

УДК 621.391.01

ЗАЩИТА ОТ ОШИБОК И ИНТЕРПОЛЯЦИЯ ПОТЕРЬ ПАКЕТОВ В НИЗКОСКОРОСТНЫХ РЕЧЕВЫХ КОДЕКАХ

В. В. Бабкин, начальник лаборатории ЦОС СПб ГУТ, к.т.н.; dspcenter@hotmail.ru

Ключевые слова: низкоскоростное кодирование речи, вокодеры, помехоустойчивое кодирование, интерполяция потерь пакетов.

Введение. Низкоскоростные (0,6–16 кбит/с) речевые кодеки (вокодеры) широко используются в системах цифровой речевой связи (ЦРС), организуемой по радиоканалам в диапазонах КВ и УКВ или в сетях передачи данных (ПД), в том числе с коммутацией пакетов, например в системах VoIP (Voice over IP).

Типовой вокодер состоит из кодера (анализатора) и декодера (синтезатора) речевого сигнала. В реальной системе ЦРС декодер разделен с кодером цифровым каналом связи, в котором могут возникать ошибки. Обработка речи ведется, как правило, по кадрам фиксированной длительности с формированием на выходе кодера речевых пакетов, состоящих из фиксированного или переменного числа бит.

Основными техническими характеристиками вокодеров являются: скорость цифрового потока на выходе анализатора, разборчивость и качество речи на выходе синтезатора, алгоритмическая задержка обработки сигнала, полоса передаваемых звуковых частот, чувствительность к акустическим шумам на входе вокодера, чувствительность к битовым ошибкам в цифровом канале связи, вычислительная сложность алгоритмов и их реализации для работы в реальном масштабе времени.

Основными параметрами качества передачи речи принято считать: разборчивость и качество речи, узнаваемость голоса диктора, отсутствие специфических искажений и артефактов звучания. Методы оценки этих параметров можно разделить на объективные (измеряемые инструментально) и субъективные (построенные на основе мнений людей-экспертов). Граница разделения достаточна условна, так как во многих измерениях, например артикуляционных испытаниях, принимают участие специально обученные люди, а в оценках субъективного качества речи широко используются компьютерные модели [1]. За эталон цифрового качества звучания узкополосной (0,3–3,4 кГц) речи обычно принимается качество цифровой телефонии по рекомендации ITU-T G.711 со скоростью 64 кбит/с (8 кГц, 8 бит на отсчет, логарифмическое компрессирование).

Основными причинами снижения показателей качества речевых сигналов на выходе реальных систем ЦРС по сравнению с исходными сигналами на входе являются: ограниченная пропускная способность цифровых каналов связи, приводящая к использованию низкоскоростного кодирования речи с существенными потерями информации; акустический шум на стороне передачи, снижающий точность оценки параметров исходного речевого сигнала; акустический шум на стороне приема, затрудняющий восприятие синтезированного речевого сигнала человеком; битовые ошибки и потери пакетов в цифровых каналах связи, искажающие передаваемую информацию.

Таким образом, при разработке вокодеров с заданными характеристиками необходимо решить следующие основные задачи: рациональный выбор модели описания речевого сигнала, разработка алгоритмов помехоустойчивой оценки параметров модели на основе анализа текущего речевого сигнала, выбор способов эффективного квантования параметров модели, распределение ограниченного ресурса скорости цифрового канала между передаваемыми параметрами, выбор способов защиты передаваемых

параметров от ошибок в канале. В статье последняя задача будет рассмотрена более подробно.

Ошибки, возникающие в цифровых каналах связи. По основным свойствам их можно разбить на несколько классов.

1. Случайные битовые ошибки без группирования. Характерны при низком отношении сигнал/шум (ОСШ) в физических каналах связи. Описываются постоянной во времени вероятностью возникновения ошибки в принятом бите P_{BER} (Bit Error Rate).

2. Битовые ошибки с группированием. Характерны для импульсных радиопомех, для мобильных абонентов или связи в диапазоне КВ, когда величина P_{BER} изменяется во времени вследствие перемещения абонента или замираний несущего радиосигнала.

3. Случайные потери отдельных пакетов речевых данных целиком (блоковые ошибки без группирования). Характерны для работы вокодеров в сетях ПД с коммутацией пакетов, с асинхронной доставкой пакетов или с проверкой их целостности на основе контрольных сумм. Описываются постоянной во времени вероятностью пропадания или возникновения ошибочного речевого пакета P_{BLER} (Block Error Rate).

