Из истории исследований преобразования речи

В.Г. Михайлов,
доктор филологических наук

Введение

За прошедшие годы после изобретения телефона А. Беллом в 1876 г. электросвязь стала неотъемлемой частью социально-экономической жизни современного мира. Сети электросвязи играют особую роль в развитии и прогрессе страны. Так, каждый рубль, вложенный в средства электросвязи России, дает до шести рублей прироста общественного продукта. Устойчиво развиваются сети сотовой связи. Число их пользователей в России уже достигло 70 млн.

Новые информационные технологии породили концепцию мобильных сетей третьего поколения 3G. Новые информационные технологии называют золотым ресурсом нации в XXI веке. Число абонентов телефонной сети России превышает 40 млн, а пользователей сети Internet — 8 млн человек. Наблюдается тенденция к изменению объема нынешнего трафика речь/данные равного 4/1 на обратное. (Для сравнения: число пользователей сети Internet, обслуживаемых провайдерами включают видеоконференции, банковские операции, телешопинг, видео передачи по запросу пользователя, мобильные ТВ и развлекательные программы, лотереи. Требуемая скорость передачи составляет 2 Мб/с (в ныне действующей сотовой сети — 9,6–13 кбит/с). Протокол беспроводных приложений WAP, разработанный компаниями Nokia, Ericsson и PhoneCom переводит модель сети Internet в мир подвижной связи. Уже выданы лицензии на создание сети 3G в Японии, Финляндии, Англии, Южной Корее. Принято решение о строительстве опытных зон сети 3G в Москве, Санкт-Петербурге. К тестированию системы 3G на базе протокола WAP приступают операторы фирмы Мобильные системы (Москва). Фирма ВымпелКом готовится к освоению пакетной передачи сообщений на базе технологий GPRS со скоростью передачи 40,2…115 кб/с. Ожидается поступление в Россию абонентских терминалов подвижной связи GPRS, которые обеспечат поступления через Internet видео и музыкальных файлов. В сети 3G на смену обычной микротелефонной трубке придет устройство персональной интерактивной связи, использующее наземные и спутниковые технологии и глобальные сети связи.

Новые информационные технологии становятся движущей силой страны. Без существенного прироста капвложений в отрасль и широкого притока свежих сил квалифицированных специалистов России будет трудно дать адекватный ответ мировым лидерам в сфере мультимедийных коммуникаций XXI века.

* * *

Сети электросвязи включают в себя каналы проводной и радио связи, каналообразующее и абонентское оборудование. Стандартный телефонный канал — ТЧ-канал обеспечивает передачу речевого сигнала в полосе частот на 300…3400 Гц при уровне помех

–40 дБ. Передача информации по ТЧ-каналу в цифровой форме (при установке на входе-выходе канала специального модема) ограничена скоростью около 4,8 кбит/с., а по коротковолновому КВ-радиоканалу — 2,4 кбит/с (при применении модема системы Kiniplex). Из-за присущих каналам связи ограничений в пропускной способности, наличию помех и других видов искажений речевого сигнала, а также их возрастающей стоимости уже в 30-х годах прошлого века оказались востребованы методы эффективного кодирования речевого сигнала и, в частности, методы параметрического компандирования — вокодеры (англ. — voice coder). *Назначение вокодера — сократить (компрессировать) на передаче путем анализа объем речевого сигнала в 5 — 50 раз по сравнению с исходным (с полосой частот 300 — 3400 Гц в аналоговой форме или около 64 кбит/с в цифровой форме) и восстановить (экспандировать) сигнал на приеме путем синтеза при допустимых потерях качества звучания.* Нормы допустимого снижения качества речи регламентированы в России государственным стандартом ГОСТ Р 50840 — 95. Отметим, что вокодеры в аналоговом режиме работы для уплотнения ТЧ каналов не нашли широкого применения из-за высокой стоимости оборудования и сравнительно низкого качества речи.

В развитии методов параметрического компандирования речи можно выделить четыре этапа: **1936-1952** гг. — начальный (первичный поиск); **1953-1974** гг. — отработка известных решений и развитие новых методов; **1975-1987** гг. — оптимизация моделей на базе линейного предсказания; после **1987** г. — внедрение вокодеров в глобальную систему голосовой связи по сети Internet со скоростью передачи 4,3 кбит/с (**IP**-телефония) и в радио — каналы мобильной телефонной связи МТС (GSM — Global System for Mobil — европейский стандарт; DAMPS — Digital Advanced Mobile Phone Service — стандарт США) со скоростью передачи 8…13 кбит/с. Приведем характеристики указанных этапов.

Первый этап (1936 — 1952 гг.)

Первый вокодер был продемонстрирован 11 сентября 1936 г. на Гарвардской научной конференции (США).

Спектр речевого сигнала в анализаторе делился с помощью полосовых фильтров равной ширины по 300 Гц на 10 спектральных полос, в каждой из которых измерялся текущей уровень сигнала (т.н. сигнал — параметра). Сумма указанных параметров вместе с сигналом основного тона ОТ (частотой колебаний голосовых связок) и видом спектра сигнала — тональный или шумовой ТШ поступала в синтезатор. Такой вокодер получил название спектрально-полосный. Для передачи сигнал — параметров требовалось около 400 Гц, то есть спектрально-полосный вокодер обеспечивал сжатие полосы передаваемых частот речевого сигнала в 7-8 раз и создавал возможность уплотнить в несколько раз стандартный телефонный канал.

Предложенная Г. Дадли (H.Dudley) схема вокодера была в дальнейшем значительно усовершенствована. Спектр речевого сигнала на передающем конце разделяется узкополосными фильтрами ПФ на частотные полосы (спектральные каналы), в которых путем детектирования и сглаживания фильтрами нижних частот определяются временные огибающие. На передающем конце выделяется частота основного тона с помощью выделителя ос-

новного тона ВОТ и определяется характер спектра сигнала возбуждения (звонкий-глухой) с помощью выделителя тон-шум ВТШ. Эти сигнал-параметры передаются в аналоговой или импульсной форме с помощью каналообразующего оборудования (включая модем) по каналу связи на приемную сторону в синтезатор, где имеются генератор импульсов ГИ и генератор шума ГШ. Сигнал основного тона управляет частотой ГИ, а входные фильтры переключаются на выход ГИ или ГШ по сигналу тон-шум. Широкополосный сигнал, созданный одним из генераторов (сигнал возбуждения), разделяется на частотные полосы гребенкой полосовых фильтров, аналогичных фильтрам на передаче. С выхода ПФ частотные составляющие сигнала возбуждения подаются на модуляторы М, в которых временные огибающие управляют их амплитудами. Для устранения нежелательных продуктов модуляции на выходе модулятора включается еще одна гребенка полосовых фильтров, в результате получается синтезированный сигнал, приближенно отображающий исходный естественный сигнал.

Число спектральных каналов, на которые разделяется спектр речевого сигнала составляет 7-20. С увеличением числа каналов повышается разборчивость синтезированной речи и ее качество (натуральность звучания). Так, 20-канальный полосный вокодер позволяет воспроизводить спектральную огибающую сигнала с точностью до ширины частотных групп слуха, в пределах которых ухо не замечает перемещения максимума спектра. Увеличение числа каналов свыше десяти сказывается, главным образом на улучшении качества.

Были предложены новые области эффективного использования вокодеров: изучение процессов порождения и восприятия речи; синтез звучащей речи по печатному тексту; автоматическое распознавание звуковых образов; коррекция речи водолазов (при дыхании гелио-кислородной смесью происходит смещение спектра речи вверх по частоте); средства помощи лицам с нарушениями речи и слуха и др.

