

Из истории исследований преобразования речи

Часть 2

В.Г. Михайлов,

доктор филологических наук

IV этап (после 1987 года)

Этот этап характеризуется интенсивным развитием новых информационных технологий, включая вокодеры в сети Интернет (IP-телефония), ведомственной телефонной связи с защитой информации TETRA и мобильной телефонной связи — МТС. Обзор применённых методов и соответствующих технологий дан в работах О.И. Шелухина, Н.Ф. Лукьянцева (2000) и В.Г. Михайлова (2002). Приведём описание некоторых вокодерных алгоритмов.

IP-телефония

История развития IP-телефонии насчитывает около пяти лет. В 1994 г. во время полёта космического челнока Endeavor агентство США NASA передало по сети Internet изображение корабля и голос космонавта. В следующем году несколько зарубежных фирм предложили кодеки для речевой связи в Internet (один из первых — Internet Phone фирмы VocalTec, Израиль). Наконец, в 1996г. фирма Dialogic (США) совместно с упомянутой фирмой разработал шлюз — устройство сопряжения абонентов сети общего пользования с коммутацией каналов ТфОП с пользователями сети с коммутацией пакетов Internet и тем самым положили начало голосовой связи в последней — IP-телефонии¹ [Михайлов 2002]. Вскоре появились шлюзы фирм Vienna Systems, Inter-Tel, Cisco и др. Первые шлюзы, которые использовались в России, были изготовлены фирмой Dialogic, а установку шлюзов выполняли фирмы Tario, RGC, Comptek и др. Фирма Tario к 1999 г. установила оборудование в 30 городах России и СНГ и через шлюз в США обеспечила выход пользователей VoIP на все страны мира. Фирма RGC имеет собственные серверы в Москве, Санкт-Петербурге, Владивостоке, Хабаровске, Новосибирске, Южно-Сахалинске, Нью-Йорке, Мюнхене, Берлине.

¹ IP-Internet Protocol; VoIP — IP-телефония; ISTP — Internet Short Time Providers (провайдеры сети IP-телефонии с малым временем задержки); ITSP — Internet Telephony Service Providers (провайдеры сети IP-телефонии с предоставлением телефонных услуг).



Сеть IP-телефонии обслуживает передачу мультимедийных приложений. К ним относятся речь, ФМ, передача данных и видео.

Оборудование сети IP-телефонии обеспечивает:

1. Кодирование, компрессию и упаковку речевого сигнала в IP- пакеты;
2. Управление потоками IP-пакетов в сети Internet;
3. Интерфейс с телефонными сетями общего пользования ТфОП и сетями подвижной связи, включая сотовые сети.

Эти функции реализованы в виде плат и программного обеспечения для персонального компьютера пользователя (звуковая карта, кодек) и оборудования провайдера (шлюзы, маршрутизаторы и пр.). В 1998 г. Госкомсвязь России официально отнёс услуги IP-телефонии к телематической службе пакетных голосовых соединений. Уже выдано около 700 лицензий отечественным и зарубежным фирмам на развитие и внедрение этой службы (фирмы **CISCO Systems, RGC** (США), **VocalTEC** (Израиль) и др.). В 1999 г. были провайдеры в 30 городах России и СНГ, которые обеспечивали выход по сети IP-телефонии на многие страны мира. К их числу относятся провайдеры сети Tario.Net (оборудование фирмы DM3), объединяющей 55 городов России и стран СНГ и имеющей выход на 237 стран мира. ЗАО «Корпорация OCC» имеет собственные серверы в Москве, С-Петербурге, Нью-Йорке и более 50 узлов на территории СНГ, обеспечивает соединения через сети ITXC, Teleglobe по всему миру. Фирма RGC развивает глобальную сеть на базе оборудования фирм Cisco, VGW, имеет серверы в России, США, Германии. В ряде сетей применяется также оборудование шведской фирмы Ericsson (например, Sitek), фирмы Clarent (сеть «Элвис-Телеком») и других производителей.

Отметим следующие виды услуг IP-телефонии:

1. Дистанционное обучение по схеме: офисный мультимедийный компьютер — персональный компьютер дистанционно удалённого пользователя или обычный телефон.
2. Деловые местные, междугородние или международные переговоры, включая конференц-связь из четырёх, восьми и более абонентов с обычных телефонов по тарифам, значительно более низким по сравнению с обычной телефонной связью.
3. Продажа, консультации и коммерческая реклама в сети IP-телефонии и в сети ТфОП.
4. Связь территориально разнесённых офисов фирмы для пересылки финансовых отчётов, собраний, деловых обсуждений.
5. Голосовой сервис, в том числе голосовая почта (запись и воспроизведение входящих речевых сообщений), чтение текстов, распознавание устной речи, режим «белая доска» и т.д.

Исходя из потенциальных возможностей технологии IP-телефонии, следует ожидать в последующие годы быстрого развития и совершенствования её средств с постепенным замещением и вытеснением услуг традиционной телефонии.

По оценкам зарубежных специалистов (фирма **Philips Group**) объём услуг рынка IP-телефонии в ближайшие годы вырастет в мире от одного миллиарда до 20–30 миллиардов долларов США в год.

Вместе с тем ещё не решены до конца вопросы качества речевой связи, организации конференц-связи, совместимости технологического оборудования и программного обеспечения разных фирм... Ожидается много трудностей на пути развития IP-телефонии в России: значительная стоимость офисного оборудования и программного обеспечения (порядка 30 тысяч долларов США на 10 абонентских оконечных пунктов), ограниченная пропускная способность многих из действующих каналов связи, несовместимость установленного и нового оборудования. При этом уровень цифровизации основных междугородних и особенно местных телефонных линий в России остаётся крайне низким. (По данным компании «Связьинвест», соответствующие показатели в 2000г. составляли 70% и 28,5%. Отметим, что в России ещё в 2003 г. почти 54 тыс. населённых пунктов вообще не были телефонизированы.)