4. Блоковые ошибки с группированием. Характерны для периодических кратковременных или долговременных срывов или пропадания отдельных каналов связи. Борьба с этим явлением на уровне организации системы ЦРС выходит за рамки возможностей вокодеров и далее не рассматривается.

Оценку поведения вокодеров в составе систем ЦРС для данных характерных ситуаций эксплуатации необходимо проводить еще на стадии их разработки, используя моделирование на компьютере. Для этого необходимо построить программные модели — симуляторы каналов, имитирующие поведение реальных цифровых каналов связи. В соответствии с приведенной выше классификацией ошибок в симуляторах каналов можно использовать следующие упрощенные модели их возникновения.

1. Двоичный симметричный канал без памяти (ДСКБП). задается единственной вероятностью битовой ошибки P_{BER} .

2. Двоичный канал с памятью (ДКСП), моделирующий группирование битовых ошибок. Канал имеет два состояния: «хорошее», обозначаемое G (от слова good), с малой вероятностью битовых ошибок P_{BER-G} и «плохое», обозначаемое B (от слова bad), с большей вероятностью битовых ошибок P_{BER-B} . Состояния P и B могут случайно переключаться между собой в любой момент с вероятностями переходов P_{GB} и P_{BG} .

3. Потери отдельных речевых пакетов целиком без группирования. Описываются ДКСП со следующими ограничениями: $P_{BER-G} = 0$, $P_{BER-B} = 0,5$; переключение между состояниями происходит только на границе пакетов, состоящих из фиксированного числа бит, потери пакетов независимы друг от друга и описываются вероятностью потери пакета P_{BLER} . Интересно отметить, что при $P_{BER-B} = 0,5$ испорченный пакет представляет собой случайный набор бит с вероятностью заполнения нулем или единицей равной 0,5, а значение коэффициента битовых ошибок, усредненного за длительное время, стремится к половине величины вероятности блоковых ошибок P_{BLER} .

Основные методы борьбы с ошибками в цифровых каналах связи в приложении к вокодерам. Для борьбы со случайными битовыми ошибками без группирования используется помехоустойчивое канальное кодирование, основанное на внесении до-

полнительной избыточности в передаваемый цифровой поток на стороне кодера с последующей коррекцией возникших в канале ошибок на стороне декодера. Такие методы хорошо разработаны для систем ПД. Особенностью их применения в вокодерах, по сравнению со стандартным применением при ПД, является использование специфических свойств кодированного речевого сигнала, приводящее к общему повышению эффективности защиты передаваемых речевых сообщений от ошибок.

К этим свойствам вокодеров можно отнести, во-первых, локальный характер воздействия отдельных битовых ошибок в речевых пакетах на синтезируемый речевой сигнал с постепенным плавным уменьшением его качества при росте числа ошибок. Большинство алгоритмов речевого кодирования сохраняют приемлемую разборчивость синтезируемой речи при уровне битовых ошибок, достигающих до $10^{-3} \dots 10^{-2}$. Во-вторых, передаваемые параметры имеют, как правило, различную чувствительность к битовым ошибкам с точки зрения их влияния на качество синтеза речи, что позволяет использовать схемы неравномерной защиты бит от ошибок в речевом пакете. В-третьих, передаваемые параметры вокодеров обладают определенными остаточными статистическими свойствами, позволяющими использовать на приемной стороне схемы мягкое декодирование.

Для борьбы с группированием битовых ошибок применяется помехоустойчивое канальное кодирование с дополнительной перестановкой бит перед их передачей по каналу (для снижения степени группирования ошибок при декодировании). Основным отрицательным фактором для систем ЦРС в этом случае может быть возрастающая задержка в передаче речевых сигналов, если глубина перестановки бит превышает размер речевого пакета.

Потери речевых пакетов целиком (блоковые ошибки) представляют собой безвозвратные потери передаваемой информации, не подлежащие точному восстановлению. При ПД эта проблема решается путем повторной передачи пакетов с разнесением во времени или по запросу через обратный канал. В вокодерах возможен совсем иной подход, а именно интерполяция (реконструкция) потерянных речевых данных, исходя из свойств самого источника информации — речевого сигнала. Таким образом, при синтезе речи на выходе вокодера становится возможной реализация субъективного сглаживания (маскирования) артефактов звучания, вызванных потерями одиночных пакетов.