Значительную роль на первом этапе разработки сыграли работы: R.K. Potter (1947) ; W. Koenig, H. Dunn, L. Locy (1946); H. Dudley (1939); R. Halsey, J. Swaffield (1948) и др. Обширная библиография работ других зарубежных ученых приведена в монографии М.А Сапожкова (1963).

В послевоенные годы получили развитие методы цифрового кодирования и передачи речевого сигнала. Цифровые методы обработки обладали существенными достоинствами по сравнению с аналоговыми: математически точное описание процедур обработки речевого сигнала, устранение процессов накопления помех и искажений сигнала на длинных линиях связи и при переприемах по низкой частоте, обеспечение хранения и перезаписи сигнала без искажений. Работы К. Шеннона (1948) и В.А. Котельникова стимулировали развитие методов цифровой защиты информации в каналах связи.

О разработке вокодеров в СССР в эти годы известно немного. Так, Л.Л. Мясников (1943) опубликовал в блокадном Ленинграде статью по методу автоматического распознавания гласных с использованием 8-канального анализатора спектра речи. Свой вклад в разработку вокодерной техники на первом этапе внесли: А.М. Васильев; Ю.Я. Волошенко; А.Н. Кабатов; Ю.К. Калинцев; В.А. Котельников; Р.Д. Лейтес; В.С. Мартынов; А.Н. Нанос; А.П. Петерсон; Н.А. Кокорева; А.М. Трахтман и др. Уже в 1952 г. была создана, возможно, первая в мире цифровая система вокодерной связи с защитой информации, работавшая по протяженным стандартным проводным ТЧ -каналам Москва-Пекин, Москва-Берлин [Калачев 2001]. Был использован 8-канальный вокодер с добавлением к сигнал-параметрам полосы спектра естественной речи 200-600 Гц — основного канала, что обеспечило приемлемое качество звучания синтезированной речи. Такая комби-

нация параметров получила название полувокодер. Позднее было показано, что для получения коммерческого качества вокодерной связи полоса естественной речи должна быть расширена примерно до 900 Гц. [Schroeder, David 1960; Кубышкин, Халышкин 1971].

Полувокодер М. Шредера содержал основной канал в полосе частот 80 ... 2000 Гц и девять спектральных каналов в полосе частот 2000 ... 10 000 Гц. Сигналы полувокодера передавались по каналу связи в аналоговом виде в полосе частот 300 ... 2700 Гц. В синтезаторе спектр сигнала возбуждения выравнивался с помощью предельных ограничителей. В результате была получена синтезированная речь с более высоким качеством по сравнению со звучанием речи в полосе частот 300 ... 2700 Гц. При основном канале в полосе частот до 900 Гц качество синтезированной речи считается отличным и в полосе до 500 Гц — хорошим. Нетрудно подсчитать, что пропускная способность канала связи для передачи сигнала основного канала составляет в рассмотренных случаях примерно 900*2*8 ~ 16 кбит/с и 500*2*6 = 6 кбит/сек. Для снижения требуемой пропускной способности передачи до 2400 бит/с сигнал основного канала подвергается специальной обработке, включая фильтрацию и клиппирование. Однако, как показано ниже, качество синтезированной речи при скорости передачи сигнала основного канала ниже 5...6 кбит/с не превышает качества речи вокодера с точным значением сигнала ОТ и даже уступает ему.

В 80-х годах усовершенствованные полувокодеры заняли лидирующее положение в коммерческих сетях цифровой связи, например УКВ радио — сетях сотовой связи МТС со скоростью передачи 8…13 кбит/с. Вокодеры в аналоговом режиме работы для уплотнения ТЧ каналов не нашли широкого применения из-за высокой стоимости оборудования и сравнительно низкого качества речи, оцениваемого как механический (вокодерный) голос. Качество синтезированной речи страдало из-за неточности выделения сигнала основного тона (при использовании пикового или фильтрового методов); сильные линейные и частотные искажения в речевой сигнал вносил угольный микрофон; дополнительные искажения создавали полосовые фильтры с большой крутизной характеристик затухания, нелинейные амплитудные характеристики модуляторов и других узлов вокодера. Техническая реализация вокодера на существовавшей элементной базе отличалась громоздкими массогабаритными характеристиками и большим энергопотреблением.

Второй этап (1953 — 1974 гг.)

Была показана возможность существенного улучшения качества и разборчивости синтезированной речи путем точного выделения основного тона при использовании контактного микрофона, расширения анализируемой полосы частот с 300 ...3400 до 70 ... 7000 Гц путем применения динамического микрофона и увеличения числа каналов в спектрально-полосном вокодере до 16-20. Значительный выигрыш в требуемой скорости передачи сигналов вокодеров был получен с помощью методов эффективного кодирования. Например, при использовании метода динамического кодирования скорость передачи могла быть снижена с 2,4 до 1,2 кб/с при разборчивости S = 82 % и приемлемом качестве синтезированной речи.

В этот период были опубликованы монографии: М.А. Сапожкова (1963); А.А. Пирогова (1974); Дж. Фланагана (1968); Н.Б. Покровского (1962) Н.Г. Загоруйко (1970) и Г. Фанта (1970). Приведем перечень отечественных ученых, работы которых приведены ниже в списке литературы: Н.Н. Акинфиев; С.П. Баронин; И.М. Литвак; Г.А. Коротаев; В.Е. Муравьев; А.А. Пирогов; Р.К. Потапова; В.Н. Соболев; Г.И. Цемель.

Значительно улучшились схемотехнические решения большинства узлов вокодера: усилителей, детекторов, модуляторов. В качестве полосовых фильтров (ПФ) стали использоваться фильтры Бесселя второго-третьего порядка с затуханием в точке пересечения характеристик смежных фильтров 3 дБ и с линейными фазовыми характеристиками. В методах обработки и передачи сигнал — параметров вокодера по каналам связи лидирующее положение заняли цифровые методы. Было обеспечено кодирование параметров речевого сигнала в динамическом диапазоне 60 дБ с шагом 2 ... 3 дБ. Хорошая разборчивость при приемлемом качестве была получена при скорости передачи 2,4 кбит/с.

Ортогональные и формантные методы параметрического кодирования обеспечивают более эффективное сокращение избыточности речевого сигнала, что приводит, однако, к усложнению оборудования. Устройства непосредственного кодирования обеспечивают более высокое качество передачи по сравнению с вокодерами при скорости передачи выше 16 кбит/с, а при скорости менее 8 кбит/с проигрывают последним.

Качество синтезированной речи и скорость передачи

Рассмотрим факторы, воздействующие на качество синтезированной речи на примере спектрально-полосного вокодера. Согласно статистическим данным для передачи огибающей каждого спектрального канала требуется полоса частот шириной около 25 Гц, для передачи сигнала основного тона — 50 Гц. Таким образом, для передачи сигнал — параметров десятиканального вокодера требуется канал связи с полосой пропускания 10 * 25 + 50 = 300Гц и с учетом реальных частотных характеристик фильтров — около 450 Гц. Для передачи этих же сигналов в дискретной форме требуется около 2400 бит/с (при кодировании временных огибающих спектральных каналов трехзначным кодом с частотой отсчетов 50 Гц, сигнала функции возбуждения восьмизначным кодом с частотой отсчетов 100 Гц.).

Полоса пропускания фильтров спектрально-полосных вокодеров должна быть близка к ширине полос равной разборчивости [Покровский 1962; Михайлов 1972]. Такое распределение спектральных каналов по частоте обеспечивает более высокую разборчивость синтезированной речи. Однако из-за того, что в полосе пропускания фильтра на звонких звуках может оказаться несколько гармоник основного тона, качество речи снижается. Например, при равномерном спектре и попадании в канал одной-двух гармоник измеренные значения амплитуды будут отличаться на 3 дБ. При узкополосных фильтрах число каналов может оказаться непомерно большим. Точность измерений временной огибающей можно повысить использованием фильтров с равной полосой пропускания, имеющих одинаковое время замедления. Это существенно при измерениях на участках сигнала с быстрыми изменениями спектра.

