

ПРОБЛЕМЫ ПОСТРОЕНИЯ СОВРЕМЕННЫХ СИСТЕМ ЦИФРОВОЙ РЕЧЕВОЙ СВЯЗИ

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

Основной функцией систем цифровой речевой связи является передача на расстоянии речевых сообщений между людьми. Основными показателями качества передачи речевых сообщений являются разборчивость речи, определяющая смысл сказанного и степень напряжения слушателя, и субъективное качество речевого сигнала, определяющее естественность, узнаваемость и другие характеристики голоса диктора. Для тестирования качества и разборчивости речи разработано множество субъективных и объективных методов (DRT, MOS, DAM и др.), в частности, используются ГОСТ В 20775-75 и рекомендация ITU-T P.862 (PESQ-MOS).

Главным фактором, определяющим качество передачи речевого сигнала в низкоскоростных цифровых системах связи, является пропускная способность цифровых каналов, как правило, ограниченная, особенно для радиоканалов в диапазонах КВ и УКВ. Она задает способ преобразования и степень сжатия речевого сигнала при формировании цифрового потока. Помимо этого, при эксплуатации аппаратуры связи, часто возникают второстепенные мешающие факторы, дополнительно снижающие качество речевой связи, такие как:

- внешний акустический шум на стороне передачи сообщения,
- битовые ошибки или потери речевых пакетов в цифровых каналах связи,
- внешний акустический шум на стороне приема сообщения.

Поэтому, при проектировании или модернизации систем связи имеет смысл рассматривать в целом всю цепочку преобразований и передачи речевых сообщений от человека к человеку, учитывая обстоятельства реальной эксплуатации, особенно для систем мобильной, подвижной и служебной связи.

В статье рассматривается только один из множества вопросов построения таких систем – вопрос организации цепочки обработки и преобразования речевого сигнала, включая акустический интерфейс с пользователем, с целью достижения максимального сквозного показателя качества с учетом всех мешающих факторов. Другие, не менее важные вопросы, затрагивающие проблемы организации и коммутации цифровых каналов, системы синхронизации, протоколы обмена, защиту информации и т. д., но не относящиеся непосредственно к качеству передачи речевых сигналов, не рассматриваются.

Предлагаемая для обсуждения, обобщенная из различных источников блок-схема преобразований речевого сигнала в современной цифровой системе связи, представлена на рис. 1. Она традиционно состоит из передающей и приемной частей, разделенных цифровым каналом связи с ошибками. Каждая из частей, в свою очередь, состоит из ряда устройств с различной функциональностью, которые могут быть использованы как совместно, так и самостоятельно.

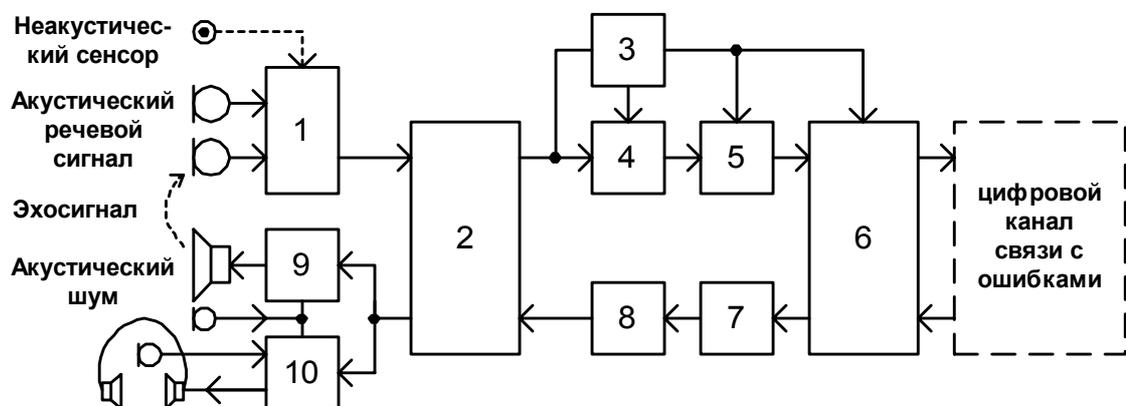


Рис.1. Обработка речи в современной системе цифровой речевой связи.