Существует множество алгоритмов маскирования потерь речевых пакетов (PLC—Packet Loss Concealment) [2]. Все они решают задачу генерации вместо отсутствующего или испорченного фрагмента оригинальной речи синтетического речевого сигнала с близкими спектральными характеристиками и отсутствием заметных на слух переходов между оригинальными и интерполируемыми речевыми кадрами.

Для речи без кодирования возможна реконструкция пропущенного кадра во временной области методом повторения предыдущей речи синхронно с периодом основного тона (ОТ) [3]. Многие стандартные гибридные CELP (Code Excited Linear Prediction) кодеки, такие как ITU-T G.723.1, G.729, имеют встроенные механизмы PLC. Однако в них ошибка от потерянного кадра может распространяться на последующие кадры, принятые без ошибок, вследствие использования дифференциального кодирования параметров. Поэтому для систем VoIP предпочтительнее применение специально разработанных CELP-кодеков [4]. Для низкоскоростных параметрических вокодеров удобнее осуществлять интерполяцию в области параметров модели описания речевого сигнала. Этим обеспечивается непрерывность и плавность переходов между смежными фрагментами при синтезе речи.

Помехоустойчивые вокодеры. В соответствии с изложенными выше принципами защиты от битовых ошибок и блоковых потерь в Центре ЦОС СПб ГУТ были разработаны два вокодера со скоростью в канале связи 4320 бит/с и длительностью кадра

25 мс на основе MELP (Mixed Excitation Linear Prediction) модели описания речевого сигнала. Вокодеры предназначены для работы в составе цифровой УКВ радиостанции или системы VoIP.

Вокодеры типа MELP [5] используют параметрическое описание речевого сигнала на основе модели линейного предсказания (ЛП) 10-го порядка со смешанным возбуждением тон/шум в пяти частотных полосах. Помимо этого, отличительной особенностью MELP модели является использование: одиночных импульсов ОТ в сигнале возбуждения для моделирования переходных участков речи и взрывных звуков; адаптивного формантного фильтра для обострения формантных пиков при синтезе; дополнительного кодирования амплитуд гармоник ОТ сигнала остатка ЛП для вокализованных сигналов для повышения узнаваемости голоса; всепропускающего фазового звена для снижения пик-фактора синтезированного сигнала и др.

Первый помехоустойчивый вокодер RMELP-4320 (Robust MELP) построен на основе речевого кодирования со скоростью 2160 бит/с и двух вариантов помехоустойчивого канального сверточного кодирования: со скоростью 1/2 для закрытия всех бит речевого пакета или с использованием подхода неравномерного закрытия бит, исходя из их субъективной значимости при синтезе речевых сигналов [6]. Вокодер реализован в виде программы для ПЭВМ в арифметике с фиксированной точкой и на цифровом процессоре обработки сигналов TMS32VC55xx (38 MIPS, 46 Кбайт).

Второй вокодер MELP-4320 использует только речевое кодирование. Он имеет два режима работы: без интерполяции потерь пакетов и с интерполяцией таких потерь. В режиме без интерполяции декодер всегда синтезирует выходной речевой сигнал на основании принятого пакета вне зависимости от его целостности и содержания. В режиме с интерполяцией весь процесс состоит из следующих шагов.

1. Обнаружение испорченных (потерянных или опоздавших) пакетов с речевыми данными. Эта задача целиком ложится на внешние по отношению к синтезатору вокодера процедуры протокола и доставки речевых пакетов с проверкой их наличия или целостности. На входе синтезатора всегда должен присутствовать текущий речевой пакет и флаг, сигнализирующий о его целостности.

2. Если принятый пакет не поврежден, то осуществляется синтез выходного речевого кадра.

3. Если принятый пакет поврежден, то параметры модели описания речевого сигнала для текущего кадра экстраполируются на основе прошлого кадра или интерполируются на базе прошлого и будущего кадров (при введении задержки синтеза на один кадр). Простейшим способом экстраполяции является повторение параметров предыдущего кадра с понижением общей энергии сигнала. Интерполяцию спектральной огибающей сигнала на основе модели ЛП с точки зрения устойчивости лучше всего проводить в области линейных спектральных пар (ЛСП). В вокодере интерполируются следующие параметры: энергия кадра, ЛСП, величина ОТ для вокализованной речи, признаки тон/шум по частотным полосам. Далее производится синтез пропущенного кадра на основе интерполированных параметров.

