

ПАСЮК А. О., БЕРКО Г. А., ГАЛИЧ С. А., СЕМЕНОВ Е. С., ТЮХТЯЕВ Д. А.

ПРИМЕНЕНИЕ АЛГОРИТМА PESQ

ДЛЯ ОЦЕНКИ КАЧЕСТВА ПЕРЕДАЧИ РЕЧИ ПО IP-СЕТЯМ

Аннотация. Дается описание субъективных и объективных оценок качества речи. Приводятся результаты исследования качества речи для различных кодеков на основе алгоритма PESQ.

Ключевые слова: оценка качества, PESQ, IP-телефония.

PASYUK A. O., BERKO G. A., GALICH S. A., SEMENOV E. S., TYUKHTYAEV D. A.

APPLICATION OF THE PESQ ALGORITHM

FOR SPEECH QUALITY ESTIMATION IN IP NETWORKS

Abstract. The article considers the subjective and objective scores of the speech quality. The results of a study of the speech quality for different codecs based on the PESQ algorithm are presented.

Keywords: quality estimation, PESQ, VoIP.

Обеспечение качества обслуживания (QoS) в сетях передачи данных сегодня – одна из самых актуальных практических задач, особенно для приложений IP-телефонии, стремительно вытесняющих традиционную телефонию с рынка телекоммуникационных услуг. Параметрами, влияющими на качество передачи речи через IP-сети, являются задержка, тип кодека, потери пакетов, а также джиттер (вариации задержки пакетов). К параметрам, связанным с оконечным оборудованием, можно отнести тип кодека и задержку. Задержка в терминале обусловлена затратами времени на алгоритмическую обработку кодеком, аналого-цифровое преобразование, а также пакетизацию. К параметрам, которые связаны с сетью передачи данных, относятся потери пакетов, задержка и джиттер. В свою очередь, задержка при передаче по IP-сети складывается из задержки на распространение (которая зависит от расстояния между терминалами) и задержки, вызванной обработкой в очередях промежуточных устройств сети (коммутаторов и маршрутизаторов).

В настоящее время QoS не в состоянии решить некоторые из проблем при передаче речи по IP-сети, такие как задержка на распространение, задержка обработки кодеком, задержка на осуществление выборки и на преобразование в цифровую форму.

Голосовые приложения чувствительнее любых других к потерям пакетов и требуют тщательного планирования, чтобы гарантировать соответствие соглашению уровня

обслуживания (Service Level Agreement – SLA). Одним из элементов такого планирования является выявление «приемлемого» объема задержки. Некоторыми из составляющих задержки можно управлять и настраивать их, в то время как остальные обусловлены физическими явлениями. В таблице 1 показаны элементы, составляющие управляемую часть задержки [1, с. 181].

Таблица 1

Составляющие задержки

	Фиксированная задержка	Переменная задержка
Задержка кодека G.729	5 мс	
Задержка кодека G.729 (по 10 мс на кадр)	20 мс	
Задержка очереди в магистральной 64 кбит/с		6 мс
Задержка на сериализацию в магистральной 64 кбит/с	3 мс	
Задержка на распространение (закрытые линии)	32 мс	
Буфер компенсации джиттера		2-200 мс

Рекомендация G.114 сектора стандартизации ITU-T определяет качество голоса как «хорошее» при общей задержке не большей, чем 150 мс.

При разработке голосового приложения или терминала IP-телефонии необходимо учитывать все вышеперечисленные факторы, влияющие на качество речи. Соответственно, возникает актуальный вопрос – каким образом можно оценить влияние того или иного критического параметра IP-сети на качество речи при использовании конкретного приложения или устройства?

До недавнего времени в течение многих лет использовались т.н. субъективные оценки как единственный подход к оценке качества в телефонных сетях. Такой подход был предложен в рекомендациях МСЭ Р.800, в его основе лежали субъективные оценки качества речи (например, «отличное качество», «хорошее качество», «приемлемое качество» и т.д.). Наиболее широко используемая методика субъективной оценки качества описана в Рекомендации МСЭ Р.800 и известна как методика MOS. В соответствии с ней качество речи, получаемое при прохождении сигнала от говорящего (источник) через систему связи к слушающему (приемник), оценивается как арифметическое среднее от всех оценок, выставяемых экспертами после прослушивания тестируемого тракта передачи [2].