Рассмотрим передающую часть. Первым ее элементом является устройство преобразования акустического речевого сигнала в цифровую форму (1). Основными требованиями являются: линейность преобразования, низкий собственный шум, динамический диапазон, достаточный для передачи уровней входных речевых (и шумовых) сигналов, формирование необходимой АЧХ

сигнала до АЦП. По диапазону передаваемых частот системы делятся на узкополосные (300–3400 Гц) и широкополосные (50–7000 Гц), с частотами дискретизации 8 и 16 кГц, соответственно.

В условиях внешнего акустического шума важную роль приобретают характеристики пространственной направленности микрофона, которые могут формироваться как конструктивно (например, в микрофоне ДЭМШ-1А), так и программно, с использованием нескольких микрофонов с последующей цифровой обработкой сигналов [1, 2]. В последнем случае, можно адаптивно формировать различные диаграммы направленности, с максимумом в направлении источника речевого сигнала, или с минимумом в направлении источника помех, и тем самым существенно улучшать отношение сигнал/шум (ОСШ) на входе системы, по сравнению с традиционными решениями на базе одного ненаправленного или направленного микрофона.

Также в последнее время активно исследуется вопрос об использовании совместно с микрофонами сенсоров неакустического типа, например, ларингофонов или электрических датчиков колебаний голосовых связок. Ведутся разработки специальных алгоритмов по обработке таких сигналов для помехоустойчивого ввода речи в особо тяжелых шумовых условиях, например, для пилотов реактивных самолетов, с последующим низкоскоростным кодированием [3].

В системах громкоговорящей (hands-free) связи, рекомендованных для водителей транспорта и удобных при организации конференций с несколькими участниками, для подавления мешающего акустического эхосигнала необходим акустический эхокомпенсатор (АЭК) (2) с длиной эхотракта не менее 64–512 мс. Основными проблемами построения таких АЭК, реализуемых как во временной, так и в частотной областях [2, 4], являются: обеспечение скорости сходимости и устойчивости адаптивных фильтров с числом коэффициентов, достигающем нескольких тысяч; проблемы детектирования двойного разговора; нелинейные искажения и нестабильность акустического эхосигнала; вычислительная сложность алгоритмов. Основой для тестирования устройств эхокомпенсации являются рекомендации ITU-T G.164-G.168.

В условиях внешних акустических шумов со спектром, перекрывающимся со спектром речи, можно использовать устройства шумопонижения (ШПУ) (4), построенные на основе нелинейных адаптивных алгоритмов [2,5,6]. Необходимым условием работы таких ШПУ является квази-стационарность, аддитивность и статистическая независимость шума от речевого сигнала, что достаточно часто выполняется на практике, например, для шума равномерно едущего автомобиля.

Считается, что такие ШПУ, используемые автономно, увеличивают ОСШ для речи, но не повышают ее разборчивости слушателями, а лишь снижают их утомляемость. Однако, в тандеме с низкоскоростными речепреобразующими устройствами (РПУ) с параметрическим способом преобразования речи, ШПУ способны повысить точность оценки речевых параметров в анализаторе вокодера при работе в шумах и, тем самым, повысить разборчивость зашумленной речи на выходе системы ШПУ+РПУ по сравнению с использованием только РПУ без ШПУ [5].

Алгоритмы шумопонижения могут быть реализованы как во временной области, так и с помощью различных кратковременных преобразований сигнала, чаще всего, в частотной области, с дальнейшей обработкой кратковременного спектра амплитуд сигнала. Основными проблемами построения ШПУ являются: способ оценки сглаженных спектральных характеристик шумового сигнала и мгновенных значений спектра амплитуд речевого сигнала на основе анализа смеси речи и шума, использование алгоритмов адаптации фильтра, не приводящих к возникновению артефактов звучания («музыкальных» шумов), поиск компромисса между степенью подавления шума и степенью искажения речи. В настоящее время, наиболее развит подход к построению ШПУ на основе метода оценки мгновенных значений логарифма спектра амплитуд речевого сигнала по критерию минимума среднеквадратической ошибки (MMSE-LSA), дополненный статистической оценкой вероятности присутствия речи в шуме [5]. Для субъективного и объективного тестирования ШПУ разрабатываются рекомендации ITU-T P.835 и G. VED.

В Центре ЦОС СПб ГУТ разработано ШПУ для совместного использования с помехоустойчивым РПУ 4400 бит/с. В зависимости от вида шума и величины ОСШ, выигрыш в качестве речи PESQ-MOS, измеренный по рекомендации ITU-T P.862, достигает 6-8 дБ [7].