Аппаратные и программные средства VoIP

В качестве абонентских оконечных пунктов сети VoIP выступают ПК, оборудованные звуковой картой типа AWE-64 (Advance Wave Effects) фирмы Creative Labs (США), кодеком (картой или ПО) и модемом со скоростью передачи не менее 14.4 кб/с. ПК подключается через IP-провайдера к сети VoIP. Абоненты сети ТфОП включаются в сеть VoIP также через IP-провайдеров, шлюзы (Gateway) которых обеспечивают оцифровку, кодирование и упаковку речевого сигнала в IP-пакеты. В небольших и корпоративных сетях VoIP маршрутизацию и контроль соединений обеспечивают серверы под управлением стандартного ПО (например, Unix, Windows NT). В крупных сетях для этой цели используют специальные сетевые средства — маршрутизаторы (Gatekeeper), которые обеспечивают необходимую производительность, качество и надёжность соединения.

Взаимодействие мультимедийных оконечных устройств определено рекомендацией H.323 Международного союза электросвязи ITU (International Telecommunication Union). Выработаны соответствующие рекомендации для кодеков речевого сигнала на скорости передачи 6,3; 8,0; 13,0 и 32,0 кб/с (протоколы G.723, G.729, G.728, G.726). Следует отметить конкурирующие предложения разработчиков стандартов для сети Internet по использованию протокола SIP (Session Initiation Protocol — протокол по управлению вызовами и коммутацией), который позволит централизовать все голосовые службы сетей с пакетной коммутацией и положит конец традиционной телефонной системе. По запросу пользователя IP-сети появится возможность поиска абонента в сети, упростятся решения вопросов конференц-связи и совместимости оборудования разных производителей. Протокол SIP способен обслуживать любые виды приложений и коммуникации в реальном времени поверх протокола IP.

Качество речевой связи. Кодеки

К достоинствам сети VoIP следует отнести низкую по сравнению с сетью ТфОП стоимость междугородних и международных соединений, поддержку мультимедийного трафика и развития сервиса. Вместе с тем следует отметить и присущие сети VoIP недостатки, важнейший из которых связан с качеством передачи. Рассмотрим этот вопрос подробнее.



В сети с коммутацией сообщений время прохождения пакетов оказывается непостоянным: пакеты на ЦКС выстраиваются в очередь, время прохождения зависит от нагрузки в сети ПД; замедление пакетов в виртуальном канале перемененно и определяется их маршрутом в сети. В результате время прохождения пакетов по сети **VoIP** может составлять значительную величину (одна и более секунд при использовании каналов спутниковой связи) и сопровождается нарушением порядка и интервала между временем получения пакетов (т.н. джиттером). Кроме того, некоторая часть пакетов оказывается потерянной, так как для уменьшения времени передачи голосовых данных УЗО не используют. Условно считается допустимым потеря до 5% пакетов, при этом вероятность потери подряд 2–3 пакетов достаточно мала. При задержке, превышающей 150 мс, требуется принимать меры против возникновения в разговорном тракте явления эхо (абонентская сеть ТфОП в основном состоит из двухпроводных линий). Эхозаградители, а также детекторы речь/пауза, применяемые в сетях с уплотнением, вносят свою долю в искажение речевого сигнала.

Качество предоставленных услуг сервиса обслуживания **QoS** (Quality of Service) в цепи «точка — точка» системы пакетной коммутации мультимедийных приложений **ATM** (Asynchronous Transform Mode) определяется долей потерянных пакетов, временем задержки и джиттером. В соответствии с рекомендацией **G.114** ITU качество предоставленных услуг **QoS** классифицируется так.

0...25 мс — малая задержка, качество предоставленных услуг **QoS** по оценке методом мнений **MOS** (Mean Opinion Scores) равно пяти ;

25...150 мс — нормальное интерактивное взаимодействие между пользователями. **MOS** = 4...5;

150...400 мс — эффективное взаимодействие затруднено, но ещё допустимо, требуется эхокомпенсация. **MOS** = 2,9...3,8;

более 400 мс — интерактивное взаимодействие крайне затруднено, требуется эхокомпенсация. Режим полудуплексных переговоров. **MOS** = 2...2,9.

В целом для уменьшения времени задержки голосовых пакетов в сети VoIP современными сетевыми технологиями этим пакетам присваивается приоритет, оптимизируется маршрут и предусматривается определённый резерв по загрузке сети. Джиттер пакетов снижается на порядок при увеличении скорости передачи данных в сети Internet от 28кбит/с до 112 кбит/с.

Для передачи голосового трафика по сети ПД аналоговый речевой сигнал оцифровывается методом ИКМ со скоростью передачи 64 кб/с и кодируется — компрессируется кодеком с помощью методов устранения избыточности. В результате удаётся понизить скорость передачи до 13,0...4,8 кб/с. Отметим, что согласно требованиям к качеству речи в цифровых средствах теле/радиовещания и к хранению аудиостереосигнала на CD для кодирования методом ИКМ используется скорость передачи 1,5 мбит/с, а после преобразования по методу **MPEG** (Motion Picture Experts Group), рекомендованному экспертной комиссией ITU, — до 8 кб/с.

Эффективность использования кодеков в сети **VoIP** растёт при понижении скорости передачи голосового сообщения от 13 к 4,8 кб/с, хотя одновременно увеличивается задержка сигнала в кодеке от 0,1мс до 40...100мс. Поэтому предпочтительны те кодеки, которые обеспечивают минимальную задержку при данной скорости передачи и заданном качестве речи.