Но субъективные оценки было сложно точно соотнести с сетевыми характеристиками, которые используются при проектировании и эксплуатации сетей. Поэтому в 1998 г. МСЭ стандартизовал подход, основанный на объективных оценках

качества обслуживания, который позволяет описать показатели качества при передаче речи в пакетной форме (Рекомендация МСЭ G.107).

В основу объективного метода положена так называемая *E*-модель, которая, связана с измерением характеристик терминалов и сетей. После создания *E*-модели было проведено большое число испытаний, в которых менялся уровень воздействия искажающих сетевых факторов. Данные этих тестов были использованы в *E*-модели для вычисления объективных оценок. Результатом вычислений в соответствии с *E*-моделью является число, называемое *R*-фактором («коэффициентом рейтинга»).

E-модель является многокритериальной оценкой качества передаваемой речи в сетях IP и *R*-фактор определяется в соответствии с ней в диапазоне значений от 0 до 100, где 100 соответствует самому высокому уровню качества. Но теоретическое значение *R*-фактора уменьшается от 100 до 93,2, что соответствует оценке MOS, равной 4,4.

На практике величина *R*-фактора меняется от 0 до 93,2, что соответствует изменению оценок MOS от 1 до 4,4. Значение *R*-фактора определяется по следующей формуле:

$$R = R_0 - I_s - I_d - I_e + A,$$

где: $R_0 = 93,2$ — исходное значение *R*-фактора;

I_s — искажения, вносимые кодеками и шумами в канале;

I_d — искажения за счет суммарной сквозной задержки («из конца в конец») в сети;

I_e — искажения, вносимые оборудованием, включая и потери пакетов;

A — так называемый фактор преимущества [4].

Большинство методов основано на сравнении оригинального (эталонного) и кодированного (искаженного) сигналов с помощью психоакустической модели. Оценивается степень заметности искажений в кодированном сигнале для человека. Психоакустическая модель – это модель, преобразующая звуковой сигнал в его внутреннее представление с точки зрения слухового аппарата человека, которое и сравнивается с внутренним представлением эталонного сигнала.

Наиболее распространенными является оценка PESQ, определенная в рекомендации МСЭ-T P.862. Она представляет собой объективный метод определения качества речевой связи в телефонных системах, который прогнозирует результаты субъективной оценки качества этого вида связи слушателями-экспертами. Для определения качества передачи речи в PESQ предусмотрено сравнение входного, или эталонного, сигнала с его искаженной версией на выходе системы связи [3].

Результатом сравнения входного и выходного сигналов является оценка качества связи, которая аналогична усредненной субъективной оценке MOS. Далее полученные

результаты оценки PESQ калибруются с использованием огромной базы данных оценок MOS.

Процесс определения качества передачи речи с алгоритмом PESQ схематично изображен на рисунке 1.