Параметры спектральных каналов кодируют путем поочередного подключения выходов ФНЧ через коммутатор к кодирующему устройству — аналого-цифровому преобразователю (АЦП). При последовательных отсчетах сигнал-параметров огибающая речевого сигнала может передаваться с размыванием фронта сигнала. При этом сигнал на выходе вокоде-

ра удлинится, а паузы сократятся. Поэтому отсчеты сигналов всех спектральных каналов в анализаторе осуществляют одновременно, а в синтезаторе одновременно подают на модуляторы. Конечно, требование равного времени замедления и линейности фазовой характеристики гребенки ПФ остается.

Приведем результаты испытаний 19-канального полосного вокодера, полоса пропускания фильтров которого соответствовала полосам равной разборчивости для мужских голосов русской речи. Было проведено сравнение разборчивости и качества синтезированной речи для двух вариантов распределения числа каналов вокодера: шесть каналов в области частот ниже 800 Гц, семь в области 800...2200 Гц и три в области частот выше 2200 Гц (1-й вариант); три, семь и шесть соответственно (2-й вариант). Число каналов изменялось путем объединения соседних фильтров по два. Получено, что вокодер, построенный по первому варианту, дал разборчивость слогов на 3 % ниже, а качество речи по пятибальной шкале оценок методом парных сравнений — почти на целый балл выше, чем вокодер с полосными фильтрами по варианту 2. В связи с этим обычно в области частот ниже 1000 Гц используют пять-восемь равноширокых узкополосных фильтров, а выше этих частот полосу пропускания фильтра расширяют в соответствии с шириной полос равной разборчивости. При выборе характеристик полосовых фильтров исходят из того, что с увеличением крутизны затухания вне полосы пропускания повышается точность измерения статистических спектров. Одновременно возрастает время переходных (неустановившихся) процессов в ПФ, что искажает быстрые спектральные изменения и приводит к реверберации синтезированной речи. По этой причине широкое применение в вокодерах находят фильтры Бесселя третьего-шестого порядков, которые при хорошей частотной разрешающей способности имеют монотонную временную характеристику без всплесков, а фильтры Баттерворта второго порядка и фильтры Чебышева неприемлемы.

Рекомендуемые фильтры имеют затухание около 16...20 дБ на средней частоте смежного фильтра, в точке пересечения характеристик затухания смежных фильтров 3 дБ, линейную фазовую характеристику в полосе пропускания с отклонениями 5 %. Неравномерность времени замедления гребенки ПФ — не более 1 мс. Полоса пропускания ФНЧ должна обеспечить передачу временных изменений сигналов в спектральных каналах при достаточно сильном подавлении флуктуаций гармоник основного тона с частотой около 60 Гц. Вместе с тем полоса пропускания ФНЧ определяет степень компрессии полосы спектра речи. Статистические измерения спектра временной огибающей на выходе детекторов спектральных каналов показали, что 99 % всей энергии лежит в полосе 25 Гц, поэтому полоса пропускания ФНЧ должна составлять 25 Гц.

Результаты синтеза в значительной мере зависят от характеристик полосовых фильтров. В целом к этим фильтрам предъявляются те же требования, что и к фильтрам анализатора. В качестве примера рассмотрим характеристики двух выходных смежных фильтров. Характеристики затухания и фазовые характеристики должны обеспечивать такое сложение сигналов с фазами, чтобы их векторная сумма равнялась сигналу в полосе пропускания каждого фильтра. При этом спектральная огибающая сигнала на выходе синтезатора не будет искажена. Фаза составляющей сигнала на выходе фильтра приблизительно равна фазе составляющей сигнала возбуждения на его входе.

Кратковременные флуктуации спектра, вызванные изменением длительности соседних периодов основного тона, проходят на выход модулятора и вносят искажения в синтезированную речь, поэтому предложено перед модулятором выравнивать спектр с помощью предельного ограничителя. Тогда на выходе модуляторов будет серия прямоугольных импульсов равной энергии. Это позволит упростить схему фильтров (на выходе ограничителя нет четных гармоник), применить импульсный модулятор и одновременно улучшить качество звучания синтезированной речи. На вход ограничителя к тональному сигналу иногда подмешивается шумовой. При наличии сигнала ОТ этот шум подавляется ограничителем, а без ОТ проходит через него. В результате переход от звонких звуков к глухим происходит более плавно, что улучшает звучание.

Отметим, что по сигнал-параметрам 19-канального полосного анализатора могут быть измерены положения (моменты) первых трех формант — в области частот 180 ... 800 Гц (каналы 1-7), 800 ... 2100 Гц (каналы 8-14) и 2100 ... 3900 Гц (каналы 15-19), которые управляют настройкой формантного синтезатора. Такой метод анализа-синтеза получил название формантного вокодера (возможны и другие методы поиска спектральных максимумов, например метод линейного предсказания, рассмотренный ниже). Требующаяся скорость С для передачи сигнал-параметров 19-канального вокодера при применении для кодирования амплитудных значений 4-элементного кода и включая затраты на кодирование сигнала ОТ и ТШ около 1 кбит/с составит:

$$C_n = 19 \cdot 50 \cdot 4 + 1000 = 5 \ кбит/с,$$

а 3— 5 формантного

$$C_ф = (3...5) \cdot 50 \cdot 4 + 1000 = 2 кбит/с$$

Для передачи и хранения в ЗУ сигнал — параметров вокодера могут быть использованы методы эффективного кодирования, в частности метод динамического кодирования, позволяющий снизить скорость передачи в 1,5–2 раза при сохранении качества речи.

Сигнал возбуждения, сформированный по сигналу тон-шум в виде последовательности импульсов ОТ или шума, обеспечивает синтез гласных и глухих звуков речи. Для синтеза звуков речи со смешанным спектром (звонких согласных, например [*б, в, г, д, з*]) необходим сигнал возбуждения смешанного типа с гармоническим спектром в области нижних частот и шумом в области верхних частот. Поэтому для улучшения звучания звонких согласных синтез области частот свыше 3 кГц всегда осуществляют на шумовом сигнале возбуждения. Заметного зашумления гласных при этом не происходит, так как их энергия лежит в более низкой области частот.

В *табл. 1* приведены результаты анализа звуковых ошибок при приеме слоговых таблиц, переданных через 8-канальный полосный полувокодер и трехформантный вокодер. Использовался широкополосный динамический микрофон с полоса анализируемых частот речи в первом случае 300-3400 Гц, во втором — 150-7000 Гц.

Из таблицы видно, что количество ошибок в восприятии звуков синтезированной речи по всем группам твердых согласных больше в формантном вокодере, а по группам мягких — в полувокодере, что объясняется отсутствием в спектре синтезированной речи последнего частотных составляющих выше 3400 Гц.

Была проведена оценка качества синтезированной речи на выходе указанных вокодеров по методу парных сравнений. В испытаниях участвовали пять дикторов и 6-8 слушателей.

