ВОКОДЕР 600 БИТ/С

Бабкин В.В.
Центр ЦОС СПб ГУТ, Санкт-Петербург, пр. Большевиков д.22 корп.1.
vb@dsp-sut.spb.ru, (812)589-51-85, www.dsp.sut.ru

Основным фактором, определяющим качество передачи речевого сигнала в низкоскоростных цифровых системах связи, является пропускная способность цифровых каналов, задающая необходимую степень сжатия, а, следовательно, и способ преобразования речевого сигнала при формировании цифрового потока.

Так, при скоростях передачи ниже 4 кбит/с для сжатия речи обычно используются параметрические вокодеры, описывающие входной сигнал на основе различных параметрических моделей синтеза речевых сигналов. Основными характеристиками вокодеров являются: выходная скорость цифрового потока; качество и разборчивости речи; чувствительность к ошибкам в цифровом канале связи; чувствительность к акустическим шумам; зависимость от диктора; алгоритмическая задержка обработки сигнала; вычислительная сложность реализации.

Если рассматривать только вопросы речевого кодирования, то основными задачами при разработке вокодеров являются: выбор модели описания и синтеза речевого сигнала; помехоустойчивая оценка параметров модели на основе анализа текущей речи; выбор числа бит и способа квантования параметров модели; выбор способа интерполяции квантованных параметров на приемной стороне.

Существуют приложения, в которых, по целому ряду причин желательно иметь скорость выходного цифрового потока вокодера не более 600–800 бит/с, а в некоторых случаях – 300 бит/с. Например, это может быть связь с морскими судами, цифровая связь по узкополосным КВ радиоканалам в условиях помех и замираний, разработка помехоустойчивых вокодеров на принципах объединения и совместной оптимизации схем речевого и канального кодирования для работы в каналах с высоким уровнем битовых ошибок [16,17] и т. д.

Основной проблемой построения сверх низкоскоростных вокодеров (300–800 бит/с) является крайне малое количество бит, остающихся для описания речевого сигнала по отдельным кадрам с фиксированной длиной 20–30 мс. Например, при стандартном размере кадра 22.5 мс и скорости 800 бит/с на 1 кадр приходится всего 18 бит. Такого числа бит уже недостаточно для передачи информации о следующем минимальном наборе параметров: энергии сигнала, форме огибающей амплитуд кратковременного частотного спектра сигнала, структуре спектра (признак тон/шум) и частоте основного тона (ОТ) речи для вокализованных кадров. Поэтому качество и разборчивость речи параметрических вокодеров с кадрами фиксированной длины 20-30 мс на данных скоростях вырождаются и становятся неприемлемым, даже для служебных систем связи.

По этой причине для сверх низкоскоростной компрессии речи были разработаны другие подходы: вокодеры с обработкой нескольких кадров фиксированной длины, объединенных в один большой суперкадр [1,2,3,4], вокодеры с переменной длиной кадра на основе сегментации речи [5,6] и фонемные вокодеры [7,8]. Первые используют векторное квантование траекторий параметров для всего суперкадра и обеспечивают повышение качества кодирования за счет динамического перераспределения информационных бит между квантуемыми параметрами и кадрами, входящими в суперкадр. Вторые сегментируют речь и описывают большие однородные речевые фрагменты целиком в пределах естественных границ, а не отдельные небольшие кадры с фиксированными границами. Третьи используют методы теории распознавания образов для выделения элементов звукового алфавита из текущей речи на передающей стороне с последующим синтезом по алфавиту на приемной стороне. Во всех случаях становится возможным создавать приемлемое по точности описание для фрагментов речи с длительностью около 100 мс с помощью 40-80 бит.

Наибольшее развитие в мировой практике получили вокодеры первого типа, использующие суперкадр. В настоящее время за рубежом существует стандарт на вокодеры LPC 800 бит/с [2], вокодеры MELP 600 бит/с находятся в стадии стандартизации [4], а на вокодеры со скоростью 300 бит/с объявлен международный конкурс [9]. Однако, такие стандарты в силу специфики применения являются закрытыми. Поэтому разработка собственных алгоритмов компрессии речи со скоростями 300-800 бит/с представляет большой научный и практический интерес. С

коммерческой точки зрения, разработка оригинальных алгоритмов так же желательна для обеспечения независимости от держателей патентов стандартных решений.