Вокодер реализован в виде программы для ПЭВМ в арифметике с фиксированной точкой.

Для проведения сравнительных испытаний вокодеров использовались имитаторы цифровых каналов связи 1-го и 3-го типов, реализованные в виде программ для ПЭВМ. Для них можно задавать вероятность битовой ошибки P_{BER} или блоковой ошибки P_{BLER} . Качество речи на выходах вокодеров оценивалось количественно на основе объективного критерия PESQ-MOS по рекомендации ITU-T P.862 [1] с использованием тестовых записей двух дикторов — мужчины и женщины, читающих фонетически сбалансированные фразы по ГОСТ Р 50840—95.

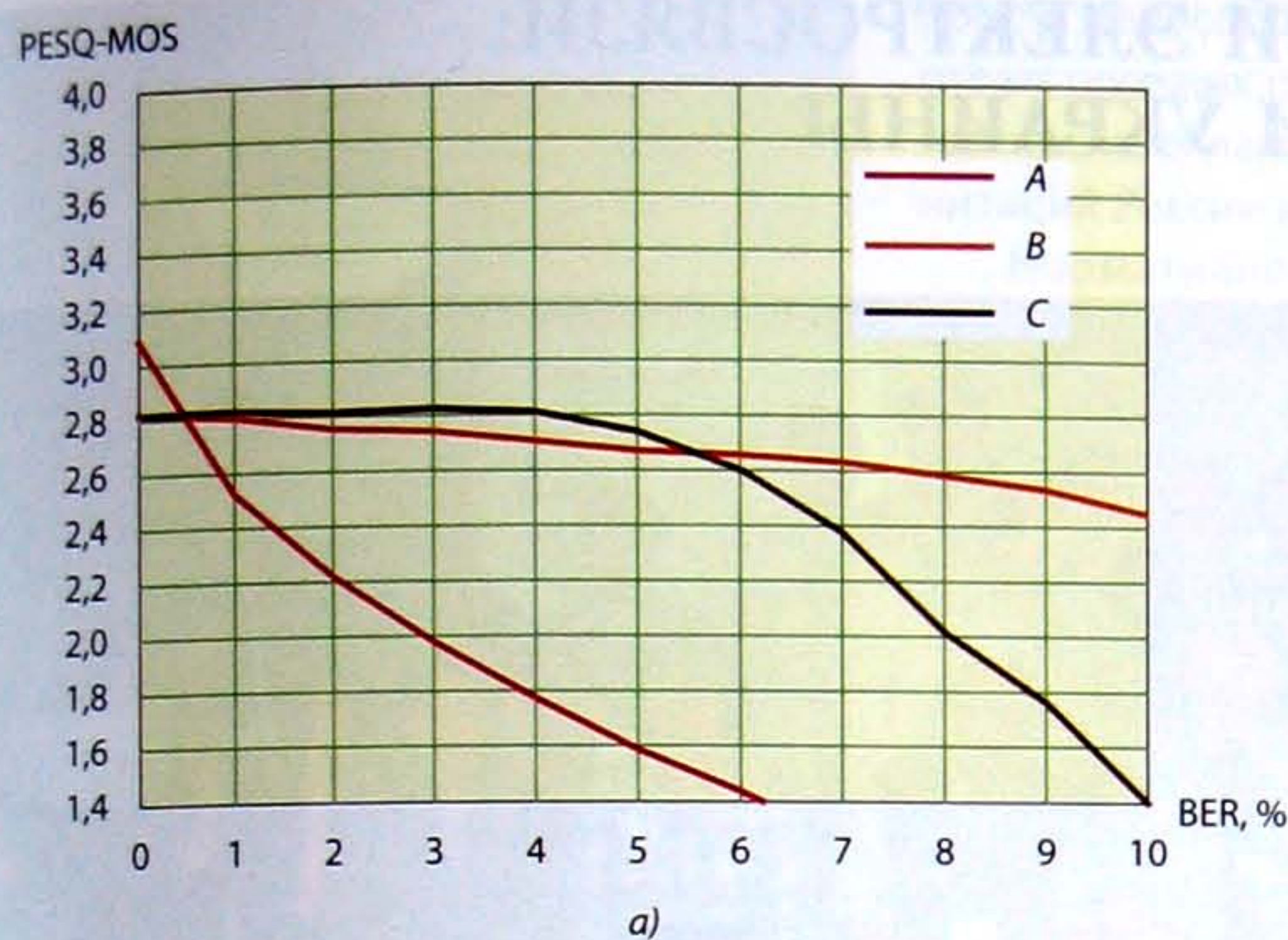


Рис. 1

Зависимости оценок качества речи по шкале PESQ-MOS на выходах вокодеров от вероятности битовых ошибок без группирования P_{BER} в цифровом канале связи приведены на рис. 1 для дикторов «Ася» (а) и «Саша» (б). Используются следующие обозначения: А — вокодер MELP-4320 в режиме без интерполяции, В — вокодер RMELP-4320 с неравномерной защитой бит, С — вокодер RMELP-4320 с равномерной защитой бит.

На рис. 2 представлены зависимости оценок качества речи по шкале PESQ-MOS на выходе вокодера MELP-4320 от вероятности блоковых ошибок без группирования P_{BLER} в цифровом канале связи для двух дикторов и двух режимов работы: с интерполяцией потерь пакетов (А, В) и без интерполяции (С, D).

Из анализа зависимостей видно, что при отсутствии ошибок в канале связи помехоустойчивый вокодер RMELP-4320 уступает по качеству простому вокодеру MELP-4320, однако в отличие от последнего сохраняет работоспособность при битовых ошибках, достигающих до 7–10%. Причем для $P_{BER} > 6\%$ неравномерная схема защиты бит вокодера от ошибок имеет явное преимущество над равномерной защитой всех бит со скоростью 1/2.

Для вокодера MELP-4320 в условиях случайных потерь пакетов без группирования в режиме их экстраполяции (без задержки обработки сигнала) субъективное качество речи сохраняется вплоть до 25% потерь пакетов, а при воспроизведении поврежденных пакетов со случайным заполнением «как есть» значительно снижается.

Заключение. На основе полученных результатов можно сделать следующие выводы.

1. Целесообразно объединять схемы речевого и канального кодирования в один помехоустойчивый вокодер с неравномерным закрытием бит, разрабатываемый на основе их совместной оптимизации, исходя из требований пропускной способности цифрового канала, минимально допустимого качества синтези-

