индивидуальных особенностей речи говорящего.

На рис. 1 представлены результаты тестирования фирмой AT&T различных систем цифровой связи на основе субъективного восприятия качества речи, передаваемой по цифровым каналам связи [3].

В табл. 3 указаны основные стандарты кодирования.

Как следует из графика, качество некоторых систем параметрического кодирования приближается к качеству ADPCM. Подчеркнем, что такой подход позволил снизить информационную емкость речевого сигнала с исходных 96000 бит/с до 2400 бит/с, то есть в 40 раз. При этом сохраняется не только разборчивость речи, но и индивидуальные особенности речи человека.

Следующим шагом в направлении дальнейшего увеличения сжатия речевого сигнала является создание фонемного вокодера. Как известно, минимальной слогоразличительной (и словоразличительной) единицей речи является фонема. Ожидается, что создание устойчивого метода распознавания фонем позволит снизить скорость передачи данных до 100 бит/с, что соответствует информационной скорости передачи текста. Заметим, что на приемном конце речь будет восстановлена синтезатором речи по фонемному тексту, при этом информация об индивидуальности диктора будет утрачена.

И наконец, последним этапом совершенствования систем кодирования может явиться создание системы на основе автоматического распознавания слов и, возможно, целых фраз. В этом случае по

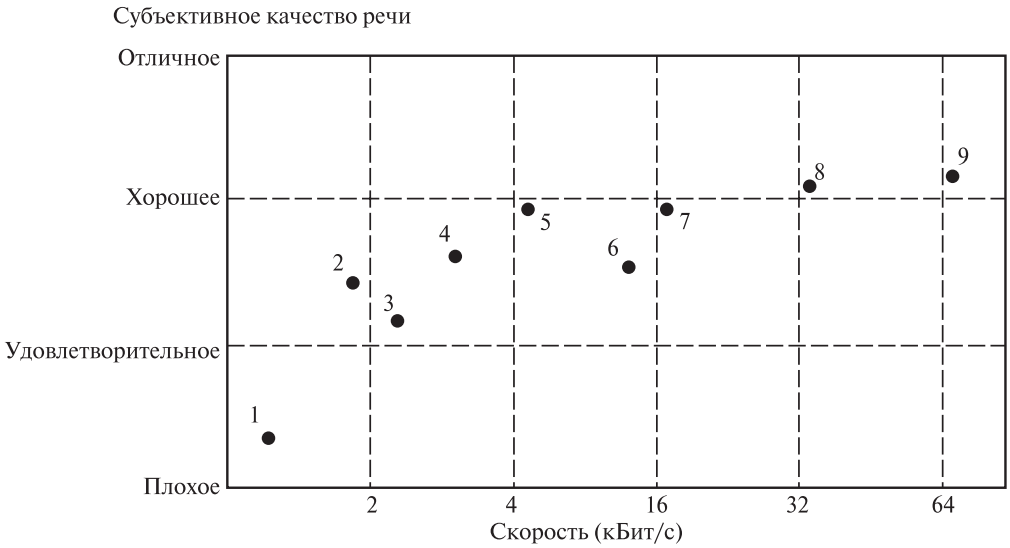

Рис. 1. Субъективное восприятие качества речи

Таблица 3

№ пп	Стандарт кодирования	Организация	Тип кодирования	Скорость Кбит/с
1	FS1015	Secure Voice Communication USA	LSF	2,4
2	JDC-HR	Japanese Cecular	PSI-CELP	3,67
3	FS1016	Secure Voice Communication USA	CELP	4,8
4	GSM-HR	Eropean Cecular	RPE-LTP	7,8
5	ITU	ITU-T	VCELP	8
6	GSM	Eropean Cecular	RPE-LTP	13
7	G.728	ITU-T	LD-CELP	16
8	G.726	ITU-T	ADCPM	32
9	G.711 m, A закон	ITU-T	PCM	64

каналу связи может быть передан только его код, а на приемном конце будет получен номер слова из некоторого ограниченного словаря и с помощью синтезатора преобразован в речевой сигнал.

Рассмотрим основные методы кодирования речевого сигнала.