Таблица 1

Данные для некоторых видов кодеков

Кодек ²	Стандарт	Быстрод. DSP, Mips	Скорость, Кб/с	MOS исх	MOS 2 транз.нч
ADPCM	G.72	8	32	3,7	3,49
RPE-LTP (GSM)	ETSI		13	3,5	3,06
VSELP DAMPS	TIA	20	8	3,4	2,53
MP-MLQ	G.72	17	6,4	3,9	3,41

Из приведённых данных наилучшими показателями обладает кодек, реализующий алгоритм многоимпульсной максимально подобной квантизации **MP-MLQ** (Multipulse Maximum Likelihood Quantization). Алгоритм разработан фирмами **AudioCodes** (Израиль) и **DSP Group** (США) для передачи речи со скоростью 4.8, 6.4, 7.2 и 8 кбит/с. В основу алгоритма положен липредер **LPC-10**. При низкой скорости используются алгебраические коды линейного предсказания **ACELP**, при более высокой — **MP-MLQ**. Время задержки — 37,5 мс. В структуре алгоритма поддерживается программирование с плавающей точкой и кодирование с переменной скоростью. Алгоритм даёт возможность снизить скорость передачи ниже 4 кбит/с и получить время задержки 20 мс. Требуемая скорость обработки составляет 17 mips (млн. инструкций в сек.). Коммерческая реализация алгоритма **MP-MLQ** осуществлена фирмой **RAD Data Communications** (США). Речевой мультиплексер Kilotmux — 2000 содержит несколько плат низкоскоростных кодеков KVC –3, обеспечивающих ведение по линии 64 кб/с одновременно 13 разговоров. Использование такого оборудования особенно эффективно на дорогостоящих каналах, например линиях спутниковой связи.

В отличие от других кодеков алгоритм **MP-MLQ** обеспечивает минимальные искажения речевого сигнала при тандемном соединении вокодеров по низкой частоте, например на стыке сети ПД с сетью ТфОП: при двух транзитах по низкой частоте оценка MOS для алгоритма MP-MLQ практически равна оценке для ADPCM (3.41 против 3.49).

Рассмотрим характеристики некоторых кодеков подробнее.

² ADPCM (Adaptive Pulse Code Modulation) — адаптивная ИКМ; RPE-LTP (Regular Pulse Excitation-Long Term Prediction — система импульсного возбуждения с долговременным предсказанием; GSM (Global System for Mobile) — европейская система подвижной сотовой связи; VSELP (Vector Sum Excited Linear Prediction) — линейное предсказание с возбуждением векторной суммой, DAMPS (Digital Advanced Mobile Phone) — система сотовой и спутниковой связи США.



Кодеки системы GSM

В системе **GSM** для кодирования речевого сигнала используется регулярное импульсное возбуждение и долгосрочное предсказание **RPE-LTR**. В блоке предварительной обработки осуществляется коррекция (предыскажение) спектра входного сигнала при помощи цифрового фильтра, подчёркивающего верхние частоты. Далее на сегментах по 20 мс производится измерение восьми коэффициентов линейного предсказания ki , которые перед передачей в канал связи преобразуются в логарифм отношения площадей ri , причём для функции логарифма используется кусочно-линейная аппроксимация. Сигнал с выхода блока предварительной обработки поступает на фильтр — анализатор кратковременного линейного предсказания и по его выходному сигнал-остатку предсказания eh оцениваются параметры долгосрочного предсказания: коэффициент предсказания g и задержка t . При этом сегмент-остаток из 160 отсчётов кратковременного предсказания e^* разделяется на четыре подсегмента длительностью по 5 мс из сорока выборок в каждом. Параметры t оцениваются для каждого из подсегментов в отдельности, причём для оценки задержки t для текущего подсегмента используется скользящий подсегмент из 40 выборок, перемещающийся в пределах предшествующих 128 выборок сигнал — остатка предсказания e^* . Сигнал e^* фильтруется фильтром-анализатором долгосрочного линейного предсказания, а выходной сигнал последнего fh фильтруется сглаживающим фильтром, и по нему формируются параметры сигнала возбуждения в отдельности для каждого из подсегментов по 40 выборок. Сигнал возбуждения одного подсегмента состоит из 13 импульсов, следующих через равные промежутки времени (втрое большие, чем интервал дискретизации исходного сигнала) и имеющих различные амплитуды.

Таблица 2

Число бит для кодирования параметров по системе GSM

Передаваемые параметры	Число бит	Примечание
Параметры фильтра кратковременного предсказания (логарифм отношения площадей $r_i, i = 1 \dots 8$)	36	R_1, r_2 — по 6 бит; r_3, r_4 — по 5 бит; r_5, r_6 — по 4 бита; r_7, r_8 — по 3 бита;
Параметры фильтра долгосрочного предсказания (коэффициент предсказания g , и задержка t для каждого из четырёх подсегментов)	36	g — 2 бита; t — 7 бит;
Параметры сигнала возбуждения (номер последовательности n , максимальная амплитуда v , нормирование амплитуды импульсов $b_i, i = 1 \dots 13$, для каждого из четырёх подсегментов)	188	n — 2 бита; v — 6 бит; b_i — 3 бита
Всего	260	

Для формирования сигнала возбуждения 40 импульсов подсегмента сглаженного остатка f_n обрабатываются следующим образом. Последний (сороковой) импульс отбрасывается, а первые 39 импульсов разбиваются на три последовательности по 13 импульсов: в первой — импульсы 1,4,7,...,37, во второй — импульсы 2,5,8,...,38, в третьей — импульсы 3,6,9,...,39. В качестве сигнала возбуждения выбирается та из последовательностей, энергия которой больше.