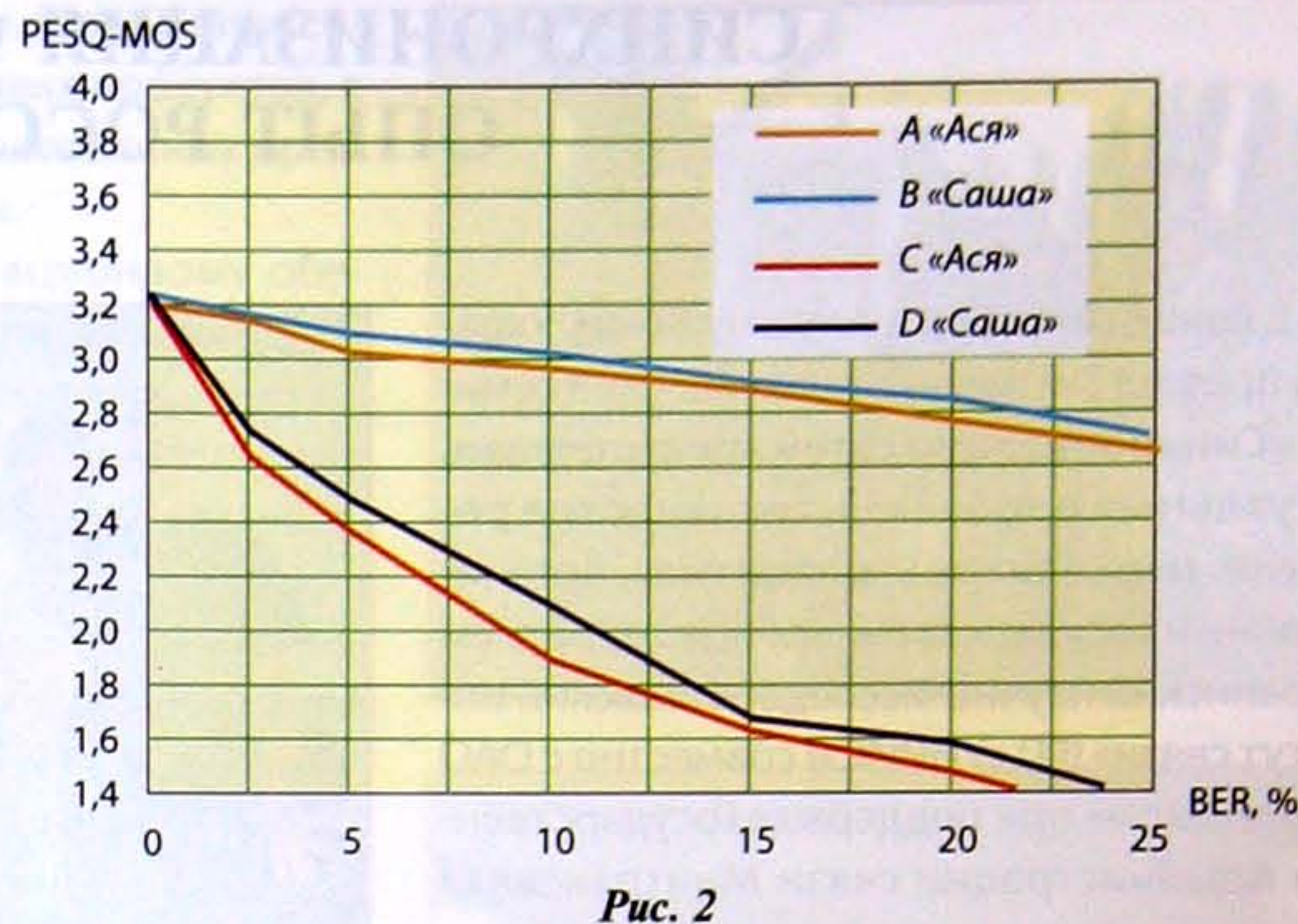


Рис. 2

руемой речи и предельной вероятности битовых ошибок. Схемы с неравномерной защитой передаваемых бит в речевом пакете дают определенный выигрыш по сравнению со схемами с равномерным закрытием бит при высоких уровнях битовых ошибок, однако могут проигрывать им в качестве синтетической речи при низких уровнях ошибок.

2. Для сглаживания восприятия невозвратных потерь отдельных речевых кадров при синтезе речи удобно использовать интерполяцию речевого сигнала на основе его параметрического описания. Маскировка потерь пакетов улучшает субъективное качество речи, однако ее влияние на слоговую разборчивость требует дальнейшего изучения.

Объединение двух подходов, т. е. использование техники интерполяции потерь кадров в условиях значительных битовых ошибок, особенно при организации цифровой связи по радиоканалам, достаточно противоречивая задача. Интеграция механизма контроля целостности (например, CRC) внутрь речевого пакета приводит к значительному росту числа обнаруженных поврежденных пакетов даже при незначительных P_{BER} , что может привести к нежелательным пороговым эффектам в снижении качества речи при росте P_{BER} . Это явление можно наблюдать в системе связи TETRA.

Звуковые файлы разработанных вокодеров со скоростями от 600 бит/с и выше, а также шумопонижающих устройств можно прослушать на странице Центра ЦОС СПб ГУТ www.dsp.sut.ru.

ЛИТЕРАТУРА

1. ITU-T Recommendation P.862. (02/2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.
2. Perkins C., Hodson O., Hardman V. A survey of packet loss recovery techniques for streaming audio. Proc. of Network//IEEE. — 1998. — Vol. 12. — Issue 5. — P. 40–48.
3. ITU-T Recommendation G.711 (1999). Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711.
4. Babkin V.V., Ivanov V.N., Lanne A.A., Pozdnev I.B. Internet Telephony Vcoders/Proc. of Second European DSP E&R Conference. — Paris. — 1998. — Sept. — P. 83–87.
5. Supplee L.M., Cohn R.P., Collura J.S., McCree A.V. MELP: The New Federal Standard at 2400 bps//IEEE ICASSP. — 1997. — P. 1591–1594.
6. Бабкин В. В., Ланне А. А., Шапгала В. С. Оптимизационная задача выбора речевого и канального кодирования/7-я Международ. конф. и выставка «Цифровая обработка сигналов и ее применение» (DSPA-2005). — Москва, 16–18 марта 2005 г.

Получено 22.10.09