1.1. Базовое преобразование речевого сигнала – метод импульсно-кодовой модуляции

При разработке практически любой технической системы кодирования необходимо представить речевой сигнал в цифровом виде. Этот процесс осуществляется на основе метода импульсно-кодовой модуляции (pulse code modulation – PCM).

Процесс базируется на теореме Котельникова (Найквиста), в соответствии с которой цифровой сигнал, полученный выборкой с частотой в два раза выше максимальной частоты аналогового сигнала, с помощью интерполяции обратно преобразуется в аналоговую форму. Человеческая речь воспроизводится с приемлемым качеством в полосе частот 100-4000 Гц, чему соответствует частота выборки 8 кГц (8000 отсчетов в секунду), каждый отсчет преобразуется в 8-битовый цифровой код. Общая скорость цифрового потока PCM-сигнала равна 8×8000 отсчетов в секунду, то есть 64 кбит/с.

Метод импульсно-кодовой модуляции принят в 1960 г. в качестве международного стандарта кодирования речи для телефонного канала (стандарт ITU G.711), работающего на скорости 64 кбит/с. Этот алгоритм используется при передаче речевой информации в коммерческих телефонных сетях. Оцифровка голосового сигнала включает измерение уровня аналогового сигнала через равные промежутки времени. В соответствии со стандартом G.711 необходимо обеспечить передачу частотных составляющих речевого сигнала в диапазоне от 200 до 3400 Гц. Стандарт обеспечивает неискаженную передачу сигнала в полосе человеческого голоса (до 4 кГц) с отношением сигнал/шум 40 дб.

Хотя PCM-сигнал со скоростью 64 кбит/с и гарантирует качество речи аналогового телефонного сигнала, ограниченная общая ширина канала, особенно в спутниковых и радиочастотных системах, вынуждает снижать скорость битовых потоков, отводимых для каждого речевого сигнала. С этой точки зрения весьма эффективны алгоритмы сжатия речи, дополняющие PCM-кодирование математическими функциями, такими как фильтры, квантизаторы и предсказатели. Они преобразуют PCM-сигнал так, чтобы передавать его более эффективным способом, обеспечивая тем самым не менее точное воспроизведение сигнала на приемном конце. Соответствующие устройства называют кодерами (при прямом преобразовании), декодерами (при обратном преобразовании) или кодеками (другое название – вокодер VOice CODer). Основной технической характеристикой кодеков является скорость передачи кодированного речевого сигнала по каналу связи, которая измеряется в бодах (бит/с). Скорость обработки измеряется в миллионах инструкций в секунду (*millions of instructions per second* – mips).

1.2. Метод нелинейного квантования

Метод нелинейного квантования основан на уменьшении числа уровней квантования по амплитуде. На рис. 2 представлен сигнал, который квантуется по времени в 11 точках с использованием 8 уровней квантования по амплитуде сигнала.

Уровни квантования (по амплитуде) кодируются по правилу, представленному в табл. 4:

Номер уровня квантования преобразован в код с 3 битами.

Таблица 4

Кодирование уровней квантования для несжатого сигнала

Уровень квантования	0	1	2	3	4	5	6	7
Код	000	001	010	011	100	101	110	111

Сигнал кодируется следующим образом: 101 111 110 001 010 100 111 100 011 010 101. Общее количество – 33 бита.

Уменьшение числа уровней квантования по амплитуде состоит в том, чтобы соединить каждые два соседних уровня в один. Результатом такого преобразования является только 4 уровня квантования вместо 8. Они могут быть закодированы с использованием только кодов с 2 битами (см. табл. 5).

Таблица 5

Кодирование уровней квантования для сжатого сигнала

Уровень	0	1	2	3
Код	00	01	10	11

Восстановленный сигнал после сжатия имеет те же самые базисные контуры, но искажения в таком