Амплитуды импульсов нормируются по отношению к импульсу с наибольшей амплитудой. Нормированные амплитуды кодируются тремя битами каждая по линейной шкале квантования. Абсолютное значение наименьшей амплитуды кодируется шестью битами в логарифмическом масштабе. Положения начального импульса 13-элементной последовательности кодируется двумя битами, т.е. формируется номер последовательности, выбранной в качестве сигнала возбуждения для данного подсегмента. Таким образом, выходная информация кодека для 20 мс сегмента речи включает в себя: параметры фильтра кратковременного линейного предсказания — восемь коэффициентов логарифма отношения площадей $r_i, i=1, \dots, 8$; параметры фильтра долговременного линейного предсказания — коэффициент предсказания q и задержку t для каждого из четырёх подсегментов; параметры сигнала возбуждения, номер последовательности, максимальную амплитуду v , нормированные амплитуды $b_i, i = 1 \dots 13$, импульсов последовательности для каждого из четырёх подсегментов. Всего для одного 20 мс сегмента речи отводится 260 бит, т.е. кодер осуществляет сжатие информации в пять раз ($1280 : 260 = 4,92$).

Перед выдачей в канал связи выходная информация кодера подвергается дополнительно каналному кодированию. В декодере блок формирования сигнала возбуждения, используя принятые параметры возбуждения, восстанавливает 13-импульсную последовательность сигнала возбуждения для каждого из подсегментов, включая амплитуды импульсов и их расположение во времени. Сформированный таким образом сигнал возбуждения фильтруется фильтром-синтезатором долговременного предсказания, на выходе которого получается восстановленный остаток предсказания фильтра-анализатора кратковременного предсказания. Последний фильтруется решётчатым (лестничным) фильтром-синтезатором кратковременного предсказания, причём параметры фильтра предварительно преобразуются из логарифма отношений площадей r_i в коэффициенты частной корреляции k_i .

Выходной сигнал фильтра — синтезатора кратковременного предсказания фильтруется (в блоке постфильтрации) цифровым фильтром, восстанавливающим амплитудные соотношения частотных составляющих сигнала речи, т.е. компенсирующим коррекцию, внесённую входным фильтром блока предварительной обработки кодера. Сигнал на выходе постфильтра является восстановленным цифровым сигналом речи. GSM-кодек выдаёт информацию со скоростью 13 кбит/с. Главным для этого кодека является то, что он может быть легко реализован для работы в реальном времени при малых вычислительных затратах.

При передаче в системе **GSM** используется техника прерывистой передачи **DTX** (Discontinuous Transmission). При такой системе передачи в групповом радиоканале каждый речевой канал активен не непрерывно. В дуплексном режиме переговоров каждый участник говорит менее 50 % времени. Кроме того, во время разговора между словами и фразами также есть паузы. Для определения интервалов активности используется детектор активности речи (**VAD**); при этом групповой канал связи в обнаруженных паузах может быть предоставлен для передачи других переговоров. При использовании **VAD** в паузах речи могут передаваться неречевые данные. Эффективность таких систем зависит от алгоритма **VAD**, который работает в условиях воздействия внешних шумов, типичных, например, для подвижной автомобильной радиосвязи.

Прерывистая передача DTX

DTX — эффективный способ повышения эффективности подвижных систем передачи речи. Основной принцип DTX — включение передатчика только в те временные интервалы, когда присутствует речевой сигнал. Для устранения интерференции с соседними каналами и для



предохранения аккумуляторных батарей носимых радиостанций от разряда передатчик может выключаться. Основной проблемой **DTX** является потенциальное снижение качества речи из-за того, что речь может идентифицироваться как шум. При использовании детектора активности канала возможны следующие нежелательные явления: пропадания участков речи и возможность того, что шум будет неправильно идентифицирован как речевой сигнал. Пропадания могут существенно снизить общее качество речи.

В случае, когда VAD используется для включения и выключения передатчика, шум на приёмной стороне может изменяться по уровню. Это явление связано с тем, что при включённом передатчике фоновый шум передаётся вместе с речью. Однако когда речевой сигнал отсутствует, передатчик выключается, что приводит к снижению фоновых шумов до очень низкого уровня. Это случайное изменение в уровне шумов неприятно для слушателя и может повлиять на разборчивость речи. Для уменьшения этого эффекта на время выключения передатчика в декодере производится генерация шума. Этот шум должен быть похож, например, на шум машины или поезда на передающей стороне. Поэтому передатчик периодически передаёт информацию о среднем уровне фонового шума (т.н. «комфортный шум»).

Передатчик состоит из кодека речи, VAD и измерителя уровня фонового шума. Когда на входе есть речь, передатчик включён. Во время речевых пауз передатчик выключается. Через определённое небольшое время передатчик включается на один фрейм для передачи информации о среднем уровне фона и генерации на приёмной стороне комфортного шума. На приёмной стороне при наличии речевого сигнала происходит нормальный синтез. Если не поступает новой информации о фоне, используются существующие параметры шума и генерируется комфортный шум. Когда на приём поступают новые параметры фонового шума, начинается генерация нового комфортного шума. Обычно на стороне декодера используется индикатор «хороший/плохой» фрейм, чтобы показать, верны или нет декодированные параметры, и если верны, то производится замена фрейма. Эффективность DTX зависит от точности VAD.

Кодеки стандарта DAMPS

Цифровой стандарт мобильной радиосвязи DAMPS (Digital Advanced Mobile Phone Service), принятый в США в 1990 г., по своим функциональным возможностям и предоставляемым услугам приближается к стандарту GSM. Стандарт DAMPS не принят в европейских странах, за исключением России, где он ориентирован в основном на региональное использование. В стандарте DAMPS используется метод кодирования VSELP. Блок предварительной обработки выполняет цифровую фильтрацию входного сигнала с подъёмом верхних частот. Для каждого 20 мс сегмента оцениваются параметры фильтра кратковременного линейного предсказания — 10 коэффициентов частотной корреляции $r_i, i = 1 \dots 10$, которые непосредственно кодируются для передачи в канал связи без каких-либо дополнительных преобразований, и определяется энергия сегмента речи p . Сигнал с выхода блока предварительной обработки фильтруется фильтром-анализатором кратковременного линейного предсказания $A(z)$, имеющего форму инверсного линейного фильтра,

для чего коэффициенты частной корреляции преобразуются в коэффициенты линейного предсказания a_i .