Одним из ключевых вспомогательных устройств (3) является помехоустойчивый детектор речь/пауза (Voice Activity Detector - VAD). Он необходим, например, для управления выходной скоростью цифрового потока в вокодерах, организации передачи данных во время пауз речи (DTX/CNG), для снижения энергопотребления передатчиков, при управлении адаптацией в АЭК, ШПУ, АРУ и др. блоках. Система VAD может применяться также на приемной стороне для отсеки шума незанятого канала. Основной проблемой построения алгоритмов VAD являются

ошибки классификации речи в присутствии шума, особенно в начале и в конце речевых фрагментов [6]. Универсальных стандартных методик тестирования VAD пока не предложено.

Система автоматической регулировки уровня (APU) речи на входе РПУ (5) при работе в шумах должна быть чувствительна только к изменению уровня речи, а не шума паузы. В зависимости от задач ее функциональность может быть усилена от простого компандирования, до многополосной компрессии динамического диапазона речи [2]. Основой для тестирования систем APU является рекомендация ITU-T G.169.

Центральным устройством цифровой системы речевой связи является речепреобразующее устройство (РПУ) (6). Оно построено на основе алгоритма низкоскоростной компрессии речи (вокодера), состоящего из анализатора речи на передающей стороне и синтезатора речи на приемной стороне. 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. Основными проблемами при разработке вокодеров являются: выбор модели синтеза речевого сигнала; помехоустойчивая оценка параметров модели на основе анализа текущей речи (например, оценка траектории основного тона для речи в шумах [8]); выбор способа квантования параметров; выбор схемы защиты от ошибок в канале [6].

В Центре ЦОС СПб ГУТ разработаны и реализованы в виде моделей для ПЭВМ и программ для DSP семейств TMS320VC54xx/55xx, как собственные, так и стандартные алгоритмы низкоскоростной компрессии речи со скоростями от **600 бит/с** и выше [9]. Параметрические MELP вокодеры со скоростями от 600 до 4000 бит/с, построенные на основе линейного предсказания со смешанным многополосным возбуждением, обладают значительно более высоким качеством речи по сравнению с классическими LPC вокодерами с бинарным решением тон/шум. Нишу скоростей от 4.6 до 16 кбит/с занимают гибридные CELP вокодеры, как стандартные (G.723.1, семейство G.729, G.728 и др.), так и собственной разработки (TCELP 4.6 кбит/с [10]). Нишу скоростей от 16 кбит/с и выше занимают кодеры формы волны сигнала (ADPCM G.726, G.711) и широкополосные аудио кодеки (G.722.1). Особенный интерес для систем цифровой речевой связи представляют низкоскоростные помехоустойчивые вокодеры (1200-7200 бит/с), разработанные на основе совместной оптимизации схем речевого и канального кодирования [11,12] и сохраняющие работоспособность до 5–15 % битовых ошибок в дискретных каналах связи, в условиях когда стандартные вокодеры (например, системы TETRA) не работают.

Перспективными в настоящее время направлениями в области разработки вокодеров являются: сверх низкоскоростные (300-600 бит/с) вокодеры; вокодеры с переменной (адаптивной) скоростью цифрового потока; высококачественные широкополосные вокодеры; системы транскодирования цифровых потоков без промежуточного преобразования в речевую форму (например, MELP \leftrightarrow CELP); создание помехоустойчивых вокодеров на основе совместной оптимизации схем шумоподавления, речевого и канального кодирования.

При наличии ошибок в цифровых каналах связи, на приемной стороне возникает задача интерполяции отдельных битовых ошибок или потерь речевых пакетов целиком. В первом случае, проблема решается совместным использованием схем речевого и канального кодирования в вокодерах, путем внесения избыточности для наиболее значимых бит в речевом кадре. Во втором – путем интерполяции потерянных речевых параметров по прошлым и, возможно, будущим кадрам. Для систем с низким уровнем компрессии (ITU-T G.711, G.726) целесообразно использовать отдельные устройства интерполяции потерянных речевых пакетов (Packet Loss Concealment – PLC)(7). Одной из известных рекомендаций в этой области является ITU-T G.711 Appendix 1. При размере пакетов не более 30 мс удается сохранить приемлемое качество до 15% случайных потерь пакетов без группирования.

Одним из возможных вспомогательных устройств при переходе от узкополосных к широкополосным системам связи является устройство выполняющее искусственное расширение (восстановление) частотного диапазона (50–300 Гц и 3.4–7 кГц) для узкополосной синтетической речи на основе ее анализа и классификации (8) [13].