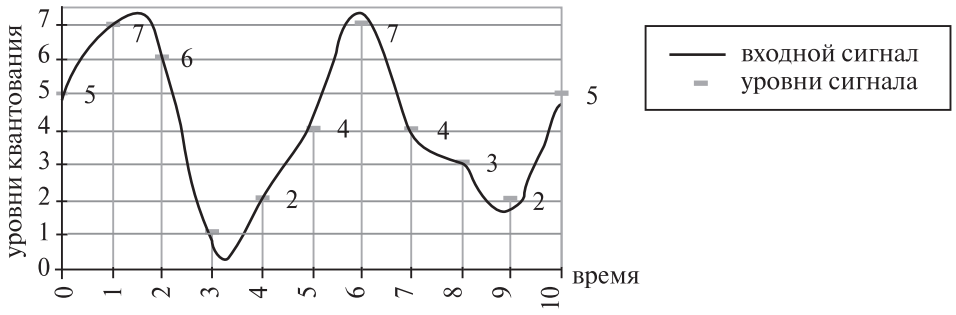

Рис. 2. Исходный несжатый сигнал

сигнале больше из-за грубого приближения, так как шум квантования увеличился. Это – следствие того, что шаг квантования стал двойным по сравнению с первым случаем (несжатым сигналом).

Сигнал закодирован следующим образом (см. рис. 3): 10 11 11 00 01 10 11 10 01 01 10, общее количество 22 бита (11 отсчетов, 2 бита на отсчет), коэффициент сжатия – 1,5:1 (уменьшение 33% в размере). Уровни квантования неравномерно распределены внутри диапазона квантования – они плотнее вблизи уровня нуля и реже вблизи максимального уровня.

Стандарт кодирования, применяемый на практике с коэффициентом сжатия 2:1, получил название A-law.

1.3. Метод кодирования на основе уменьшения числа выборок

Метод основан на уменьшении числа выборок квантования по времени.

Например можно заменить каждые две соседних выборки только на одну, равную их среднему значению. Плотность выборок становится меньше, чем прежде (число выборок уменьшается в два раза). В результате восстановленный по выборкам сигнал может быть существенно искажен за счет потери части гармоник.

При использовании этого метода необходимо отобрать те частоты (гармоники), которые не имеют существенного значения и могут не учитываться, не внося существенной погрешности. Сигнал закодирован следующим образом (см. рис. 4): общее количество битов 18 (6 отсчетов, 3 бита на отсчет), коэффициент сжатия – 1,8:1 (уменьшение 45,5 % в размере).

1.4. Метод адаптивно-дифференциальной импульсно-кодовой модуляции

Прямое аналого-цифровое преобразование (PCM-преобразование) является низкоэффективным, но 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.

Значительные результаты в области эффективного кодирования речи достигнуты на базе общего подхода «кодирования с предсказанием». Большая часть стандартизированных ITU алгоритмов кодирования относится именно к этому направлению.

Метод адаптивной дифференциальной импульсно-кодовой модуляции (ADPCM) принят в качестве стандарта в 1984 г. под названием G.726. Он воспроизводит речь почти с такой же субъективной оценкой качества, как и PCM, используя только 32 кбит/с, и обеспечивает на порядок более высокую помехоустойчивость. Однако он теряет работоспособность при вероятности одиночной ошибки, составляющей около 5×10^{-3}, и передаче пакетов ошибок малой длительности.

Метод основан на том обстоятельстве, что в аналоговом речевом сигнале 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. Таким образом, ADPCM снижает скорость битового потока вдвое путем обработки разности между двумя соседними отсчетами, а не самих отсчетов. Позволяет снизить скорость с 64 кбит/с до 24-48 кбит/с (в зависимости от того, насколько точно передается приращение). Приводит к снижению отношения сигнал/шум и менее точному воспроизведению исходного сигнала.

Среди кодеров формы сигнала первыми появились методы адаптивной дельта-модуляции. Анали-