Выходной сигнал фильтра кратковременного предсказания (сигнал-остаток предсказания e_n) используется для оценки параметров фильтра $P(z)$ долговременного предсказания. Оценки даются для каждого из четырёх подсегментов по 40 выборок, на которые разделяется сегмент из 160 выборок. Для каждого из подсегментов определяются параметры сигнала возбуждения. Для этого в составе кодера используется схема, аналогичная входящей в состав декодера, которая включает фильтры-синтезаторы кратковременного $H(z)$ и долговременного $R(z)$ предсказания и две кодовые книги и реализуется метод «анализа через синтез».

Каждая из кодовых книг сигнала возбуждения содержит 128 кодовых векторов, по 40 элементов в каждом. Все кодовые векторы одной книги являются элементами 7-мерного линейного подпространства в 40-мерном пространстве. Каждая кодовая книга, содержащая 128 векторов, задаётся семью базисными векторами и 128 кодовыми словами (7-элементными векторами коэффициентов линейных комбинаций) с однобитовыми элементами.

Сигнал возбуждения фильтра-синтезатора кратковременного предсказания является суммой векторов возбуждения из двух кодовых книг и вектора с выхода фильтра-синтезатора долговременного предсказания. Векторы возбуждения из кодовых книг до подачи на сумматор умножаются на соответствующие коэффициенты усиления t_1 и t_2 , а входным сигналом фильтра — синтезатора долговременного предсказания является, в зависимости от участка сегмента, выходной сигнал того же фильтра или суммарный сигнал возбуждения фильтра-синтезатора кратковременного предсказания. Параметры сигнала возбуждения — номера векторов возбуждения l_1 и l_2 из первой и второй кодовых книг и соответствующие коэффициенты усиления t_1 и t_2 — определяются по критерию минимума среднеквадратичной ошибки на выходе фильтра-синтезатора кратковременного предсказания, входящего в состав кодера. Предварительно базисные векторы обеих кодовых книг декоррелируют: для первой книги — по отношению к выходному вектору фильтра-синтезатора долговременного предсказания, для второй книги — по отношению к тому же выходному вектору и к базисным векторам первой книги.

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

- параметры фильтра кратковременного линейного предсказания — 10 коэффициентов частной корреляции r_i ; $i = 1 \dots 10$, и амплитудный множитель ρ — один выбор на весь сегмент;
- параметры фильтра долговременного линейного предсказания — коэффициент предсказания g и задержку τ — для каждого из четырёх подсегментов;
- параметры сигнала возбуждения — номера l_1 и l_2 векторов возбуждения из двух кодовых книг и соответствующие коэффициенты возбуждения t_1 и t_2 для каждого из четырёх подсегментов.

Перед передачей в канал связи выходная информация кодера речи подвергается дополнительному каналному кодированию, причём разные параметры в зависимости от их важности для обеспечения качества речи кодируются с различной степенью избыточности.



Таблица 3

Характеристики кодека (согласно стандарту DAMPS)

Передаваемые параметры	Число бит	Примечание
Параметры кратковременного предсказания (коэффициенты частичной корреляции $r_i, i = 1 \dots 10$)	38	k_1 — 6 бит; k_2, k_3 — по 5 бит; k_4, k_5 — по 4 бита; k_6, \dots, k_9 — по 3 бита; k_{10} — 2 бита
Амплитудный множитель (энергия сегмента) p	5	
Задержка фильтра долговременного предсказания t	28	7 бит на каждый подсегмент
Номера векторов возбуждения l_1 и l_2	56	l_1 и l_2 по 7 бит
Коэффициенты усиления g, γ_1 и γ_2	32	8 бит на каждый подсегмент; векторное квантование
Всего на 20 мс сегмент	159	

Общий объем информации, выдаваемой в канал связи равен 8000 бит/с. Поскольку исходный объем информации на выходе кодека составляет 1280 бит (160 выборок по 8 бит), кодек осуществляет сжатие информации более чем в восемь раз.

В декодере сигнал возбуждения фильтра-синтезатора кратковременного предсказания формируется таким же образом, как и в синтезирующей схеме кодека: по номерам l_1 и l_2 из кодовых книг выбираются векторы возбуждения, которые умножаются соответственно на коэффициенты усиления t_1 и t_2 и складываются с выходным вектором фильтра-синтезатора долговременного предсказания, определяемого параметрами g и t .

Окончательно сигнал возбуждения фильтруется фильтром-синтезатором кратковременного предсказания, выполненного в виде инверсного фильтра, т.е. параметры фильтра преобразуются из коэффициентов частной корреляции r_i в коэффициенты предсказания a_i . Для улучшения субъективного качества синтезированной речи выходной сигнал фильтра-синтезатора подвергается цифровой адаптивной пост-фильтрации, и с выхода пост-фильтра получается восстановленный цифровой речевой сигнал.

Стандарт TETRA

TETRA разработан как единый общеевропейский цифровой стандарт на основе технических решений и рекомендаций стандарта GSM и ориентирован на создание систем связи, эффективно и экономно поддерживающих совместное использование сети различными группами пользователей с обеспечением засекречивания информации. При разработке стандарта ориентировались, прежде всего на создание профессиональных систем связи, хотя системы общего пользования также могут быть созданы на основе этого стандарта.

Системы по стандарту TETRA предназначены для организации связи с абонентами телефонных сетей, радиосвязи, передачи данных. В стандарт входят спецификации беспроводного интерфейса, интерфейсов между сетью TETRA и цифровой

сеть с интеграцией услуг (ISDN), телефонной сетью общего пользования, сетью передачи данных, учрежденческими АТС.