Таблица 1

Результаты анализа звуковых ошибок при приёме слоговых таблиц

	% ошибок, тв/м		
	Формантный вокодер	*Полу - вокодер*	*Транзит по параметрам*
	20,2 / 20,8	*4,9 / 3,8*	*14,8 / 24,4*
	6,3 /14,0	*1,8 /11,2*	*3,0 / 12,6*
	1,3 / 2,6	*6,3 / 4,5*	*3,8 / 6,8*
	- / 5,1	*0,7 / 6,3*	*- / 5,0*
	- / 7,4	*- / 5,1*	*- / 10,6*
	9,6 / 1,2	*5,8 / 3,7*	*14,8/ -*
	4,9 / 5,9	*4,9 / 6,2*	*3,1 / 7,8*
	13,1 / 69,0	*0,9 / 15,5*	*5,2 / 59,7*
	5,7 / 12,3	*2,3 / 8,1*	*4,6 / 8,5*
	6,6 / -	*11,1/ -*	*12,1/ -*
	5,1 / 6,9	*1,5 / 55,6*	*3,5 / 19,5*
	23,5 / 8,0	*13,0 / 6,6*	*18,4 / 4,6*
	3,0 / 13,7	*5,4 / -*	*3,8 / 26,7*
	3,6 /15,5	*2,9 / 24,0*	*6,1 / 19,3*
	5,4 / 6,4	*10,8 / 12,6*	*12,0 / 24,2*
	5,4 / 19,5	*10,2 / 22,0*	*8,0 / 31,3*
	4,9 / 39,0	*4,0 / 11,8*	*7,8 / 18,8*
	- / 55,0	*1,2 /24,2*	*1,0 / 38,5*
	2,9 / 13,9	*9,1 / 2,7*	*4,8 / 6,9*

Примечание. Слоговая разборчивость в исходном режиме 80,8% и 82% и в режимах переприёма: по параметрам — 74%, по НЧ — 58%. Тв/м — твёрдый / мягкий.

Дикторы прочитывали фразы длительностью 5…8 с. Оказалось, что при почти одинаковой разборчивости (80,8 и 82%) качество синтезированной речи на выходе полувокодера в 84% случаев было признано более высоким.

В табл.1 приведены также данные по восприятию синтезированной речи для случая тандемного включения вокодеров — транзитной передачи сигнал-параметров между разными вокодерами. В режиме перекодирования сигналов формантного вокодера в сигнал-параметры полувокодера (переприема по параметрам) качество речи оказалось в 78% случаев более высоким по сравнению с режимом переприема сигнал-параметров по низкой частоте (транзит по НЧ) и лишь в 22% случаев — одинаковым или более низким. Отмечалось, что речь в режиме переприема по низкой частоте оказывается хриплой, а ее прослушивание — утомительным. Гармоническая структура речи при переприеме по НЧ сильно искажена, и гармоники основного тона сливаются с шумом. По точности передачи основного тона режим переприема по параметрам приближается к исходному.

Джиттер основного тона

Известно много элементарных ВОТ на основе выделения основного тона по пикам речевого сигнала (пиковые ВОТ) или по переходам через нуль в клиппированной частотно-ограниченной полосе (фильтровые ВОТ). Выделители ВОТ оценивают частоту колебаний голосовых связок по речевому сигналу с некоторыми ошибками [Мартынов 1957]. Эти выделители весьма чувствительны к частотным и другим видам искажений речевого сигнала Так, точность работы пикового ВОТ существенно снижается при подавлении в речевом сигнале области частот ниже 400 Гц. В отсутствие составляющих в речевом сигнале ниже 400 Гц пиковый ВОТ с детектором на входе дает до 15-20% ошибок в выделении основного тона. Появление ошибок объясня-

ется недостаточной величиной первой гармоники основного тона, а также действием шумов, переносимых при нелинейных преобразованиях в область низких частот. При сохранении низкочастотных составляющих в речевом сигнале в полосе до 160 Гц число ошибок уменьшается в 3-4 раза.

В фильтровом ВОТ полоса пропускания фильтра выбирается такой, чтобы в нее попадала только первая гармоника основного тона. Полоса пропускания полосового фильтра может быть нерегулируемой или регулируемой. В последнем случае значение сигнала, управляющего частотой настройки полосового фильтра, пропорционально частоте основного тона, причем при повышении частоты ОТ полоса пропускания фильтра расширяется. По сигналу тон-шум в паузах речевого сигнала и на глухих звуках сохраняется настройка полосового фильтра, что минимизирует ошибку перестройки фильтра на начальные периоды очередного тонального участка речи. Частота основного тона определяется по числу переходов через ноль в клиппированном сигнале.

Исследования разновидностей фильтрового выделителя дали неудовлетворительные результаты. Ошибки, возникающие при использовании фильтровых схем, объясняются наличием низкочастотных акустических шумов на входе; схемы с управляемым фильтром подвержены перестройке на вторую гармонику. Ошибки в выделении основного тона появляются в начале тональных участков речи и при быстрых изменениях ОТ. Несмотря на большое многообразие фильтровых схем ВОТ, всем им присущ существенный недостаток — низкое качество синтезированной речи из-за ошибок в выделении основного тона 8-15%, и значительных фазовых искажений, создаваемых сглаживанием флуктуаций основного тона.

Это обусловлено как несовершенством ВОТ, так и свойствами колебаний голосовых связок. По данным Л. Доланского (1968) на некоторых участках произнесенных фраз при незавершенности наблюдаются нерегулярности периодов основного тона. С другой стороны, быстрые изменения траектории движения контура ОТ достигающие 6000 Гц/с, особенно на сильно эмоционально окрашенных фразах, затрудняют измерения ОТ. Наконец, сами колебания голосовых связок не являются строго регулярными.

Наблюдаются два вида нерегулярностей: значительные изменения длительности периодов основного тона (на 30...50%) и небольшие флуктуации соседних периодов основного тона. Нерегулярности первого вида возникают из-за неполного смыкания голосовых связок на низком основном тоне в начале и в конце звонких участков, при этом периоды с неполным смыканием голосовых связок чередуются с периодами с полным смыканием. Число нерегулярных периодов в конце звонких участков примерно в 4 раза больше, чем в начале, причем нерегулярные периоды следуют сериями, а в начале — изолированно. Число нерегулярных периодов основного тона в речевом сигнале весьма мало (около 0,3% общего числа периодов ОТ), причем оно значительно колеблется на голосах разных дикторов. С вероятностью 0,36 для мужского голоса и 0,5 для женского периоды основного тона флуктуаций не имели, и с вероятностью 0,85...0,96 отклонения длительности смежных периодов основного тона не превысило 10%. Флуктуации частоты основного тона речи могут быть объяснены двумя причинами: характером колебаний голосовых связок (включая нерегулярность) и состоянием артикуляционного аппарата при речеобразовании. Из-за зависимости (особенно на частоте первой форманты) между состоянием голосовых связок и акустических полостей голосового аппарата начало возбуждения полостей может совпасть с моментом смычки голосовых связок (пиковым значением) или находиться в другой точке периода основного тона. В связи с этим, если измерять длительности периодов ОТ по одной фразе речевого сигнала, то при изменении фазы получим частоту ОТ, заметно отличающуюся от ча-

стоты колебаний голосовых связок. Путем специальной выборки фазы пиков речевого сигнала можно получить минимальное фазовое отклонение от истинного значения фазы импульсов голосовых связок (по сигналу ларингофона), и, следовательно, значение длительности периодов ОТ речи, более точно совпадающее с длительностью периодов колебаний голосовых связок. Вероятность отклонения пикового сигнала ларингофона на величину не более 50 мкс составляет 0,57, а минимальное фазовое отклонение — 0,83. Из приведенных данных следует, что точность измерения параметров основного тона по речевому сигналу может быть повышена при учете рассмотренной особенности речеобразования. Кроме того, отношение пиковых значений речевого сигнала в соседних периодах основного Ai / Ai+1 с вероятностью 0,2 равно единице, с вероятностью 0,8 — (0,9 — 1,1) и с вероятностью 0,95 — (0,7 — 1,3).