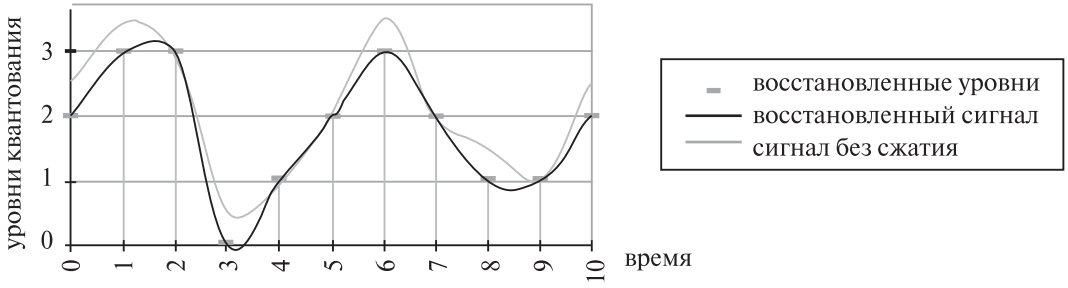

Рис. 3. Сжатый (закодированный) сигнал – уменьшение числа уровней квантования

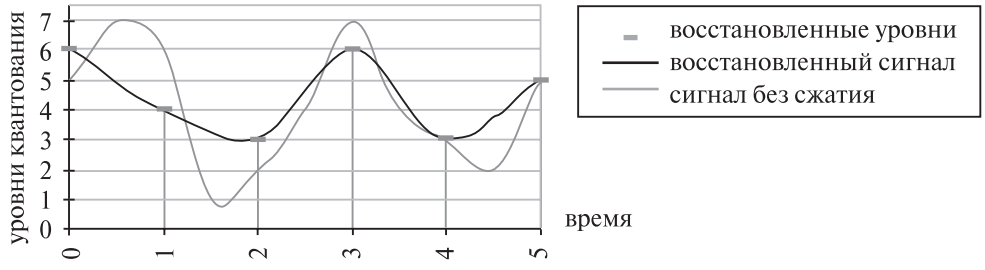

Рис. 4. Сжатый сигнал (уменьшение числа выборок)

тически они являются предельными случаями дифференциальной PCM, 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 этот метод обеспечивает более высокое качество восстановления, чем PCM. Дельта-модуляция обладает наилучшими параметрами помехоустойчивости среди всех методов кодирования. Соответствующие системы не теряют работоспособности при возникновении одиночных ошибок и их пакетов малой длительности.

Принципиальное отличие ADPCM от PCM заключается в использовании адаптивного аналого-цифрового преобразователя (АЦП) и дифференциального кодирования. Адаптивный АЦП отличается от стандартного PCM-преобразователя тем, что в любой момент времени уровни квантования расположены однородно (а не логарифмически), причем шаг квантования меняется в зависимости от уровня сигнала. Применение адаптивного метода базируется на том, что в речи последовательные уровни сигнала не являются независимыми. Поэтому, преобразуя и передавая лишь разницу между предсказанным и реальным значением, можно заметно снизить загрузку линии, а также требования к широкополосности канала. Следует иметь в виду, что метод не лишен серьезных недостатков: уровень шумов, связанный с квантованием сигнала, выше, а при резких изменениях уровня сигнала, превышающих диапазон АЦП, возможны серьезные искажения.

PCM и ADPCM – методы кодирования волновой функции речевого сигнала. Это означает, что они рассматривают входной речевой сигнал как чисто аналоговый. Однако для получения высокого качества сигнала при скоростях ниже 32 кбит/с такое кодирование неэффективно.

1.5. Метод кодирования с линейным предсказанием

При кодировании с линейным предсказанием (*linear predictive coding* – LPC) моделируются различные параметры речи, которые передаются вместо отсчетов или их разности, требующих значительно большей пропускной способности канала. LPC работает с блоками отсчетов, а не с отдельными отсчетами, как PCM или ADPCM. Для каждого блока алгоритм LPC вычисляет и передает частоту основного тона, его амплитуду, «флаг» речевого или неречевого происхождения сигнала и другие параметры.

При таком подходе к кодированию речи, во-первых, возрастают требования к вычислительным мощностям микропроцессоров, используемых для обработки сигнала, а во-вторых, увеличивается задержка при передаче, поскольку кодирование применяется не к отдельным значениям, а к некоторому их набору, который перед началом преобразования следует накопить в определенном буфере. Подчеркнем, что задержка в передаче речи при использовании этого метода связана не только с необходимостью обработки цифрового сигнала (эту задержку можно уменьшать, увеличивая мощность процессора), но и непосредственно следует из специфики метода.