TETRA — полностью цифровая система, поддерживает обслуживание речевых сообщений и данных различного формата с обеспечением выбора скорости передачи данных и уровня защиты от ошибок. В системе используется технология TDMA с четырьмя каналами на одной несущей при разносе между несущими 25 кГц, что обеспечивает высокую эффективность использования частотного спектра.

Время установления вызова в системе составляет 300 мс. TETRA поддерживает работу в режиме полудуплекса для связи групп и дуплекса для индивидуальных вызовов. Возможности группового и циркулярного вызовов соответствуют требованиям большинства пользователей. Схема многих приоритетов обеспечивает эффективное распределение ресурса для самых важных соединений в сети.

Основные характеристики протокола радиointерфейса стандарта TETRA. Рабочие частотные каналы отстоят друг от друга на 25 кГц. На каждом частотном канале располагаются четыре временных интервала, которые и являются физическими каналами связи, т.е. основным элементом временной структуры является MVDR-кадр (Minimum Variance Distortionless Response), который содержит четыре пакета. Скорость передачи данных одного канала (пакета) 7,2 кбит/с. Речевой сигнал кодируется со скоростью передачи 4,8 кбит/с с использованием метода ACELP. Цифровые данные с выхода кодера подвергаются блочному сверточному кодированию, перемежению и шифрованию, после чего формируются информационные каналы. Пропускная способность одного информационного канала составляет 7,2 кбит/с, а скорость цифрового информационного потока данных — 28,8 кбит/с. При этом общая скорость передачи символов в радиоканале за счёт дополнительной служебной информации и контрольного кадра в мультикадре равна 36 кбит/с.

В стандарте TETRA предусмотрены различные способы защиты пользовательских данных от помех в радиоканале, а также несколько вариантов использования полосы пропускания радиоканала. Если пользователь сам обеспечивает целостность и достоверность пакетов данных, система TETRA может представить «прозрачный» канал связи, не внося дополнительных символов корректирующего кода. В этом случае скорость передачи данных при использовании только одного временного интервала в кадре составит 7,2 кбит/с. Для повышения скорости передачи данных может быть предоставлено от одного временного интервала до четырёх. В этом случае скорость передачи данных пропорционально увеличивается и составляет 28,8 кбит/с. Если пользователь не обеспечивает достоверность информации собственными средствами, можно использовать систему помехоустойчивого кодирования системы TETRA. При этом можно работать с использованием кодовых скоростей 2/3 (4,8 кбит/с — низкая степень защиты) или 1/3 (2,4 кбит/с — высокая степень защиты).

MVDR — кадр содержит четыре временных интервала (пакета). Пакет в кадре соответствует независимому каналу передачи информации. Каждый пакет, в зависимости от его назначения, имеет свою внутреннюю структуру. Пакет содержит 510 бит цифровой информации, что соответствует 255 символам модуляции.

Восемнадцать MVDR-кадров объединены в мультикадры, которые, в свою очередь, образуют гиперкадр длиной 60 мультикадров. В стандарте TETRA для организации связи между подвижным абонентом и базовой радиостанцией предусматривается выделение дуплексной пары радиочастот. Так как при этом используется временное уплотнение до четырёх независимых каналов, для снижения взаимных помех в системе применяется жёсткая синхронизация пакетов подвижных станций относительно пакетов, передаваемых базовой станцией; при этом



последовательность пакетов мобильных станций задерживается на две позиции относительно пакетов базовой станции.

Практически в любом пакете от базовой станции к абонентской имеются поля, предназначенные для передачи команд управления и сигнализации. Таким образом, помимо канала управления в 18-м кадре сигнала базовой станции, элементы управления присутствуют во всех информационных кадрах. Метод линейной модуляции, применяемый в TETRA — $t/4$ — QDPSK.

Определены классы мощности для радиостанций, используемых в системе: 25, 10, 3 и 1 Вт. Радиостанции могут автоматически регулировать выходную мощность в соответствии с нужной напряжённостью поля. Для увеличения зоны действия носимых или возимых абонентских радиостанций предусматривается их использование как ретрансляторов для выхода во внешние сети для работы в сети или организации локальных сетей в режиме двухчастотного симплекса.

Сети стандарта TETRA предполагают распределённую инфраструктуру управления и коммутации, обеспечивающую быструю передачу вызовов и сохранение локальной работоспособности системы при отказе её отдельных элементов. Основными элементами сетей TETRA являются базовые и мобильные станции, устройства управления базовыми станциями, контроллеры базовых станций, диспетчерские пульты, терминалы технического обслуживания и эксплуатации. Функции сетевого обслуживания и межсистемного взаимодействия определяются специальными интерфейсами:

- радиointерфейсом, определяющим взаимодействие базовой станции с мобильными абонентскими радиостанциями;
- радиointерфейсом непосредственного соединения между абонентскими радиостанциями;
- интерфейсом проводной связи, связывающим контроллер базовой станции с диспетчерским пультом;
- межсистемным интерфейсом для организации связи между контроллерами базовых станций различных сетей;
- интерфейсом связи между терминалами передачи данных и мобильной станцией или диспетчерским пультом;
- интерфейсом управления сетью;
- интерфейсом для подключения к учрежденческим АТС, ТФОП, ISDN, сети с коммутацией пакетов.

Система стандарта TETRA предусматривает выполнение следующих функций:

- индивидуальный вызов между абонентскими радиостанциями и ведение радиотелефонных переговоров;
- вызов абонентских радиостанций со стороны абонентов телефонной сети и вызов абонентов телефонной сети со стороны абонентских радиостанций;
- групповой вызов;
- аварийный вызов;
- передача данных (коротких сообщений, статусных сообщений и пакетов данных);
- передача циркулярных сообщений;
- организация конференц-связи;
- присвоение приоритетов;
- переадресация вызовов;

- организация динамических групп абонентов;
- защита от несанкционированного доступа к сети;
- постановка на очередь абонентов при отсутствии свободных каналов с последующим обслуживанием;
- установка таймеров длительности вызова, соединения;
- ограничение перечня предоставляемых услуг для отдельных абонентов;
- контроль оборудования.