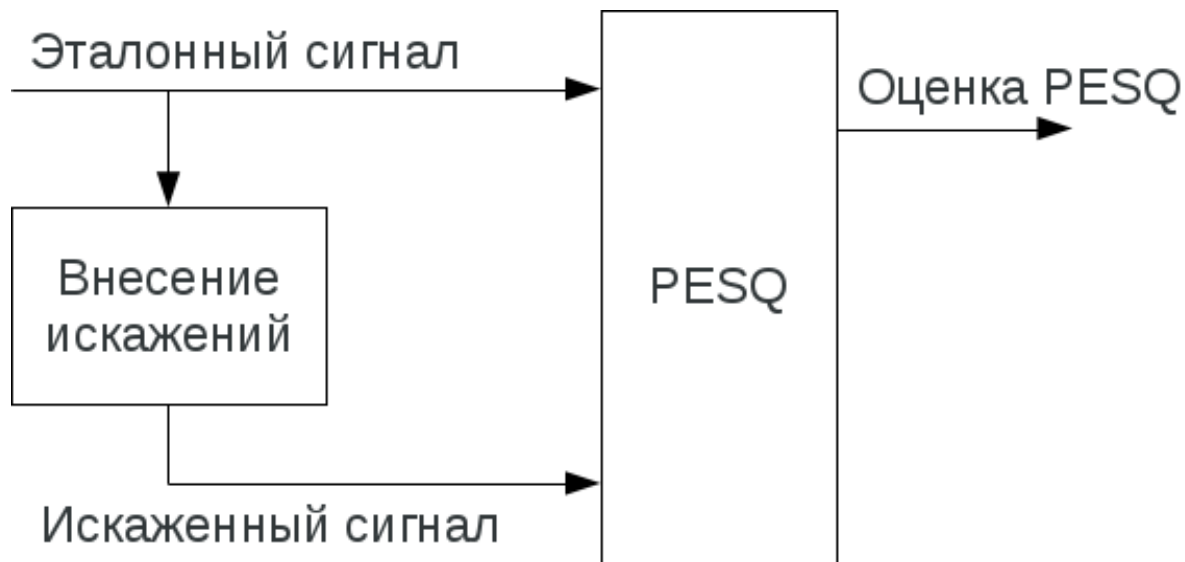


Рис. 1. Сравнение эталонного сигнала с его искаженной версией.

Результатом сравнения эталонного и искаженного сигналов является оценка качества связи, которая аналогична усредненной субъективной оценке MOS. Оценки PESQ калибруются с использованием базы данных оценок MOS. PESQ возможно использовать как для определения качества сквозной передачи речи, так и для оценки влияния на качество связи отдельных элементов сетевого оборудования, включая кодеки.

В статье описаны результаты эксперимента, в котором для оценки качества речи применялся алгоритм PESQ.

Для проведения эксперимента был собран тестовый стенд, состоящий из двух ПК, служащих оконечными терминалами, промежуточного сервера WanEM [5], являющегося эмулятором каналов глобальной сети и маршрутизатора (см. рис. 2). При совершении VoIP-звонка на обоих терминалах производилась одновременная запись голосовых сообщений (на ПК «звонящего» через микрофонный вход – запись эталонного сигнала, на ПК «принимающего звонок» с выхода звуковой платы – запись искаженного сигнала). Далее записанные файлы попарно обрабатывались алгоритмом PESQ. Обработка производилась на основе метода, предложенного в [6].



Рис. 2. Схема тестового стенда.

На выходе получается массив из двух элементов:

- 1) PESQMOS;
- 2) MOS-LQO.

MOS (Mean Opinion Score) – среднее значение оценок, т. е. значение на заранее определенной шкале, по которой субъекты оценивают качественные показатели работы телефонной системы передачи, используемой для разговора или для слушания речевого материала.

Кроме субъективного мнения, сокращение MOS также используется для оценок, которые исходят из объективных моделей или моделей планирования сети. Для разграничения области применения вместе с сокращением MOS рекомендуется использовать следующие обозначения: N – узкополосный, W – широкополосный, LQ – listening quality (качество при слушании), CQ – conversational quality (качество при разговоре), S – subjective (субъективный), O – objective (объективный), а E – estimated (плановый).

MOS-LQO – данная оценка вычисляется посредством объективной модели, целью которой является прогнозирование качества испытательной ситуации «только слушание». Результатами объективных испытаний, выполненных при помощи модели, являются показатели MOS-LQO. Процесс обработки сигналов при использовании алгоритма PESQ показан на рисунке 3.