Этот метод позволяет достигать очень больших степеней сжатия, которым соответствует полоса пропускания 2,4 или 4,8 кбит/с, однако, при этом качество звука получается низким. Поэтому в коммерческих приложениях он не используется, а применяется, в основном, для ведения служебных переговоров.

Более сложные алгоритмы на базе LPC комбинируют LPC с элементами кодирования звуковой волны. Эти алгоритмы используют замкнутый LPC-кодер (называемый также «анализ через синтез» – *analysis-by-synthesis* – AbS), в котором при передаче сигнала осуществляется оптимизация кода. Ее выполняет алгоритм, который находит наилучшую аппроксимацию каждого речевого сегмента. Закодировав сигнал, процессор пытается восстановить его форму и сличает результат с исходным сигналом, после чего начинает варьировать параметры кодировки, добиваясь наилучшего совпадения. Для использования такого метода требуются большие вычислительные мощности.

Вариант замкнутых LPC-алгоритмов – это метод линейного предсказания с кодовым возбуждением (*code-excited linear prediction* – CELP), метод регулярного импульсного возбуждения (*Regular Pulse Excitation* – RPE), используемый в европейских сотовых системах при скорости 13,2 кбит/с и метод LD-CELP с низкой задержкой (*low delay* CELP). Для кодирования речевых сигналов, например, в стандарте TETRA используется кодер с линейным предсказанием и многоимпульсным возбуждением от кода – CELP со скоростью преобразования 4,8 кбит/с. Данный метод кодирования основан на линейной авторегрессионной модели процесса формирования и восприятия речи и входит в группу так называемых методов анализа через синтез, реализующих современные и эффективные алгоритмы кодирования 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.

LD-CELP принят ITU в 1992 г. как стандарт кодирования речи G.728 на скорости 16 кбит/с. Алгоритм LD-CELP применяется к последовательности цифр, получаемых в результате аналого-цифрового преобразования речевого сигнала с 16-разрядным разрешением. Пять последовательных цифровых значений кодируются одним 10-битовым блоком, что позволяет достигнуть скорости 16 кбит/с. Для применения этого метода требуются большие вычислительные мощности, в частности, для прямо-

линейной реализации G.728 необходим процессор с быстродействием 44 mips [4].

1.6. Методы кодирования с предсказанием

Широкое распространение для различных приложений получило и множество нестандартных методов кодирования. В частности: варианты адаптивного кодирования с предсказанием (*adaptive predictive coding* – APC), разработанные в лабораториях компании Bell; метод линейного предсказания с векторным возбуждением (*vector-sum-excited linear prediction* – VSELP), предложенный фирмой Motorola в качестве стандарта для цифровых сотовых систем США, работающих на скорости 8 кбит/с; метод линейного предсказания с предикативным кодовым возбуждением (*predictive code-excited linear prediction* – PCELP), созданный DSP Group в 1992 г.

Эти высокоэффективные кодеры обеспечивают отличное качество звука при низких скоростях (2,4-8 кбит/с). Для кодирования погрешности предсказания в них используются так называемые «кодовые книги», состоящие из блоков с конечным числом символов. Эти методы описаны стандартом G.729, в котором вместо оцифрованного сигнала передаются номера выборок из хранящейся в памяти «кодовой книги», где описаны типичные элементы человеческого голоса.

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 другие) указывают на способ реализации предсказателя, синтеза квантователя или кодовой книги.

Стандарт G.729A (G.729 Annex A) – это алгоритм сжатия звука преимущественно для передачи речи, генерирующий кадры (фреймы) длительностью 10 мс. Принцип G.729A похож на G.729, но требует меньше вычислений. Это упрощение приводит к незначительному ухудшению качества речи.

Особенности этого стандарта кодирования:

– частота сэмплирования 8 кГц/16 бит (80 сэмплов в 10-ти миллисекундном фрейме);

– фиксированный битрейт (8 кбит/с 10-тимиллисекундные фреймы);

– фиксированный размер фрейма (10 байт для 10-тимиллисекундного фрейма);

– алгоритмическая задержка 15 мс на фрейм, с 5-тимиллисекундной предварительной задержкой;

– сложность алгоритма оценивается в 15 баллов, используя относительную оценочную шкалу, где G.711 равен 1 и G.723.1 равен 25 баллам;

– условно-объективная оценка MOS по методу экспертных оценок, протестированная в идеальных условиях, дает результат 4,04 для G.729A, по сравнению с 4,45 для G.711 u-law;

– условно-объективная оценка MOS по методу экспертных оценок, протестированная в перегруженной сетевой среде, дает результат 3,51 для G.729A, по сравнению с 4,13 для G.711 μ-law.