Данные любого характера и сообщения произвольной длины могут быть переданы в системе TETRA с использованием службы пакетной передачи. Это служба предоставляет пользователям сервис протокола IP. Служба пакетной передачи по протоколу IP позволяет использовать все виды протоколов транспортного уровня. Это может быть как датаграммный протокол (UDP), не предусматривающий установления логического соединения, так и ориентированный на установление сеанса связи. Пользователи могут использовать различное программное обеспечение, работающее с протоколами TCP/IP. При подсоединении к инфраструктуре TETRA внешнего оборудования обработки данных, а также при присоединении к абонентским радиостанциям абонентского терминального оборудования используются протоколы X.25, LAP.B, X.21.

Речевой кодек TETRA основан на модели кодирования CELP — линейное предсказание с кодовым возбуждением. В этой модели блок из N речевых выборок синтезируется путём фильтрации соответствующей обновлённой последовательности из кодовой книги, масштабированной коэффициентом усиления g_c , с помощью двух изменяющихся во времени фильтров. Первый фильтр является фильтром долгосрочного предсказания (фильтром основного тона), цель которого — моделирование псевдопериодического речевого сигнала, а второй — фильтр краткосрочного предсказания — моделирует огибающую речевого спектра.

Передаточная характеристика долгосрочного фильтра (или фильтра синтеза основного тона) определяются формулой:

$$\frac{1}{B(z)} = \frac{1}{1 - g_p \cdot z^{-T}},$$

где T — задержка основного тона; g_p — коэффициент усиления основного тона.

Фильтр синтеза основного тона выполнен как адаптивная кодовая книга, где для задержек, меньших чем длина подфрейма, повторяется последнее возбуждение.

Краткосрочный фильтр синтеза определяется формулой:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_i \cdot z^{-1}},$$

где $a_i, i=1, \dots, p$ — параметры линейного предсказания; p — порядок предсказателя. В кодеке TETRA $p=10$.

Для определения основного тона и параметров кодовой книги возбуждения в кодеке TETRA используется способ анализ-через-синтез. При способе анализ-через-синтез синтезированная речь вычисляется для всех кандидатов — последовательностей, составляя особую последовательность, которая и формирует выходной сигнал, наиболее близкий к исходному, в соответствии со взвешенной величиной измеренных искажений. Фильтр взвешивания определяется формулой:



$$W(z) = \frac{A(z)}{\tilde{A}(z)},$$

где $A(z)$ — обратный (инверсный) фильтр линейного предсказания; ($0 < z_1$ (используется значение $z_1 = 0,85$). Для взвешивающего фильтра $W(z)$ и фильтра синтеза $H(z)$ используются квантованные параметры линейного предсказания.

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

$$F(z) = \frac{A\left(\frac{z}{z_1}\right)}{A\left(\frac{z}{z_n}\right)},$$

где $A(z)$ — инверсный фильтр линейного предсказания (в конкретных реализациях $z_1 = 0,75$ и $z_2 = 0,85$).

В кодеке TETRA используются фреймы длительностью по 30 мс. Это требуется для того, чтобы параметры краткосрочного предсказания вычислялись и передавались в каждом речевом фрейме. Речевой фрейм разделён на 4 подфрейма по 7,5 мс (60 выборок). Основной тон и параметры алгебраической кодовой книги также передаются в каждом подфрейме. В табл. 4 представлено распределение бит для кодека TETRA. Описание каждого фрейма длительностью 30 мс занимает 137 бит, что в результате даёт скорость передачи параметров речи 4567 бит/с.