Требования к выделителям основного тона

Опишем эксперимент с оценкой заметности джиттера ОТ в вокодере при применении точного выделителя ОТ — датчика колебаний голосовых связок. Описание датчика приведено в работе Ю.Я. Волошенко (1968). Для уточнения данных по заметности искажения сигналов ОТ и ТШ нами был использован высококачественный спектрально-полосный вокодер, описанный выше. Анализатор и синтезатор вокодера имели 19 каналов равноартикуляционной ширины в полосе частот 0,1...7,0 кГц. Для формирования сигнала возбуждения использовался контактный микрофон и пиковая схема формирования импульсов ОТ. Сигнал тон-шум формировался из импульсов основного тона. Анализ осциллограмм сигналов ОТ и ТШ ошибок не обнаружил. Вокодер с указанной схемой формирования сигнала возбуждения обеспечивал высокие качественные показатели: слоговая разборчивость речи 94...96%, качество передачи по 5-балльной шкале 4,3 балла по сравнению с качеством естественной речи в полосе частот 300-3400 Гц.

Измерения заметности искажения сигнала основного тона в синтезированной речи выполнялись на голосах пяти дикторов-мужчин (со средней частотой основного тона 128 Гц и диапазоном возможных значений 58...238 Гц) и трех женщин (со средней частотой основного тона 256 Гц и диапазоном 135-522 Гц) с участием шести-десяти слушателей. Испытательный материал состоял из пяти пар фраз длительностью 5...8 с каждая. Устройство введения ошибок работало в двух режимах. В режиме 1 ошибки вводились непрерывно в течение всей длительности звонких участков, а в режиме 2 — лишь в начале звонких участков со средней длиной участка речи с ошибками в сигнале основного тона 50 или 25 мс.

Оценка заметности искажений сигнала ОТ проводилась методом парных сравнений, причем для сравнения использовался режим с передачей сигнала ОТ без искажений. В испытаниях участвовали семь дикторов (в том числе четыре мужчины и три женщины) и 10 слушателей. Заметность искажений сигнала основного тона мужских и женских голосах оказалась примерно одинаковой. Искажения сигнала ОТ становятся заметными (величина заметности 57,5%) уже при 0,6% сбитых импульсов, что соответствует примерно 1% искаженных периодов ОТ. Это свидетельствует о высокой чувствительности

слуха к искажениям сигнала ОТ. При дальнейшем увеличении числа ошибок заметность искажений продолжает расти и при 3% искаженных периодов ОТ достигает 70%.

При искажении основного тона только на начальном участке звонких звуков чувствительность слуха значительно уменьшается. Так, при искажении менее 4% импульсов ОТ (6% периодов) на начальных участках звонких звуков длиной 25...50 мс качество речи не ухудшается. При увеличении числа ошибок до 12% искаженных периодов заметность искажений равна 63% для $t_{ср} = 25$ мс и 70% для $t_{ср} = 50$ мс. Получено, что чувствительность слуха к искажению сигнала основного тона падает при уменьшении длины начального искаженного участка.

Из сказанного следует, что для оценки выделителей основного тона недостаточно иметь данные о средней точности, поскольку для качества речи не безразлично как распределены ошибки. Поэтому следует пользоваться двумя видами оценок:

1. Средней точностью выделения основного тона на начальных участках звонких звуков длительностью, например, $t_{ср} = 25$ мс.

2. Средней точностью выделения основного тона на средних участках звонких звуков (без учета ошибок на начальных участках).

Была исследована заметность флуктуаций (джиттера) временного положения импульсов ОТ. Для внесения флуктуаций использовалась схема квантования временного положения импульсов ОТ. Смещение $dt = \frac{1}{f}$, где f — частота квантования. В испытаниях участвовала та же бригада дикторов и слушателей, что и при описанном выше измерении. Искажения заметны при флуктуации временного положения импульсов ОТ на мужских голосах на 90 мкс, на женских — на 32мкс. Соответствующая частота квантования равна 11 и 30 кГц.

Разница в минимальной допустимой частоте квантования сигнала ОТ для мужских и женских голосов объясняется тем, что средняя частота основного тона женских голосов на октаву выше, чем частота основного тона мужских. Интересно отметить, что отклонение длительности периода ОТ на 90 мкс, соответствует изменению его мгновенной частоты на 1,2 Гц (при средней частоте ОТ 128 Гц), то есть на (1-0,6) % от частоты ОТ. Эти значения пороговой чувствительности слуха к изменению основного тона в 2 раза превышают значения 0,3 — 0,5%, полученные Дж. Фланаганом и очень близки к данным Э. Цвикера по заметности девиации тона с частотой 100-500 Гц, равной 2 Гц. Отличие можно объяснить тем, что первые измерения проводились на протяженных синтетических гласных без динамических изменений других речевых параметров; последние, вероятно, оказывают маскирующее действие на слуховое восприятие ОТ реального речевого сигнала.

Для построения системы передачи вокодера представляют интерес данные по заметности частоты отсчетов сигнала основного тона Fo, с которой передаются значения ОТ, а промежуточные значения в интервале 1/Fo опускаются. На приеме в синтезаторе эти периоды ОТ заменяются переданными значениями ОТ. Минимальная допустимая частота отсчетов сигнала ОТ на мужских голосах равна 60 Гц; а на женских — 150 Гц.

Представляют интерес данные по заметности совместного воздействия частоты отсчетов сигнала основного тона *Fo* и квантования положения импульсов *f*. Испытания проводи-

лись при квантовании ОТ с частотой f = 1,9; 3,8; 7,5 и 15 кГц (соответствующая значность кода равна 5, 6, 7 и 8) и F = 60, 100 и 150 Гц. По трем измерениям, относящимся к разной частоте отсчетов были вычислены средние значения заметности. Можно видеть, что при одновременном квантовании сигнала основного тона с частотой *fкв* и взятии отсчетов с частотой *Fo* = 60-150 Гц допустимая частота *fкв* снизилась для мужских голосов с 11 кГц примерно до 7 кГц, а для женских — с 25 до 15 кГц. Следует отметить, что снижение частоты отсчетов от 100 до 60 Гц мало сказалось на заметности искажений синтезированной речи, особенно при низкой частоте квантования. Видимо квантование создает искажения, которые маскируют искажения из-за частоты *Fo*. При *Fo* = 150 Гц заметность искажений на женских голосах несколько снижается. Из приведенных данных следует, что для неискаженной передачи сигнала основного тона речи на мужских и женских голосах требуемая пропускная способность канала связи равна 8 * 150 = 1200 бит/с и только на мужских голосах — 7 * 60 = 420 бит/с.

Момент начала колебаний голосовых связок не остается постоянным для одной и той же звуковой последовательности, произнесенной многократно одним диктором. В связи с этим можно предположить, что временный сдвиг момента изменения вида сигнала возбуждения (звонкий-глухой) не оказывает существенного влияния на восприятие речи. Для оценки заметности подобных искажений были проведены эксперименты с синтезированным речевым сигналом. В первой серии экспериментов источник гармонического спектра включался вместо источника шумового спектра с задержкой, равной *(0 — dt)* мс, что приводило к оглушению начальных участков звонких звуков средней длительности $t = \frac{dt}{2}$ мс.

Предварительный анализ показал, что различия в заметности для голосов мужчин и женщин нет. Заметность искажений составляет 57% при оглушении начальных участков звонких звуков речи средней длительности 10 мс и максимальной 20 мс. Отметим, что вытеснение участков сигнала длительностью до 11 мс с заменой вытесненных участков шумовым сигналом незаметно на слух, поэтому величина 20 мс представляется достаточно обоснованной. Вместе с тем для повышения качества синтеза звонких согласных и сонантов рекомендуется сигнал возбуждения в области частот выше 3 кГц оставлять шумовым.

Таким образом, параметрические модели анализа-синтеза речи могут обеспечить адекватность звучания естественной и синтезированной речи при выполнении определенных требований по точности работы узлов вокодера.

