

УДК 004.627  
DOI 10.22213/2410-9304-2018-4-109-113

## ИСПОЛЬЗОВАНИЕ СТАНДАРТА P.862 ДЛЯ СРАВНЕНИЯ КАЧЕСТВА НИЗКОСКОРОСТНЫХ ВОКОДЕРОВ

*А. В. Коробейников*, кандидат технических наук, доцент, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия  
*М. А. Бояршинов*, кандидат технических наук, доцент, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия  
*А. И. Нистюк*, доктор технических наук, профессор, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия  
*В. Н. Емельянов*, кандидат технических наук, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия

Рассматриваются объективные методы оценки качества речевого сигнала: 1) *Perceptual Evaluation of Speech (PESQ, рекомендация МСЭ-Т P.862)* – оценка восприятия качества речи, 2) *Listening Quality Objective (LQO, рекомендация, МСЭ-Т P.800.1)* – качество прослушивания. Приведено краткое описание и схемы работы методики PESQ и формулы для преобразования оценок Raw MOS в MOS-LQO и обратно. Для тестирования были выбраны низкоскоростные вокодеры: 1) MELPe, 2) Speex, 3) Codec2. Тестирование вокодеров проводилось на битовых скоростях от 700 до 4800 бит/с. Для тестирования использовались аудиофайлы артикуляционных таблиц количеством 20 записей (wav, 8000 КГц, 16 бит, моно). В результате тестирования были построены таблицы и графики для Raw MOS и MOS-LQO оценок выбранных вокодеров. При анализе результатов экспериментов сделан вывод об эффективности применения объективных методов оценки качества речи, и в качестве перспективного для дальнейших разработок вокодера был выделен MELPe, обеспечивающий на битовых скоростях 1200 и 2400 бит/с оценку качества MOS соответственно 2,9...3,2 и 3,0...3,3. Вокодер Speex показал сравнимые с MELPe результаты оценки при большей битовой скорости (4800 бит/с), а вокодер Codec2 показал результаты оценки ниже, чем MELPe.

**Ключевые слова:** сжатие речи, низкоскоростные вокодеры, оценка качества, стандарт P.862.

### Введение

При разработке аппаратуры передачи информации, в частности речи, постоянно возникает задача оценки эффективности различных вариантов построения указанной аппаратуры. Поэтому много внимания уделяется средствам, позволяющим оценить качество предложенных решений, и автоматизации измерений [1].

Однако, несмотря на широкое распространение такой услуги, как передача речи, оценка ее качества долгое время проводилась только субъективными методами. Более того, при государственных испытаниях разработанной аппаратуры должны применяться именно субъективные методы, так как основанные на них методики закреплены в нормативных документах [2].

Субъективные методы основаны на чтении группой дикторов фраз и неполных слогов из артикуляционных таблиц. Принятые фразы и слоги записываются аудитором. По записанным таблицам вычисляется процент верно принятых слогов и фраз. Подобный процесс оценки качества трудоемок, требует значительного времени, а на ранних стадиях проектирования еще и трудно реализуем.

В настоящее время появляются объективные методы оценки качества речи, которые можно применять на любых, в том числе на самых ранних стадиях проектирования, даже на стадии моделирования. Рассмотрим подобные методики.

### Методика тестирования вокодеров

#### *PESQ Raw MOS*

*PESQ (Perceptual Evaluation of Speech, рекомендация МСЭ-Т P.862)* – оценка восприятия качества речи: объективный метод для сквозной оценки качества речи узкополосных телефонных сетей и речевых кодеков (2001 г.). *PESQ Raw MOS* – необработанная (сырая) оценка MOS (*Mean Opinion Score*, средняя экспертная оценка) стандарта *PESQ* [3].

*PESQ* сравнивает оригинальный сигнал  $X(t)$  с ухудшенным сигналом  $Y(t)$ , являющимся результатом прохождения  $X(t)$  через коммуникационную систему. Значение *PESQ* – это предсказание воспринимаемого качества  $Y(t)$  экспертами в субъективном тесте на прослушивание.

На первом шаге *PESQ* вычисляется ряд задержек между исходным и ухудшенным сигналом, по одному для каждого временного интервала, для которого задержка значительно отличается от предыдущего временного интервала. Для каждого из этих интервалов рассчитывается соответствующая начальная и конечная точка. Алгоритм выравнивания основан на принципе сравнения вероятности наличия двух задержек в определенном интервале времени с уверенностью наличия одной задержки для этого интервала. Алгоритм может обрабатывать изменения задержки как во время тишины, так и во время активных частей речи.

На основе набора обнаруженных задержек *PESQ* сравнивает исходный сигнал с выровненным ухудшенным сигналом тестируемого устройства с помощью модели восприятия (рис. 1). Ключом к этому процессу является преобразование как исходного, так и ухудшенного сигналов во внутреннее представление, аналогичное пси-

хофизическому представлению звуковых сигналов в слуховой системе человека, с учетом частоты и громкости восприятия. Этапы анализа: выравнивание по времени, выравнивание уровня до откалиброванного уровня прослушивания, время-частотное картирование, частотное искажение и сжимающее масштабирование громкости.