1.7. Метод множественной импульсной многоуровневой квантизации

В марте 1995 г. ITU выбрал метод сжатия речи для своих будущих стандартов в области мультимедиа и видеотелефонов, подключаемых к коммутируемым телефонным сетям. Стандарт сжатия G.723 частично базируется на новом методе сжатия речи – множественной импульсной многоуровневой квантизации (*Multipulse Maximum Likelihood Quantization* – MP-MLQ), разработанном израильской фирмой AudioCodes, создателем передовых речевых и факсимильных технологий и ее корпоративным партнером – американской фирмой DSP Group.

Метод MP-MLQ относится к семейству алгоритмов AbS. Речевой кодер MP-MLQ использует LPC-анализатор 10-го порядка и работает на скоростях 6,4; 7,2 и 8,0 кбит/с. Его структура поддерживает перепрограммирование «на лету» для одной или нескольких скоростей. Масштабируемость алгоритма MP-MLQ позволяет разрабатывать производные реализации для скоростей вплоть до 4,0 кбит/с и более низких коммуникационных задержек (до 20 мс), осуществлять кодирование на нескольких скоростях и с переменной скоростью, выполнять многоканальную обработку (благодаря низкой вычислительной нагрузке – менее 10 mips) и достигать высокого качества.

В отличие от других кодеров с низкими скоростями, метод обеспечивает минимальный уровень искажений при парном кодировании, когда речевой сигнал проходит через два или более последовательных цикла компрессии/декомпрессии. Эта особенность имеет практическое значение в приложениях, в которых сеанс речевой связи в цифровом канале коммутируется через центральную АТС. Тесты, проведенные в фирмах AT&T Labs и France Telecom (CNET), показали, что оценка качества сигнала по шкале MOS после двух кодирований в тандеме методом MP-MLQ составила 3,409, что лучше оценки G.726 ADPCM на 32 кбит/с после четырех кодирований в тандеме (3,102) и почти эквивалентно после двух кодирований в тандеме (3,491).

2. Достигнутые результаты

В настоящее время считается достигнутой скорость передачи кодированного речевого сигнала 2400 бит/с (американский федеральный стандарт FS1015). При этом кроме разборчивости от вокодера требуется обеспечение узнаваемости говорящего, возможность определения его эмоционального состояния и тому подобное. Такие требования заставляют разрабатывать методики испытаний, отличающиеся от обычного тестирования на разборчивость. Разработка таких методик является нетривиальной задачей.

Следует заметить, что для реализации алгоритмов кодирования речевых сигналов в реальном масштабе времени требуются процессоры с производительностью в 15-20 mips.

Исследовательские работы, проводимые в этих направлениях Санкт-Петербургским университетом, позволили в настоящее время создать экспериментальные алгоритмы, программы и устройства, обеспечивающие:

– кодирование/декодирование речевого сигнала на скоростях 3600, 2400 и 1200 бит/с;

– синтеза речи по произвольному тексту.

Системы кодирования и передачи речи могут быть выполнены с использованием как стандартных технических средств, ставших неотъемлемой частью современных персональных компьютеров: средства мультимедиа (звуковые карты) и модемы, так и на базе специальных сигнальных процессоров серии ADSP-21XX.

Экспериментальный макет кодирования/декодирования речи обеспечивает компрессию речевых сигналов в реальном масштабе времени, а затем после передачи по каналу связи восстанавливает речь с сохранением разборчивости и индивидуальных особенностей диктора на компьютерах даже такого уровня, как IBM PC 486 DX4-100 без использования спецвычислителей.

Результаты испытаний экспериментального компьютерного вокодера для русскоязычных дикторов приведены в табл. 6.