В статье представлен вокодер со скоростью 600 бит/с, разработанный на принципах векторного квантования траекторий параметров для суперкадра. Вокодер состоит из анализатора и синтезатора речи (рис.1).

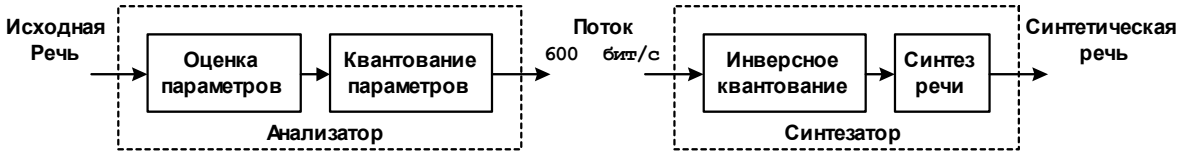

Рис.1 Структурная схема вокодера 600 бит/с.

На передающей стороне, на основе анализа текущей речи оцениваются параметры модели, которые затем квантуются и индексы квантования передаются в цифровой канал связи. На приемной стороне по принятым индексам квантования восстанавливаются квантованные параметры модели, которые далее интерполируются и синтезируется выходной речевой сигнал.

Анализатор и синтезатор можно разделить на две структурно независимые части: параметрическую модель описания и синтеза речевых сигналов, выполняющую анализ/синтез речи, и блок квантования параметров, отвечающий за дискретное описание непрерывных по своей природе параметров модели с помощью ограниченного количества бит.

Точность описания параметров, определяемая квантованием, влияет как на скорость передачи параметров, так и на ошибку синтеза речи. При понижении скорости вокодера за счет более грубого квантования параметров качество синтезированной речи снижается и, в конце концов, вырождается. Однако, максимальное качество синтетической речи (при отсутствии квантования параметров) определяется только моделью синтеза речи, лежащей в основе вокодера. Поэтому, повышая точность описания параметров за счет роста скорости не всегда можно добиться пропорционального улучшения качества, которое, выше определенного уровня перестает заметно расти.

В низкоскоростных параметрических вокодерах с фиксированной длиной кадра 20-30 мс используется множество различных моделей описания и синтеза речевых сигналов, из которых наиболее известны и широко применяются на практике следующие: LPC [10], MBE [11] и MELP [12]. Основным принципиальным отличием последних двух схем, обуславливающем более высокое качество синтезированной речи, является многополосное (мягкое) решение тон/шум в отличии от бинарного (жесткого) решения в вокодерах LPC.

В качестве модели описания и синтеза речи для вокодера 600 бит/с была выбрана MELP модель, осуществляющая анализ/синтез речи во временной области. Исходная MELP модель [12], имеющая длину кадра 22.5 мс, была подвергнута доработке и упрощению. Для описания речевого сигнала используются следующие основные её параметры: кратковременная энергия, частота ОТ для вокализованных звуков, огибающая кратковременного спектра на основе модели линейного предсказания (ЛП) 10-го порядка, признак тон/шум по частотным полосам. Для суперкадров входной речи размером 100 мс вычисляются траектории этих параметров, которые затем векторно квантуются с помощью 59 бит. К выходному речевому пакету добавляется 1 бит синхронизации. Так как хранить значения и осуществлять поиск векторов в кодовой книге с индексом из 59 бит нереально, использовалось расщепленное квантование с делением индекса на несколько кодовых книг по числу квантуемых признаков. Одну из основных трудностей работы составляла разработка схем векторного квантования и обучение квантователей различных типов.

Признак тон/шум в исходной MELP модели оценивается в 5 частотных полосах: 0-500 Гц, 500-1000 Гц, 1000-2000 Гц, 2000-3000 Гц и 3000-4000 Гц методом расчета максимума выборочной оценки коэффициента корреляции сигнала при сдвигах на величину периода ОТ с последующим сравнением с фиксированным порогом. Признак тон/шум требует для описания теоретически 5 бит на кадр, однако, как отмечено в [3], не все возможные комбинации бит встречаются одинаково часто. Для исследования этого факта использовалась обучающая выборка речи длительностью 20 минут, состоящая из фонетически сбалансированных фраз русского языка, начитанных 10 дикторами – 6 мужчинами и 4 женщинами. Статистика поведения многополосного признака тон/шум приведена в таблице, где озвученные полосы обозначены 1, а не озвученные – 0.