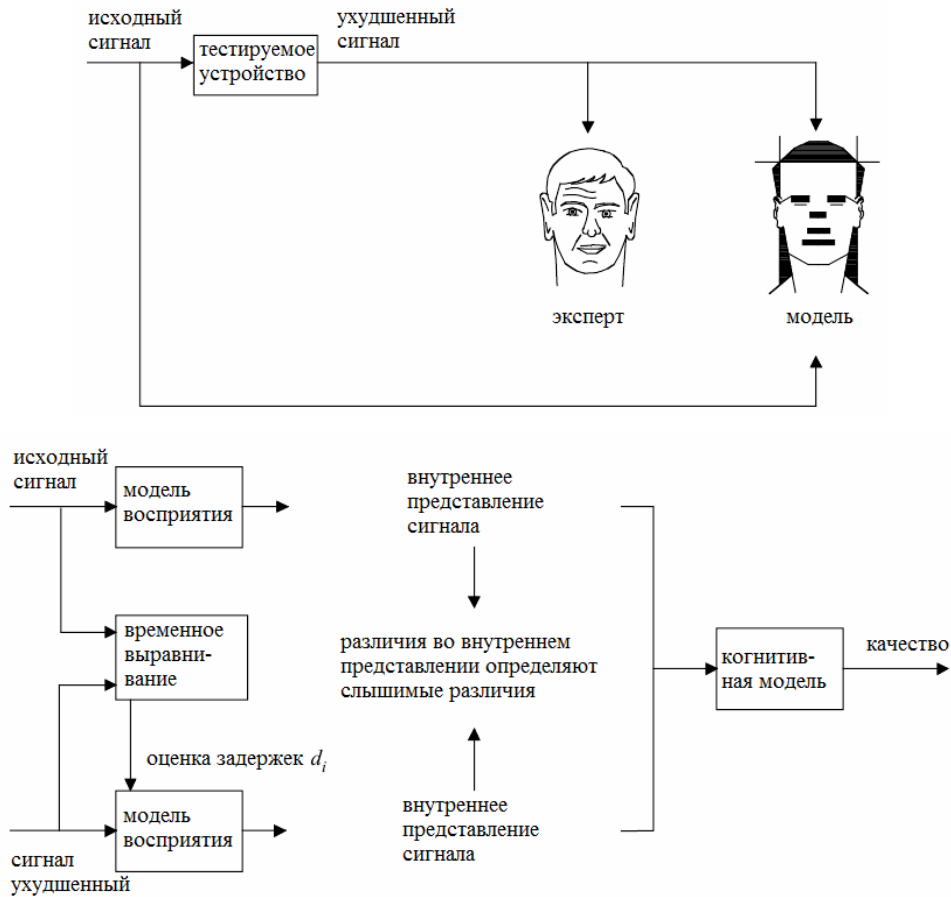


Рис. 1. Обзор методики *PESQ*

Внутреннее представление сигналов обрабатывается с учетом таких эффектов, как локальные изменения усиления и линейная фильтрация, которые могут, если они не слишком серьезны, иметь небольшое значение при восприятии. Это достигается путем ограничения степени коррекции и выполнения коррекции с задержкой после речевых различий. При этом корректируются незначительные, стационарные различия между исходным и ухудшенным сигналами. Более сильные различия, или быстрые изменения, только частично корректируются так, что остаточные различия сигналов сохраняются и вносят свой вклад в общее нарушение восприятия. Это позволяет использовать небольшое количество показателей качества для моделирования всех субъективных различий. В когнитивной модели *PESQ* вычисляются два параметра ошибок, кото-

рые объединяют для получения объективного качества прослушивания *MOS*.

*PESQ* предоставляет оценку *Raw MOS* в диапазоне 0,5...4,5.

*PESQ MOS-LQO*

*MOS-LQO (Listening Quality Objective, МСЭ-Т Р.800.1)* – качество прослушивания. Данная оценка вычисляется на основе объективной модели *Raw MOS*, целью которой является прогнозирование качества испытательной ситуации «только слушание», которая описывается рекомендацией [4, 5].

*PESQ* предлагает функции отображения оценки *Raw MOS* в *MOS-LQO*:

$$y = 0,999 + \frac{4,999 - 0,999}{1 + \exp(-1,4945 \cdot x + 4,6607)} \quad (1)$$

Обратная функция позволяет отобразить  $MOS-LQO$  в  $Raw\ MOS$ :

$$x = \frac{4,6607 - \ln\left(\frac{4,999 - y}{y - 0,999}\right)}{1,4945} \quad (2)$$

График функции отображения  $Raw\ MOS$  в  $MOS-LQO$  представлен на рис. 2.