Многоканальный выделитель основного тона

Успехи в создании высокоточных ВОТ достигнуты на основе анализа речевого сигнала с помощью нескольких элементарных ВОТ и применения для статических измерений цифровых методов обработки. Один из наиболее совершенных выделителей был предложен Б. Голдом [Gold 1962]. Схема состоит из фильт-

ра с полосой пропускания 180...900 Гц, блока измерения максимальных значений, шестиканального пикового ВОТ и блока статистической обработки и принятия решения. В блоке измерения из речевого сигнала выделяются шесть импульсных последовательностей: максимальные (положительные) значения речевого сигнала f_i; разность максимальных и минимальных (отрицательных) значений $f_i - g_i$; разность максимальных соседних значений $f_i - f_{i-1}$; минимальные значения $g_i - f_i$ и, наконец, разность соседних минимальных значений $g_i - g_{i-1}$. Из шести последовательностей с помощью элементарных выделителей, построенных по схеме управляемого пикового детектора, выделяют основные (главные) максимумы, интервалы между которыми T_{ij}. Длительность горизонтальной части бланкирующего сигнала t_i и постоянная времени затухания пропорциональны среднему значению интервала между импульсами основного тона в последовательности T_{ij}, причем $t_i = 0{,}4 \cdot T_{ij}$, $1{,}4 \cdot T_{ij}$, $1{,}4\, T_{ij}$. Амплитуда бланкирующего сигнала пропорциональна среднему значению основных пиков сигнала на входе схемы сравнения. В результате на выход индивидуального ВОТ проходят импульсы, амплитуда которых превышает амплитуду бланкирующего сигнала.

Импульсные последовательности с выхода элементарных ВОТ с частотой квантования 10 кГц подаются на блок статистической обработки и принятия решения, причем три последних интервала T_{i1}, T_{i2}, T_{i3}, каждой последовательности (всего 18 чисел) поступают в ЗУ. Дополнительно в запоминающее устройство записывается еще 18 чисел, представляющих собой суммы $T_{i1} + T_{i2}$; $T_{i2} + T_{i3}$; $T_{i1} + T_{i2} + T_{i3}$. Для примера в *табл. 2, 3* приведены исходные и вычисленные значения *Tij*.

Исходные и вычисленные значения Tij

Таблица 2

	T_{ij} при i					
J	*1*	*2*	*3*	*4*	*5*	*6*
1	*5,25*	*5,80*	*5,25*	*6,10*	*5,54*	*6,10*
2	*1,50*	*5,95*	*11,73*	*5,54*	*6,10*	*5,54*
3	*4,00*	*6,80*	*21,30*	*7,50*	*5,54*	*27,20*
4	*6,75*	*11,75*	*16,98*	*11,64*	*11,08*	*11,64*
5	*5,50*				*12,75*	*33,03*
6	*10,75*	*18,55*	*38,78*	*19,14*	*16,62*	*32,84*

Таблица 3

	ni при Ti1					
Wk, mc	*1*	*2*	*3*	*4*	*5*	*6*
0	*2*	*1*	*2*	*4*	*4*	*3*
0,2	*2*	*1*	*3*	*5*	*5*	*3*
0,4	*2*	*1*	*4*	*5*	*5*	*4*
0,6	*7*	*10*	*5*	*8*	*8*	*5*
0,8	*7*	*10*	*5*	*8*	*8*	*5*
1,0	*7*	*10*	*5*	*8*	*8*	*5*
1,2	*8*	*12*	*10*	*12*	*12*	*10*

Каждый столбец таблицы соответствует *i*-му выделителю, а строка — *j*-му интервалу между импульсами. Далее для каждого *Ti1* подсчитывается число совпадений *ni* из 36 чисел в пределах семи значений допустимого отклонения (окна) *Wk = m (0-1,2)* мс, где *k = 0 — 6*; *m = 0,5*; 1 и 2 мс для *Ti = 3,1...6,3; 6,3...12,7* и *12,7 ... 25,4* мс. Полученные результаты для *m = 0,5* мс приведены в табл. 3. Например, значение *T2*1 в пределах окна *W1= 0* встретилось один раз и в пределах *W4 = 0,6* мс — 10 раз. Далее показания табл. 3 в соответствии с весовой функцией окон Wk уменьшают на 0,1 ... 6 единиц соответственно. Наконец, за истинное значение основного тона принимается значение *Ti1*, которое получило наибольшее число ni при любом окне Wk. По данным табл. 3, например, истинным значением основного тона является интервал *T21= 5,8* мс, имеющий при *W4 = 0,6* мс, *n4 = 7*. Для остальных *Ti* число *ni* меньше при всех *Wk*

Описанный ВОТ первоначально был промоделирован на ЭВМ ТХ-2 в лаборатории Массачусетского технологического института (США), а затем реализован в виде устройства на микросхемах объемом около 5 дм3. Качество синтезированной речи вокодера оказалось сравнимым с качеством речи полувокодера с основным каналом в полосе частот до 500 Гц. Синтезированная речь при использовании ВОТ звучит более натурально, повышается узнаваемость голоса диктора. Такой результат получен благодаря использованию сложных процедур статистической обработки и принятия решения по сравнению с другими известными ВОТ. Нетрудно подсчитать, что при i = 6 (число индивидуальных выделителей), j = 6 (число основных и производных периодов основного тона, участвующих в статистической обработке), k = 7 (число групп допустимых отклонений) и при частоте отсчетов 150 Гц блок обработки должен выполнять до 0,5 млн.опер./с.

Отметим особенности схемы многоканального пикового ВОТ. Процесс обработки речевого сигнала максимально приближен к процессу визуальной обработки осциллограмм речи, принятому за идеальный. В основу ВОТ положен пиковый детектор, фиксирующий положение пиков речевого сигнала без фазовых искажений. При статистической обработке исследуется периодичность пиков речевого сигнала в шести импульсных последовательностях с учетом возможного быстрого изменения длительности смежных периодов ОТ. При использовании динамического микрофона выделитель почти не давал ошибок, хотя на выходе индивидуальных ВОТ имелись ошибки в виде пропуска, добавления и сдвига фазы отдельных импульсов. Число ошибок возрастает на начальных участках звонких звуков, при наличии шумов и подавлении в речевом сигнале низкочастотных составляющих. При использовании угольного микрофона число ошибок составило 8%. Широкое применение интегральных микросхем позволило реализовать 10-канальный спектрально-полосный вокодер с многоканальным выделителем ОТ в объёме 0,2 ... 0,5 дм3.

Благодаря успехам микроэлектроники, в разработках вокодеров стали широко использоваться ИМС и БИС, что позволило применить при ограниченных массогабаритных характеристиках аппаратуры весьма сложные алгоритмы параметрической обработки речевого сигнала на основе методов линейного предсказания и гомоморфной фильтрации. Появились предложения по созданию высококачественного вокодера на основе метода линейного предсказания. Стандартизированный в США вокодер LPC-10 характеризу-

ется удовлетворительным качеством синтезированной речи и разборчивостью односложных слов W = 94 %.

Был реализован вокодер LPC на одной плате площадью 1,2 дм2 и потребляемой мощностью 5,5 Вт. Число коэффициентов предсказания 15, число микросхем 16 (в том числе 7 микропроцессоров MPD 7720 фирмы NEC, Япония). Фирма Motorola (США) изготовила портативный полудуплексный вокодер LPC-10 массой 2,86 кг, объёмом около 2 дм3. Полоса анализируемых частот 120 ... 3800 Гц, микрофон электростатического типа.

На сравнительную оценку качества передачи вокодеров разного рода оказывают влияние многие факторы: состав дикторов (диапазон основного тона, формантное распределение на мужских и женских голосах), характер испытательного материала (дикторский текст, стандартные фразы, телефонный разговор), метод оценки (по разборчивости, качеству передачи, узнаваемости, точности работы отдельных узлов). Определенную роль играют область применения вокодера и сложность оборудования. Получение комплексных оценок сопряжено со значительными материальными затратами.