Комбинация Тон/шум	11111	10000	11100	00000	11110	11000	Остальные
Частота повторения, %	32.4	26.6	9.8	9.1	7.2	6.0	8.9

Таким образом, можно сделать вывод, что в подавляющем числе случаев достаточно лишь указать границу раздела вокализованной и невокализованной части спектра речи. При скалярном квантовании траектории упрощенного признака тон/шум с дискретизацией, например по четырем точкам во времени, требуется передать всего 6*6*6*6=1296 его возможных комбинаций, для чего достаточно 11 бит вместо 20 бит для исходной схемы. Но так как вокализация речевого сигнала изменяется относительно плавно, соседние точки траектории сильно коррелированны, поэтому векторное квантование способно дать значительный выигрыш по сравнению со скалярным квантованием. В разработанном вокодере для дальнейшего понижения скорости выходного цифрового потока используется векторная кодовая книга с полным перебором размером 4 бита.

Величина частоты ОТ в исходной MELP модели квантуется скалярно, используя 7 бит на кадр. В силу того, что потенциальный диапазон возможных частот ОТ для различных голосов лежит в пределах от 50 до 500 Гц (более трех октав), квантование для выравнивания относительной ошибки осуществляется в логарифмической области. В разработанном вокодере предварительно сглаженная траектория частоты ОТ квантуется с помощью векторной кодовой книги с полным перебором размером 7 бит.

Огибающая кратковременного спектра входного сигнала, полученная на основе модели ЛП 10-го порядка, представлена в виде 10 линейных спектральных пар (ЛСП). Во всех низкоскоростных вокодерах квантование огибающей спектра, даже с использованием наиболее удобного представления параметров модели ЛП в виде ЛСП, является самым узким местом с точки зрения большого числа бит, необходимых для адекватного на слух описания этого параметра. В исходной MELP модели [12] использовалось скалярное квантование 34 бита на кадр, аналогичное вокодеру CELP 4800 бит/с FS-1016, позднее оно было заменено многоступенчатым векторным квантованием (MSVQ) 25 бит на кадр. В [13] приводится тщательное сравнение способов расщепленного векторного квантования (SVQ) и MSVQ для ЛСП модели ЛП 10-го порядка на одиночных кадрах 20-30 мс и показано превосходство способа MSVQ с точки зрения меры спектрального искажения Итакуры при равном общем количестве бит. Это объясняется тем, что подход MSVQ полнее использует внутрикадровую корреляцию компонент векторов ЛСП. Конечно, оба способа уступают по точности квантования полноразмерным кодовым книгам с полным перебором (FSVQ). Однако, реализация последних с индексами порядка 20 бит затруднена из-за больших размеров книг и ограниченной скорости поиска. Даже для SVQ и MSVQ книг полный перебор не используется по аналогичным причинам. Эффективным (субоптимальным) методом поиска индекса наилучшего вектора в MSVQ книгах является подход с выбором M наиболее удачных кандидатов на каждой из ступеней квантования с последующим поиском наилучшего индекса по дереву. Данный метод по точности квантования превосходит последовательный поиск и приближается к полному перебору, однако лишен его недостатков. Таким образом, для передачи информации об огибающей спектра с приемлемой точностью для одиночного кадра необходимо 22–24 бита.

Дальнейшее снижение числа информационных бит возможно за счет использования межкадровой избыточности ЛСП. При квантовании траекторий ЛСП для суперкадров подходят как методы на основе предсказания, так и методы совместного квантования нескольких векторов. Первые последовательно квантуют для каждого кадра ошибку предсказания ЛСП, вторые – объединяют вектора ЛСП для смежных кадров в один и имеют большую эффективность [13]. Плюсом второго подхода является также отсутствие эффекта размножения ошибок, характерного для схем с предсказанием. В представленном вокодере используется подход совместного квантования точек траектории ЛСП и разработана MSVQ кодовая книга с выбором наилучших кандидатов (M=8). Таким образом, для описания траектории ЛСП используется 38 бит.

Величина энергии (усиления) в исходной MELP модели квантуется скалярно два раза за кадр, чтобы передать с меньшими искажениями энергетическую огибающую речевого сигнала. Используется 5+3=8 бит. Квантование осуществляется в логарифмической области для передачи динамического диапазона входных сигналов с заданной относительной точностью. В разработанном вокодере для описания траектории усиления используется векторная кодовая книга с полным перебором размером 10 бит.