Таблица 4

Распределение бит для кодека TETRA

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

Литература

1. Акинфиев Н.Н. К вопросу построения речевых сообщений // Доклады комиссии по акустике АН СССР. Апрель 1956.
2. Архипова А.Д., Сапожков М.А. Перспективы повышения качества вокодерной речи // Материалы шестой Всесоюз. Акуст. конф. М., 1968.
3. Вокодерная телефония / Под ред. А.А. Пирогова. М., 1974.
4. Волошенко Ю.Я., Михайлов В.Г., Морозов Н.А. К вопросу о регистрации колебаний голосовых связок // Вопросы радиоэлектроники. 1968. Сер. XI. Вып. 7.
5. ГОСТ Р 50840–95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М., 1995.
6. Дремов А.Н. Решительный шаг к интеграции // Технологии и средства связи. 2001. № 2.
7. Калачев К.Ф. В круге третьем. М., 2001.
8. Калинин Ю.К. Разборчивость речи в цифровых вокодерах. М., 1991.
9. Кортаев Г.А., Михайлов В.Г. Синтетическое телефонирование // Радиоэлектроника и электронная техника. 1964.
10. Кортаев Г.А., Михайлов В.Г. Современное состояние техники параметрического компандирования речи // Зарубежная радиоэлектроника. 1966. № 4.
11. Котельников В.А. Теория потенциальной помехоустойчивости. М., 1956.
12. Кубышкин Ю.И., Халышкин А.С. Полосный полувокодер для применения на междугородных линиях связи // Труды VII Всесоюз. Акуст. конф. Л., 1971.
13. Лейтес Р.Д., Соболев В.Н. Принципы цифрового моделирования вокодеров // Электросвязь. 1966. № 7.
14. Литвак И.М. О разработке систем типа вокодер // Доклады комиссии по акустике АН СССР. Апрель 1956.
15. Мартынов В.С. Выделитель основного тона // Доклады комиссии по акустике АН СССР. Апрель 1957.
16. Масленников И. Будущее реального времени // Доклады комиссии по акустике АН СССР. Апрель 1957.
17. Материалы семинара «IP-телефония и дистанционное обучение». М., 2000.
18. Михайлов В.Г. Формантное распределение для мужских голосов // Акустический журнал. 1972. Т. 1.
19. Михайлов В.Г. Аппаратурные методы измерения качества телефонной передачи // Зарубежная радиоэлектроника. 1973. № 5.
20. Михайлов В.Г. Семинар по речевой связи в Стокгольме // Электросвязь. 1975. № 4.
21. Михайлов В.Г. Информационные и статистические параметры устной речи. М., 1992.
22. Михайлов В.Г. Новые информационные технологии. IP-телефония // Системы и средства связи, телевидения и радиовещания. 2000. № 3.
23. Михайлов В.Г. IP-телефония // Акустика речи и прикладная лингвистика. Ежегодник РАО. Вып. 3. М., 2002.
24. Муравьев В.Е. Гармоническая система кодирования речи // Труды гос. НИИ Минсвязи. 1959. Вып. 1(15).
25. Мясников Л.Л. Объективное распознавание звуков речи // Журнал технической физики. 1943. Т. 13. № 3.
26. Пирогов А.А. Гармоническая система сжатия спектров речи // Электросвязь. 1959.
27. Покровский Н.Б. Расчёт и измерение разборчивости речи. М., 1962.
28. Потапова Р.К. Основные современные способы анализа и синтеза речи. М., 1971.
29. Потапова Р.К. Речевое управление роботом: лингвистика и современные автоматизированные системы. М., 1989.
30. Потапова Р.К. Речь: коммуникация, информация, кибернетика. М., 1997.
31. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. М., 1981.
32. Радиовещание и электроакустика / Под ред. Ю.А. Ковалгина. М., 1998.



33. *Распознавание слуховых образов* / Под ред. Н.Г. Загоруйко, Г.Я. Волошина. Новосибирск, 1970.
34. *Рейман Л.Д.* Россия на пути к информационному обществу // Технологии и средства связи. 2001. № 3.
35. *Росляков А.В., Самсонов М.Ю., Шибяева И.В.* IP-телефония. М., 2001.
36. *Сапожков М.А.* О методах компрессии речи // Электросвязь. 1958. № 8.
37. *Сапожков М.А.* Речевой сигнал в кибернетике и связи. М., 1963.
38. *Сапожков М.А., Михайлов В.Г.* Вокодерная связь. М., 1983.
39. *Фант Г.* Анализ и синтез речи: Пер. с англ. Новосибирск, 1970.
40. *Фланеган Дж.* Анализ, синтез и восприятие речи: Пер. с англ. М., 1968.
41. *Шелухин О.И., Лукьянцев Н.Ф.* Цифровая обработка и передача речи. М., 2000.
42. *Цемель Г.И.* Системы сокращения спектра речевого сигнала // Электросвязь. 1957. № 5.
43. *Atal B.* High — quality speech at low bit rates: multi — pulse and stochastically exited linear predictive coders // ICASSP-86. Tokyo. 1986.
44. *Dolansky L., Tjernlung P.* On certain irregularities of voiced-speech waveforms // IEEE Trans. Audio and El. 1968. V. 16. № 1.
45. *Dudley H.* A synthetic speaker // J. Franklin Inst.. 1939.
46. *Dudley H.* Remaking speech // JASA. 1939. № 11.
47. *Dudley H.* Vocoders // Bell Labs Record. 1939. № 17.
48. *Gold B.* Computer program for pitch extraction // J. Acoust. Soc. Am. 1962. V. 32. № 7.
49. *Halsey R., Swaffield J.* Analysis — synthesis telephony with special reference to the vocoder // J. of the Inst. of El. Eng. 1948. V. 95. № 34.
50. *Koenig W., Dunn H., Locy L.* The Sound spectrograph // JASA. 1946. № 18.
51. *Potter R.K.* Visible speech. N.Y., 1947.
52. *Schroeder M.R.* Vocoders: analysis and synthesis of speech // Proc. of IEEE. 1966. V. 54. № 5.
53. *Schroeder M., David E.* A vocoder for transmitting 10 kc/s Speech over a 3,5 kc/s channel // Acoustica. 1960. V. 10. № 1.
54. *Shannon C.* A mathematical theory of communication // Bell system Techn. J. 1948. V. 27. № 3; № 4.
55. *Specom — 1999.* Proc. of Int. Workshop «Speech and Computer». Moscow, 1999.
56. *Specom — 2001.* Proc. of Int. Workshop «Speech and Computer». Moscow, 2001.
57. *Speech synthesis* / Ed. by J.L. Flanagan, L.R. Rabiner. N.Y., 1973.
58. *Tremain T.* The government standard linear predictive coding algorithms LPC — 10// Speech technology. 1982. V. 1. № 2.
59. *Voice Compression Technology.* RAD data Communications. White paper. Ver. 2. 5/96. Cat. 801105.

Михайлов Вадим Георгиевич —

старший научный сотрудник, доктор филологических наук, Московский государственный университет им. Ломоносова, филологический факультет; Окончил в 1955 г. Ленинградский институт киноинженеров по специальности «звукозапись». В течение 30 лет работал в научно-исследовательских институтах и с 1985 г. — в Московском государственном университете им. М.В. Ломоносова в должности старшего научного сотрудника; В Московском государственном лингвистическом университете на должности профессора. Область научных интересов: акустико-перцептивные свойства речевого сигнала, автоматическое распознавание и разборчивость речи. Имеет около 100 научных публикаций, 3 монографии, 12 изобретений, в том числе гос. стандарт по оценке качества передачи речевого сигнала.