По сложности технической реализации вокодеры приближенно оцениваются следующим образом (по сравнению с адаптивным дельта модулятором, сложность которого принята за единицу): полувокодер — 50; вокодер полосный (ортогональный, липредер) — 100 — 200, формантный — 500; фонемный (только синтезатор) — 100 . Наиболее высокое качество обеспечили полувокодеры, а наиболее низкое — формантный вокодер. Остальные вокодеры заняли промежуточное положение.

Третий этап (1975 — 1987 гг.)

Этот этап характеризуется поиском путей совершенствования методов моделирования речевого сигнала, которые должны привести к созданию вокодера, обеспечивающего хорошее качество синтезированной речи при низкой скорости передачи. Появление в 80-х годах на рынке цифровых сигнальных процессоров (DSP — Digital Signal Processor), волоконно-оптических кабелей и лазерных технологий привело к развитию новых методов обработки речевого сигнала, и в результате к снижению весогабаритных и стоимостных характеристик аппаратуры.

Обнадеживающие решения предложены в работах по стохастическому кодированию [Atal 1986]. Высокая сложность алгоритма обработки, требующая более 30*106 опер/с, уже не является препятствием для его реализации. Такой метод улучшения качества синтезированной речи может потребовать дополнительного увеличения скорости передачи сигналов вокодера, поэтому при разработке системы передачи необходимо использовать методы эффективного кодирования, в частности метод векторного кодирования, состоящий в передаче на приемную сторону ограниченного числа образцов сигнала (около 1000 ... 1500). Этот метод позволил снизить скорость передачи модели LPC-10 с 2,4 до 0,8 кбит/с без заметного ухудшения качества. Вокодер реализован на однокристальном микропроцессоре TMS 32010 [Fette 1984].

Методы эффективного кодирования имеют существенное значение для повышения помехоустойчивости передачи. Так, в упомянутом векторном вокодере введение трехкратного избыточного кодирования для передачи со скоростью 0,8 * 3 = 2,4 кбит/с, а также методов обнаружения и исправления ошибок в сигнале на приеме путем использования лингвистических закономерностей по возможному изменению значений параметров

в течение трех последовательных интервалов отсчетов позволило обеспечить устойчивую связь при помехах в канале 5*10-2.

Параметрические модели используются в глубоководной телефонной связи с водолазами. Дело в том, что водолазы дышат гелиокислородной смесью при повышенном давлении. При давлении 6,3 атм., в составе дыхательной смеси 80 % гелия, 5 % — кислорода и 15 % — азота, при этом огибающая спектра первой форманты смещается вверх по частоте в 2 раза, вторая и третья форманты — в 1,5 раза. После нескольких дней пребывания в подводной лаборатории смещение формант становится меньше за счет артикулярной компенсации, однако речь остается полностью неразборчивой. Отметим, что частота основного тона не подвержена аналогичным изменениям. Вокодер для связи с водолазами имеет 30-34 канала шириной 100 ... 150 Гц. Величина компрессии исходного спектра не превышает 1,5 ... 1,8. При этом фильтры синтезатора речи расположены в области частот естественной речи, а анализатора — в 1,5-2 раза выше по частоте. Для образования сигнала возбуждения используется полоска непараметризованной речи.

Особой областью применения параметрического описания речевого сигнала является автоматическая диагностика характеристик каналов телефонной связи [Михайлов 1973].

Успехи микроэлектроники привели к развитию новых методов обработки речевого сигнала и снижению весогабаритных и стоимостных характеристик аппаратуры. Формантные и полосные вокодеры применяются для цифровой передачи телефонных сигналов по коротковолновым каналам связи. В качестве примеров можно привести формантный вокодер фирмы Melpar (США), полосный вокодер фирмы Philco (США) и ортогональный вокодер Пирогова. Первый вокодер предназначен для работы по КВ каналам радиосвязи со скоростью передачи 1000 бит/с. Из речевого сигнала выделяются амплитуды и частоты первой, второй и третьей формант и основной тон. Аналоговые параметры занимают полосу частот 140 Гц. Частота отсчетов 43,5 Гц, длина кода для квантования сигнал-параметров — три, для основного тона — пять. Частоты формант определяются с помощью рометров. Область применения фонемных вокодеров — линии командной связи, речевое управление и говорящие автоматы информационно-справочной службы. Вокодеры применяют для кодирования телефонных сигналов в цифровых системах связи вооруженных сил США, а также в цифровых коммерческих системах связи. Перспективно применение вокодеров в системах передачи данных для организации технологической служебной телефонной связи со скоростью передачи 1200 ...2400 бит/с.

Современные вокодеры обеспечивают хорошее качество речи при скорости передачи 4,8 кбит/с и пригодное для ведения служебных переговоров качество речи при скорости передачи 1,2 кбит/с. На этой скорости наилучшим по разборчивости и качеству звучания является формантный вокодер. Есть попытки перевести его на скорость 600 бит/с с использованием корреляционных связей между формантами (передается главная форманта, определяющая данный звук речи, а для некоторых звуков две форманты, например, при передаче звуков [э, и, ы]).

Липредер LPC=10

Липредер LPC=10 с линейным предсказанием на 10 коэффициентов со скоростью передачи 2400 бит/с [Tremain 1982] разработан и стандартизован для использования в сетях министерства обороны (DCS — Defence Communication System) и правительственной связи США. Обработка речевого сигнала осуществляется в реальном масштабе времени на основе 16-разрядного сигнального микропроцессора TMS 32010. В полудуплексном варианте исполнения LPC-10 занимает одну плату; потребляемая мощность — 10 Вт. Портативный вариант LPC-10 имеет объём 4 дм3, массу менее 2 кГ (без литиевых батарей), микрофон электростатический; потребляемая мощность — около1 Вт.

Качество синтезированной речи характеризуется как удовлетворительное, а разборчивость слов по таблицам выбора DRT по сравнению с разборчивостью естественной речи в полосе частот 0,1 ...3,6 кГц равна в тишине 90 и 95 % а при наличии акустических шумов, характерных для служебных помещений, постов наблюдения и пр., — 82 и 92 ... 93 % соответственно. Абонентская оценка приемлемости связи ДАМ равна 48 % для синтезированной речи и 65 % для естественной. Устойчивость связи обеспечивается при помехах в канале до 10-2. Возможно тандемное включение липредера в режиме переприема по импульсам с адаптивным кодеком АРС=4, работающим со скоростью передачи 9600 бит/с на основе алгоритма линейного предсказания на четыре коэффициента и совместимым с LPC-10. Сообщается также о разработке совместимого с LPC-10 алгоритма адаптивного дельта-модулятора CUSD для работы со скоростью передачи 16 кбит/с.

Речевой сигнал на входе анализатора ограничивается с помощью полосового фильтра полосой частот 0,1 ... 3,6 кГц, далее берутся отсчеты с частотой 8 кГц и 12-разрядным линейным кодом, спектр сигнала корректируется фильтром первого порядка с характеристикой 1 — 0,9375 z-1. Отсчеты поблочно записываются в буферную память синхронно с частотой ОТ. Коррекция спектра и фазирование загрузки с ОТ снижает требуемую точность вычислений коэффициентов на 4 бита. Коэффициенты вычисляются с удвоенной точностью. Для вычисления коэффициентов отражения применяется метод Чолески. При этом матричное уравнение $[R]\cdot[C]=[Q]$ преобразуется в $[S]\cdot[K]=[Q]$, где

$$r_{ij}=\sum_{n=N+1}^{M}(s_{n-i})\cdot(s_{n-j}),\ s_i-отсчеты$$

$$q_i=\sum_{n=N+1}^{M}s_n\cdot s_{n-j},\ M=130\ (длина\,сегмента)$$

$$s_{ij}=r_{ij}-\sum_{k=1}^{j-1}\frac{s_{ik}\cdot s_{jk}}{s_{kk}}$$

$$k_i=\frac{1}{s_{ij}}\left[q_i-\sum_{j=1}^{i-1}k_i\cdot\Sigma_{ij}\right]$$

— коэффициенты отражения ковариационной матрицы.