Таблица 6

Результаты испытаний разборчивости по ГОСТ-В 20775-75

Скорость (бит/с)	1200	2400	3600
Разборчивость слогов	89%	91	94%

В настоящее время завершается отладка ПО, обеспечивающего разборчивость слов на скорости 1200 бит/с не хуже 93-94 % при передаче речи в дуплексном режиме по линии связи.

Ведутся исследования по созданию алгоритмов, которые должны обеспечить кодирование речевого сигнала на скорости 600 бит/с с разборчивостью не менее 0,90.

Существующий экспериментальный макет синтезатора речи обеспечивает разборчивость синтезируемого сигнала не хуже 95% при неограниченном объеме словаря.

3. Перспективы кодирования речевой информации

Одним из перспективных направлений уменьшения информационной емкости систем кодирования речи является использование моделей речеобразования, которые не предполагают независимое функционирование источника возбуждения звука при речеобразовании и акустического фильтра, формирующего лингвистическое качество речевого сигнала. Это направление в настоящий момент развивается по линии решения обратной задачи, то есть определения формы арктикуляторного тракта (так называемые функции площади) и характеристик источника возбуждения звука по речевому сигналу. В отличие от классического параметрического описания речевого сигнала при таком подходе этот сигнал будет характеризоваться (опосредованно) с помощью медленно изменяющихся параметров артикуляторного тракта. Кроме того, характеристики взаимодействия источника звука и акустического фильтра в этом случае автоматически учитываются (в противоположность к предположению о независимости акустического фильтра и источника звука, которое привело к основным трудностям в совершенствовании малобитных и высококачественных вокодеров в последние 15 лет).

Есть еще одно преимущество при таком подходе. В случае использования в качестве параметрического описания коэффициентов линейного прогноза при интерполяции могут возникать неустойчивые состояния. В случае физического моделирования медленно меняющейся формы артикуляторного тракта такое невозможно.

В Московском техническом университете связи и информатики разработаны новые методы эффективного кодирования речи в классе линейного предсказания с анализом через синтез (ЛПАС), основанные на синтезе сигнала погрешности предсказания ортогональными полиномами Чебышева как в спектральной, так и во временной областях [5]. На скоростях 8-16 кбит/с данные методы кодирования обеспечивают первый класс качества по разборчивости в соответствии с ГОСТ Р 51061-97. Повышенное качество синтеза речи здесь обеспечивается за счет оптимизации (по критерию максимума отношения сигнал/суммарная погрешность синтеза) метода рекуррентной (со взвешиванием) оценки коэффициентов линейного предсказания для нестационарной речи. Получены оптимальные оценки множителя забывания.

Разработан новый метод низкоскоростного кодирования речи в классе ЛПАС на основе модели голосового возбуждения с динамической частотно-импульсной модуляцией (ДЧИМ). На скоростях 3,5-4,0 кбит/с метод обеспечивает первый класс качества по разборчивости в соответствии с ГОСТ Р 51061-97. По сравнению с методом ADPCM, здесь достигается сжатие цифрового представления речи в 8 и более раз.

Перспективным направлением развития систем кодирования речи представляется направление, связанное с моделью речи, предложенной К. Стивенсом. Центральной идеей этой модели является представление о том, что информационное ядро речевого сообщения привязано к ограниченным речевым участкам этого сообщения. Остальные временные отрезки являются только неким наполнением, связанным с физическими ограничениями артикуляторного тракта. Практически на этой идее основан подход, который получил название *waveform interpolation*. В этом случае в речевой волне вы-

деляются и передаются в канал только отдельные периоды основного тона, а остальные при декодировании вычисляются методом декодирования.

На аналогичной идее основан еще один интересный подход (*temporal decomposition*) к проблеме кодирования. При этом предполагается, что набор спектральных параметров может быть представлен линейной комбинацией, набора перекрывающихся компактных функций.

В последнее десятилетие в мире возникло и оформилось новое научное направление, связанное с так называемым *вейвлет-преобразованием*. Слово «wavelet», являющееся переводом с французского «ondelette», означает «небольшие волны, следующие друг за другом».