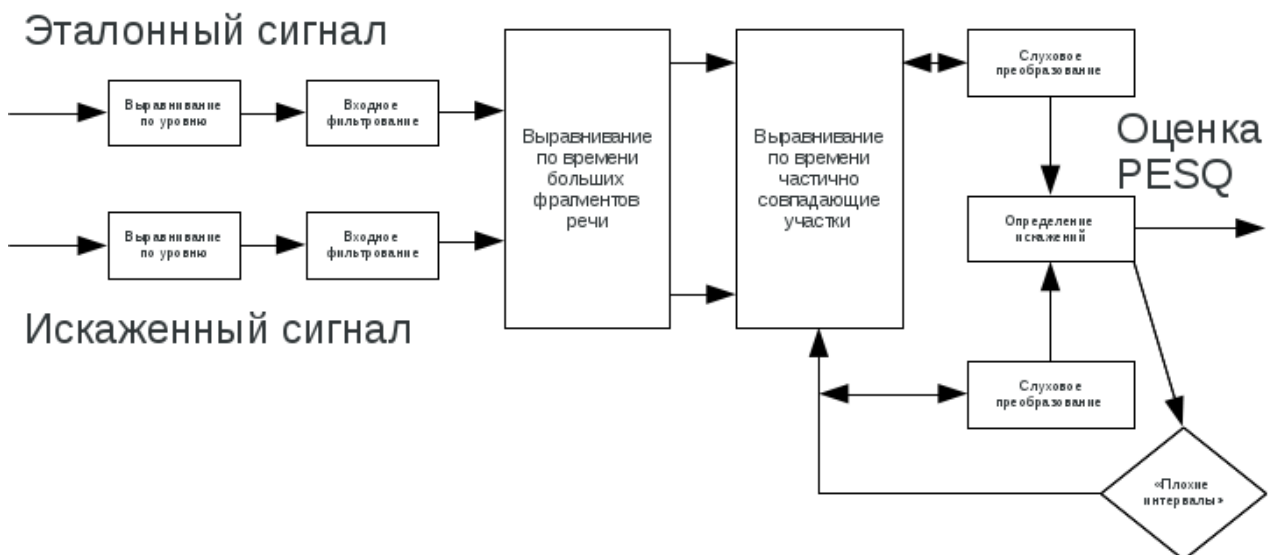


Рис. 3. Процесс обработки сигналов в PESQ.

В процессе эксперимента изменялись различные параметры канала передачи данных между двумя терминалами, а именно процент потери пакетов и джиттер. Исследование проводилось для трех типов кодеков: G.711, G.729, GSM. Результаты исследования представлены на рисунках 4–7.

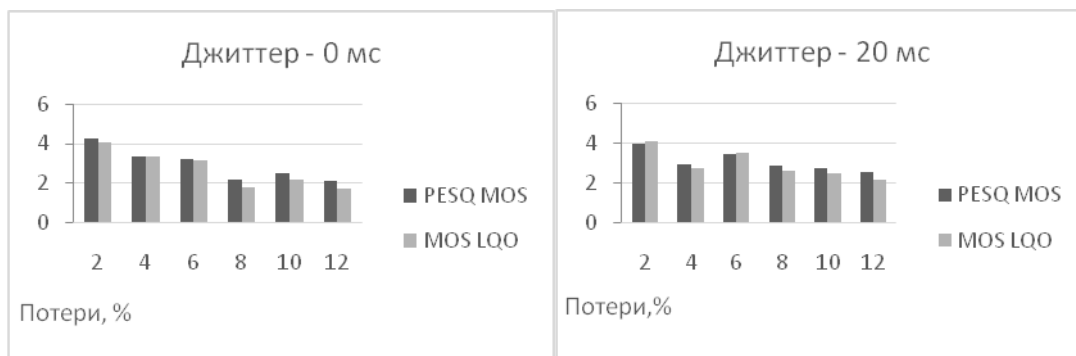


Рис. 4. Оценки для кодека G.711

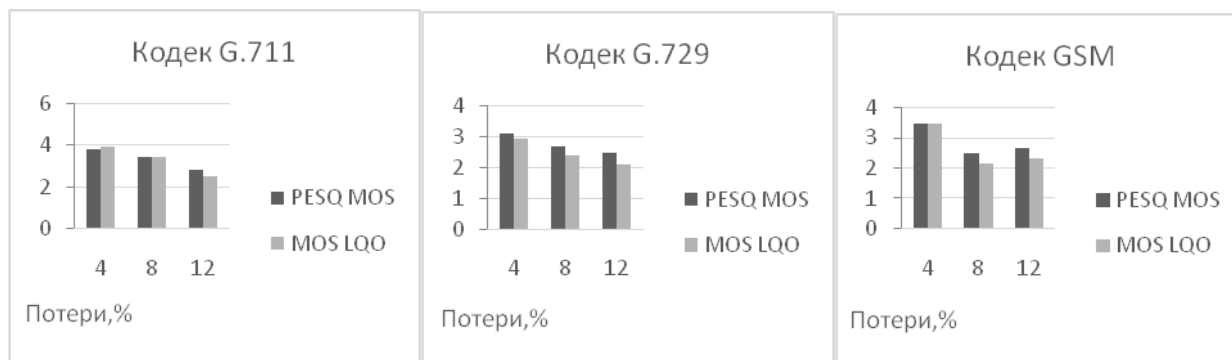


Рис. 5. Оценки для различных типов кодеков при джиттере 50 мс.