Основной тон выделяется корреляционным сдвиговым методом. Для выделения сигнала тон-шум используется метод измерения энергии в низкочастотной полосе, корреляции значений ОТ и числа переходов через нуль.

Сигнал тон-шум имеет четыре градации, которые определяются по соотношению низкочастотной энергии, отношению пик-пауза сдвиговой функции ОТ и переходам через нуль, и посылается в канал с удвоенной частотой отсчетов: 00 — глухой, 11 — звонкий, 01 — переход глухой-звонкий, 10 — переход звонкий-глухой. Коэффициент усиления передается пятиразрядным кодом с шагом квантования 1,5 дБ.

Синтезатор представляет собой рекурсивный фильтр, реализованный на 16-разрядном процессоре. На приеме параметры сглаживаются по их значениям в трех соседних кадрах. На выходе синтезатора включен фильтр с полосой пропускания 0...3600 Гц.

Отметим некоторые особенности алгоритма LPC-10.

1. Алгоритм реализован в реальном масштабе времени на основе сигнального микропроцессора TMS-32010 с быстродействием 5*106 опер/с и циклом 200 нс. Спектральный анализ занимает 37% машинного времени, синтез — 39 % и выделение основного тона — 20%. Для программирования TMS-32010 использован язык Вакс-микро. Путем моделирования алгоритма на универсальной ЭВМ был выбран ковариационный метод блочного кодирования, который позволяет получать хорошее описание формантной структуры при сокращении на 66% объема вычислений по сравнению с корреляционным методом (130 отсчетов вместо 180 на сегменте, равном 22,5 мс). Общее число ИМС — 20. Для уменьшения числа ИМС в аппаратных средствах, используемых для подключения ВЗУ к микропроцессору TMS-32010, требуется расширить его ОЗУ, равное 4 Кбайт 16-разрядных слов, а также усовершенствовать устройства ввода-вывода речевого сигнала, включая АЦП и ЦАП.

2. Элементы матрицы вычисляются синхронно с основным тоном, причем анализируемые интервалы речевого сигнала кратны длине ОТ на звонких участках и равны 22,5 мс на глухих.

3. Перед выделением ОТ речевой сигнал проходит через низкочастотный фильтр Баттерворта четвертого порядка с граничной частотой 800 Гц и через инверсный фильтр второго порядка, что необходимо для улучшения выделения ОТ из речевого сигнала с подавленными составляющими ниже 300 Гц. Затем определяется значение основного тона в момент времени t по величине корреляции между отсчетами речевого сигнала согласно

$$\sum_{n=1}^{130} \frac{s(n) - s(n+t)}{2}$$

Замена операций умножения на вычитание и квадрирования на сумму абсолютных разностей снизило объём вычислений примерно в восемь раз.

Кодирование периодов основного тона выполняется по шестиразрядной полулогарифмической шкале с 20 градациями на октаву в диапазоне частот ОТ 50 ... 400 Гц. Дополнительный седьмой разряд является корректирующим и используется на приеме для выявления одиночных ошибок. Точность передачи 2,5 Гц обеспечивается в диапазоне 50 ... 100 Гц, 5 Гц — в диапазоне 100 ... 200 Гц и 10 Гц — в диапазоне 200 ... 400 Гц.

4. В системе передачи LPC-10 приняты следующие методы повышения помехоустойчивости:

сигналы ОТ и ТШ кодируются совместно с помощью специального блочного кода; значение ОТ в блочном коде защищается избыточным кодом, обнаруживающим одиночные ошибки;

ошибки в сигнале ТШ корректируются по значениям сигнала ТШ в трех соседних кадрах;

на глухих участках порядок фильтра предсказания снижается с 10 до 4 и вводится избыточное кодирование для защиты коэффициента усиления и коэффициентов отражения K1…K4;

на звонких участках применяется адаптивный алгоритм сглаживания сигналов ОТ, G и K1…K6.

На реализацию перечисленных методов требуется только 1% машинного времени МП TMS-32010 и дополнительная память на 670 байт.

Разборчивость слов по таблицам DRT при отсутствии помех в канале сохраняется неизменной, а при помехах 5 % увеличивается от 71 до 82% (по сравнению с режимом без защиты). При этом приемлемость связи DAM характеризуется показателями 33 и 38% соответственно.

Рассмотрим особенности методов повышения помехоустойчивости передачи. Для блочного кодирования сигналов ОТ и ТШ использован 7-значный код, причем 60 значениям соответствует табличная величина ОТ, восьми значениям — табличные значения шума, восьми — переход тон-шум и остальным 50 значениям — запрещенное (ошибочное) значение ОТ. Благодаря блочному кодированию вероятность пропуска ошибочного значения ОТ снижается почти в 2 раза, а вероятность ложного перехода тон-шум почти в 1000 раз, так как решение принимается по значению параметра в трех соседних кадрах. При обнаружении запрещенной комбинации (ошибки) в сигнале ОТ для синтеза используется среднее значение ОТ:

$$y(n) = \frac{15}{16} \cdot y(n-1) + \frac{1}{16} \cdot x(n)$$

где $y(n-1)$ — среднее значение ОТ; $x(n)$ — текущее (последнее неискаженное) значение ОТ.

Повторение последнего принятого значения ОТ может привести к заметным на слух искажениям речи в случае пропуска сигнала ОТ с ошибками. На глухих участках речевого сигнала достаточная точность синтеза обеспечивается при четырех коэффициентах отражения. Освободившиеся кодовые знаки используются для избыточного кодирования параметров K1 — K4 и G. Блочный код Хэмминга (8,4) обеспечивает обнаружение до двух ошибок в 4-разрядном слове, причем одиночные ошибки исправляются, а кодовые комбинации с двумя ошибками заменяются на неискаженные значения этого параметра в соседнем кадре.

На приеме дважды за кадр осуществляется интерполяция отношения логарифма площади Ki. Среднее квадратическое отклонение сигнала G и основной тон интерполируются по линейному закону синхронно с основным тоном. Значения коэффициентов предсказания, усиления, основного тона и тон-шум обновляются каждый период ОТ. Значения

Ki передаются 14-значным кодом. В зависимости от значения помех осуществляется сглаживание параметров OT, G и K1…K6. Для этого на глухих участках речевого сигнала вычисляется средняя величина отклонения параметров xi(n) в двух соседних кадрах на величину, превышающую порог *t*:

$$2 \cdot x(n) - \frac{1}{2} x_i(n-1) > t,$$

$$2 \cdot x(n) - \frac{1}{2} x_i(n+1) > t.$$

Порог *t* изменяется пропорционально уровню помехи: малая помеха — 0,01 %, средняя — 0,1-1,0%, большая — 1-2,0% и очень большая > 2 %. При малых помехах сглаживание отсутствует, что сохраняет исходное качество речи.

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. *Дрёмов А.Н.* Решительный шаг к интеграции // Технологии и средства связи. 2001. № 2.

7. *Калачев К.Ф.* В круге третьем. М., 2001.

8. *Калинцев Ю.К.* Разборчивость речи в цифровых вокодерах. М., 1991.

9. *Коротаев Г.А., Михайлов В.Г.* Синтетическое телефонирование // Радиоэлектроника и электронная техника. 1964.

10. *Коротаев 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.* Расчет 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., Шибаева 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. *Фланаган 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.