Таким образом, подходы квантования траекторий изменения параметров для группы кадров, вместо набора параметров для одного кадра, и использование векторного квантования, вместо скалярного, позволяют понизить общую скорость цифрового потока вокодера с 2400 до 600 бит/с. Обратной стороной такого понижения скорости, помимо контролируемого на тестовых сигналах ухудшения качества и разборчивости речи, является возможное повышение чувствительности к некоторым характеристикам входного сигнала. Поэтому вопросы изучения независимости качества звучания вокодера, например от выбора произвольного диктора, типа используемого микрофона или уровня входных сигналов, требуют дальнейшего изучения.

Модель вокодера 600 бит/с реализована для ПЭВМ на языке Си в арифметике с фиксированной точкой в виде консольного приложения с файловым вводом-выводом сигналов. При желании она может быть реализована в реальном масштабе времени на целочисленных цифровых процессорах обработки сигналов, например, семейства TMS320VC5000.

При испытаниях вокодер показал хорошую разборчивость и качество речи, сопоставимое со звучанием зарубежных образцов. Демонстрационные файлы звучания вокодера 600 бит/с можно послушать на web странице Центра ЦОС СПб ГУТ в разделе «разработки» [14].

Испытания вокодера совместно с шумопонижающим устройством (ШПУ) [15] показали субъективный выигрыш от совместного использования вокодера и ШПУ по сравнению с вокодером без ШПУ при работе в акустических шумах. Особенно перспективным является объединение и совместная оптимизация схем шумопонижения, речевого и канального кодирования для работы в условиях акустических шумов по каналам со скоростями 800-1200 бит/с с высоким уровнем битовых ошибок.

Литература

1. Kemp D., Collura J., Tremain T. “Multi-frame coding of LPC parameters at 600-800 bps”. ICASSP-1991, vol. 1, pp. 609-612.
2. Mouy B., et al., "Nato Stanag 4479: A Standard for an 800 bps Vocoder and Channel Coding in HF-ECCM system", IEEE ICASSP, Detroit, pp. 480-483, May 1995.
3. Chamberlain M. “A 600 bps MELP vocoder for use on HF channels”. MILCOM-2001, vol.1, pp. 447-453.
4. Guilmin, G. Capman, F. Ravera, B. Chartier, F. “New NATO STANAG Narrow Band Voice Coder at 600 Bits/s”. ICASSP-2006, vol.1, pp. 689-692.
5. Peterson P., Jeanrenaud P., Vandegift J. “Improving intelligibility of a 300 b/s segment vocoder”, 1990.
6. Zolfaghari P., Robinson T. “Speech coding using mixture of Gaussians polynomial model”, EUROSPEECH-1999, pp. .
7. M. Padellini, F. Capman, G. Baudoin. Very low bit rate (VLBR) speech coding around 500 bits/sec. EUSIPCO 2004, pp. 1669-1672.
8. da S. Maia R. at al. “Mixed-excited phonetic vocoding at 265 bps” ICASSP-2003, vol.1, pp. 796-799.
9. DARPA ASE program. http://www.darpa.mil/ato/solicit/ASE/index.htm.
10. Tremain T. “The government standard linear predictive coding algorithm: LPC-10”. Speech Technology, April 1982, pp. 40–49.
11. Griffin D.W., Lim J.S. “Multiband excitation vocoder”. IEEE ASSP-36 (8), 1988, pp. 1223-1235.
12. McCree A., Barnwell III T. “A mixed excitation LPC vocoder model for low bit rate speech coding”. IEEE TSAP vol. 3, No. 4, July 1995, pp. 242-250.
13. Kondoz A.M. Digital Speech. Coding for low bit rate communication systems. J.Wiley & Sons, 2004.
14. Вокодеры 600-7200 бит/с. Разработки Центра ЦОС СПб ГУТ. http://www.dsp.sut.ru.
15. Бабкин В.В. Шумопонижающее устройство для вокодера. Отчеты DSPA-2007.
16. Бабкин В. В., Ланнэ А.А., Шаптала В.С. Оптимизационная задача выбора речевого и канального кодирования. Отчеты DSPA-2005, стр. 123–127, Москва 16-18 марта 2005 г.
17. Бабкин В. В., Ланнэ А. А., Шаптала В. С. Помехоустойчивые вокодеры для систем цифровой радиосвязи в КВ и УКВ диапазонах. Отчеты 1-ой межд. НПК "Исследование, разработка и применение высоких технологий в промышленности", стр. 21-22, СПб, 2005 г.