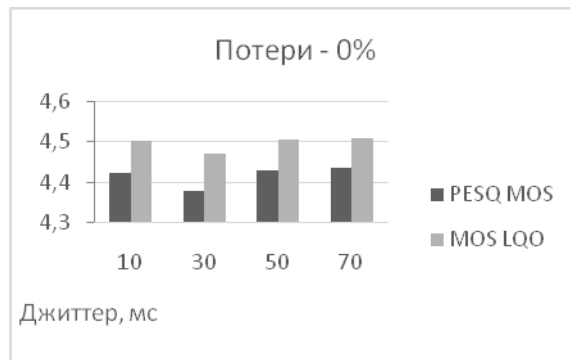


Рис. 6. Оценки для кодека G.711 при 0% потерь пакетов.

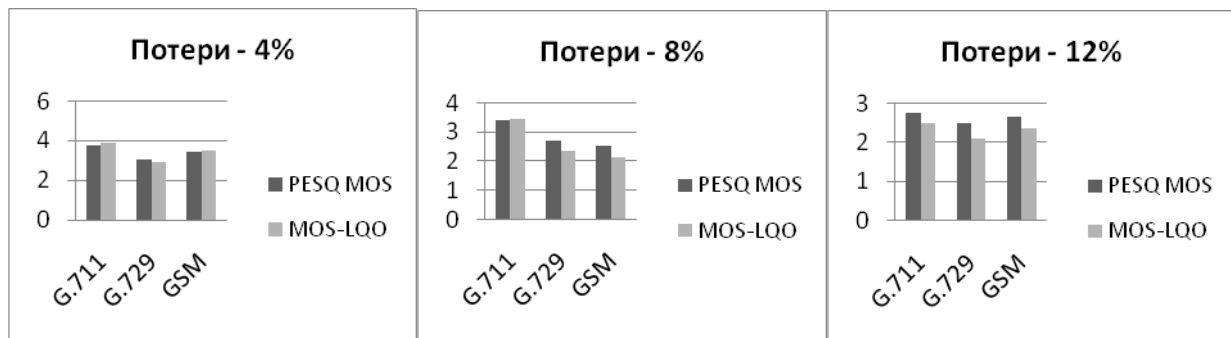


Рис. 7. Оценки для различных типов кодеков при значениях процента потерь пакетов – 4%, 8%, 12%.

На основании приведенных результатов можно сделать следующие выводы.

1. Качество речи при использовании кодека G.711 существенно зависит от потери пакетов при передаче по IP-сети. Так, на графике рисунка 4 можно увидеть, что оценка PESQ меняется в зависимости от процента потерь пакетов от 4,256 до 2,062 при нулевом джиттере, а при внесении джиттера 20 мс замечается незначительное ухудшение оценки PESQ для кодека G.711.

2. При внесении джиттера 50 мс, на графике 5 для кодека G.711, можно увидеть, что при больших процентах потерь джиттер не оказывает значительного влияния на качество речи для кодека. Также можно сказать, что при одинаковых условиях сети качество речи для кодеков G.729 и GSM ухудшается по сравнению с кодом G.711 в среднем на 17 и 11% соответственно.

3. Анализируя график на рисунке 6, можно сделать окончательный вывод о том, что джиттер незначительно влияет на качество речи для кодека G.711.

4. Из графиков на рисунке 7 следует, что в условиях больших потерь пакетов в IP-сети наилучшее качество речи обеспечивают кодеки G.711 и GSM (по сравнению с G.729).

ЛИТЕРАТУРА

1. Дэвидсон Д., Питерс Д., Бхатия М. и др. Основы передачи голосовых данных по сетям IP / пер. с англ. – 2-е изд. – М.: ООО «И. Д. Вильямс», 2007. – 400 с.
2. МСЭ-Т Р.800. Методы для объективной и субъективной оценки качества // Сектор стандартизации электросвязи МСЭ. – 2006.
3. МСЭ-Т Р.862. Оценка восприятия качества речи (PESQ): Объективный метод для сквозной оценки качества речи узкополосных телефонных сетей и речевые кодеки. // Сектор стандартизации электросвязи МСЭ. – 2003.
4. Яновский Г. Г. Оценка качества передачи речи в сетях IP // Вестник связи. – 2008. – № 2. – С. 91–94.
5. WanEM. The Wide Area Network emulator [Электронный ресурс]. – Режим доступа: <http://wanem.sourceforge.net>.
6. Woichicki K. PESQ MATLAB Wrapper [Электронный ресурс]. – Режим доступа: <http://www.mathworks.com/matlabcentral/fileexchange/33820-pesq-matlab-wrapper>.