Вейвлет-преобразование сигналов обеспечивает возможность весьма эффективного сжатия сигналов (не только речевых) и их восстановления с малыми потерями информации, а также решение задач фильтрации сигналов. Основная идея вейвлет-преобразования состоит в представлении некоторой случайной функции (исследуемого сигнала) как суперпозиции определенных базисных негармонических функций – вейвлетов (см. рис. 5).

Рис. 5. Базовые вейвлет-функции

Для того чтобы вейвлеты хорошо аппроксимировали исходный сигнал, они подвергаются масштабированию (сжатию или растяжению) и сдвигу (смещению). Результат вейвлет-преобразования – обычный массив числовых коэффициентов. Такая форма представления информации о сигнале очень удобна, поскольку числовые данные легко обрабатывать.

После масштабирования выполняется этап порогового преобразования. На этом этапе отбрасываются коэффициенты, значение которых близко к нулю. Следует помнить, что при этом происходит необратимая потеря информации, так как «отброшенные» коэффициенты участвуют в формировании (восстановлении) сигнала. Поэтому выбранное пороговое значение коэффициентов сильно влияет на качество сигнала – задание слишком высокого порога повлечет за собой снижение качества.

Уникальные свойства вейвлетов позволяют построить базис, в котором данные будут представлены всего лишь несколькими ненулевыми коэффициентами. Это означает, что массив коэффициентов можно сильно сжать обычными методами без потери информации.

Заключение

В ближайшее время можно ожидать появление систем компрессии до 600-900 бит/с. Скорее всего, они будут опираться на интерполяцию речевого сигнала во временной области. Возможность сохранения индивидуальных особенностей в этом случае весьма сомнительна. Во всяком случае, эта сторона проблемы требует дополнительного изучения, что потребует разработки специальных методик.

В настоящее время сотрудниками лаборатории исследования речи Харьковского национального университета радиоэлектроники разрабатывается новый метод кодирования речевого сигнала, который имеет принципиальные отличия от методов, описанных в статье. В основе метода лежит бионический принцип распознавания речи, в частности, продолжает свое развитие структурное (геометрическое) направление исследования речевого сигнала [6].

Список литературы: 1. *Александр Крейнес*. Как налить море в наперсток? Технологии компрессии голоса // Сети и системы связи. – 1996. – № 9-10. – С. 119-121. 2. http://www.itu.ch/itudoc/itu-t/rec/g/g700-799.html, ITU Coding Standards. 3. *О. Варламова*. Помехоустойчивые кодеки – будущее цифровой телефонии // Сети и системы связи. – 1997. – № 10. – С. 26-32. 4. ITU Recommendation G.728. Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction, November 1994. 5. *Медведев О. Н.* Автореф. дисс. на соискание ученой степени канд. техн. наук. – Москва: Технический университет связи и информатики, 2007. – 19 с. 6. *Бондаренко М.Ф., Дрюченко А.Я., Шабанов-Кушнаренко Ю.П.* Гласные звуки в теории и эксперименте. – Харьков: ХНУРЭ, 2002. – 348 с.

Поступила в редколлегию 2.10.2008

УДК 004.934

Сучасні методи кодування мовного сигналу / М.Ф. Бондаренко, А.В. Работягов, С.В. Щепковський // Біоніка інтелекту: наук.-техн. журнал – 2008. – № 2 (69). – С. 106-114.

Аналізуються сучасні методи кодування мовного сигналу. Розглянуті методи, що вже мають статус міжнародних стандартів, а також перспективні методи, які знаходяться у стадії впровадження і розвитку. В результаті проведеного дослідження зроблений висновок щодо можливості створення найближчим часом методів з більш високим ступенем компресії мови.

Табл.: 6. Іл.: 5. Бібліогр.: 6 найм.

UDC 004.934

Modern methods of encoding of speech signal /M. Bondarenko, A. Robotyagov, S. Schepkovsky// Bionics of Intelligence: Sci. Mag. – 2008. – № 2 (69). – P. 106-114.

The modern methods of encoding of speech signal are analysed. Methods, already having status of international standards, and also perspective methods which are in the stage of introduction development, are considered. As a result of the conducted research a conclusion is done about possibility of creation in the near time methods with more high degree of compression of speech.

Tab.: 6. Fig.: 5. Ref.: 6 items.