В условиях внешних акустических шумов на приемной стороне важную роль приобретает обеспечение комфортных условий прослушивания речевого сообщения. Перспективными подходами в данном направлении являются разработка адаптивных к внешнему шуму устройств регулировки выходной громкости и частотной характеристики (тембра) речи (12) [2] и разработка устройств активного подавления шумов (13) [14].

Задачей устройств первого типа является снижение эффекта маскирования прослушиваемой речи мешающим внешним шумом, путем частотно-взвешенного повышения ее громкости для формирования достаточной глубины артикуляции речевого сигнала на фоне шума во всем частотном диапазоне. Особенно это важно для систем связи и оповещения на транспорте, работающих как во время остановок, так и во время движения.

Устройства второго типа достигают аналогичной цели путем снижения уровня громкости акустического шума в точке прослушивания путем его акустической компенсации на основе адаптивных алгоритмов. Примером практической реализации такого устройства может быть оборудованный гарнитурой шлемофон, в котором, наряду с пассивной шумоизоляцией, применяется активное подавление шума в примыкающих к ушам точках пространства.

Устройства обоих типов для оценки или компенсации внешнего шума используют сигнал, получаемый от местного микрофона.

Таким образом, создание современных помехоустойчивых систем цифровой речевой связи требует комплексного подхода, одновременно учитывающего, по крайней мере, два источника возможных помех – акустических шумов и ошибок, возникающих в цифровых каналах связи.

Одним из перспективных решений является совместная оптимизация алгоритмов шумоподавления, речевого и канального кодирования, объединенных в помехоустойчивый вокодер. Одним из возможных критериев оптимизации может быть достижение максимального качества речи для фиксированной скорости цифрового канала связи при заданных параметрах максимального уровня битовых ошибок и минимального ОСШ для речи на входе системы.

Помимо разработки помехоустойчивых вокодеров, сочетающих в себе задачи шумоподавления, речевого и канального кодирования, в системах цифровой речевой связи возможно использование целого ряда дополнительных устройств, способных улучшить ОСШ для речевых сигналов в присутствии акустических шумов, как на приемной, так и на передающей стороне, и тем самым повысить надежность связи [2]. Многие из этих устройств, помимо систем связи, могут быть использованы и в других системах речевой коммуникации, например, в цифровых слуховых аппаратах или в интерфейсах человек-машина.

Литература

1. Nordholm, I. Claesson, Grbic N. Optimal and adaptive microphone arrays for speech input in automobiles. Chapter in the book "Microphone Arrays: Signal Processing Techniques and Applications", Springer, 2001.
2. Voice Communication Package (VCP). Alango Technologies 2006, <http://www.alango.com>.
3. DARPA ASE program. <http://www.darpa.mil/ato/solicit/ASE/index.htm>.
4. Breining, C. et al. Acoustic echo control. An application of very-high-order adaptive filters. Signal Processing Magazine, IEEE, Vol. 16, Issue 4, Jul 1999, pp. 42 – 69.
5. Speech Enhancement (Signals and Communication Technology). Editors: Benesty J., Makino S., Chen J. Springer, 2005, 406 pages.
6. Kondo A.M. Digital Speech. Coding for low bit rate communication systems. J.Wiley & Sons, 2004.
7. Бабкин В.В. Шумопонижающее устройство для вокодера. Отчеты DSPА-2007.
8. Бабкин В.В. Помехоустойчивый выделитель основного тона речи. Отчеты DSPА-2005, стр. 175–178.
9. Вокодеры 600-7200 бит/с. Разработки Центра ЦОС СПб ГУТ. <http://www.dsp.sut.ru>.
10. Babkin V.V., Ivanov V.N., Lanne A.A., Pozdnov I.B. Internet Telephony Vocoders, Proc. Second European DSP E&R Conference, Paris, Sept.1998, p.83-87.
11. Бабкин В. В., Ланнэ А.А., Шаптала В.С. Оптимизационная задача выбора речевого и канального кодирования. Отчеты DSPА-2005, стр. 123–127, Москва 16-18 марта 2005 г.
12. 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.
13. Vary P., Martin R. Bandwidth extension of speech signals. Chapter in the book "Digital Speech Transmission: enhancement, coding and error concealment". Wiley & Sons, 2006, pp. 361-387.
14. Hakansson L. et al. Noise canceling headsets for speech communication. Chapter in the book "Noise Reduction in Speech Applications", CRC press, 2002, pp. 305–327.