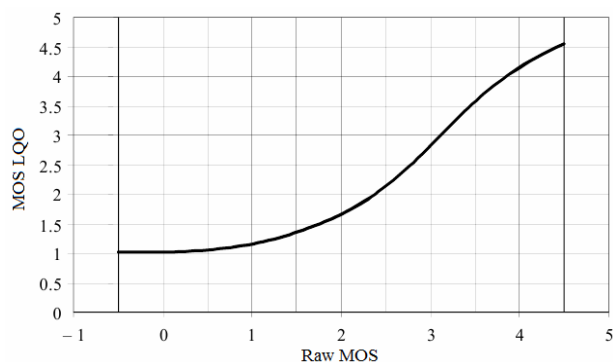


Рис. 2. График функции отображения  $Raw\ MOS$  в  $MOS-LQO$

### Тестируемые вокодеры

При кодировании речи входными данными служит файл формата  $RAW$  или  $WAV$  [6] с параметрами: аудиоформат = 1 ( $PCM$ ); количество каналов = 1 (моно); частота дискретизации = 8000 (Гц, отсчетов/секунда); количество бит в отсчете = 16; количество байт для одного отсчета = 2; количество байт за секунду воспроизведения = 1600. Результатом кодирования является файл, предназначенный для передачи по каналу связи.

По техническому заданию требовалось обеспечить битовые скорости кодирования речи: 1200, 2400, 4800 бит/с.

Для обеспечения этих скоростей были выбраны следующие вокодеры:

1.  $MELPe$  [7, 8] – расширение стандарта  $MELP$ .  $MELP$  (*Mixed-Excitation Linear Pre-*

*diction*, линейное прогнозирование со смешанным возбуждением) – стандарт Министерства обороны США для кодирования речи, используемый в основном в военных приложениях и спутниковой связи, защищенном голосовом и радиооборудовании. Вокодер  $MELPe$  обеспечивает битовые скорости 2400, 1200 и 600 бит/с. В открытых источниках доступен исходный код вокодера  $MELPe$  на языке Си, обеспечивающий битовый скорости 2400 и 1200 бит/с.

2.  $Speex$  [9] – это свободный кодек для сжатия речевого сигнала.  $Speex$  относится к классу так называемых  $CELP$ -кодеков [10] (*Code Excited Linear Prediction*), то есть кодеков, построенных на основе так называемого линейного предсказательного кодирования (ЛПК). Вокодер обеспечивает широкий набор доступных битовых скоростей: 2000...44 000 бит/с. Доступны исходные тексты вокодера на языке Си. Для тестирования использована битовая скорость 4800 бит/с.

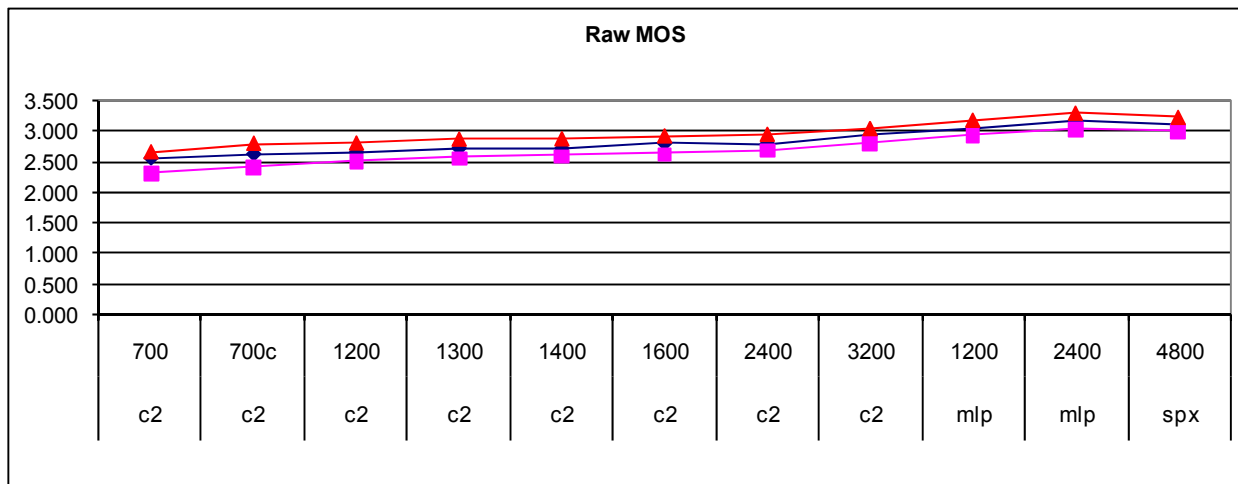
3.  $Codec2$  [11] – свободный речевой вокодер, направленный на общедоступную замену проприетарного вокодера  $MELP$ . Доступны исходные тексты вокодера на языке Си. Набор битовых скоростей вокодера: 700, 1200, 1300, 1400, 1600, 2400, 3200 бит/с.

### Результаты тестирования вокодеров

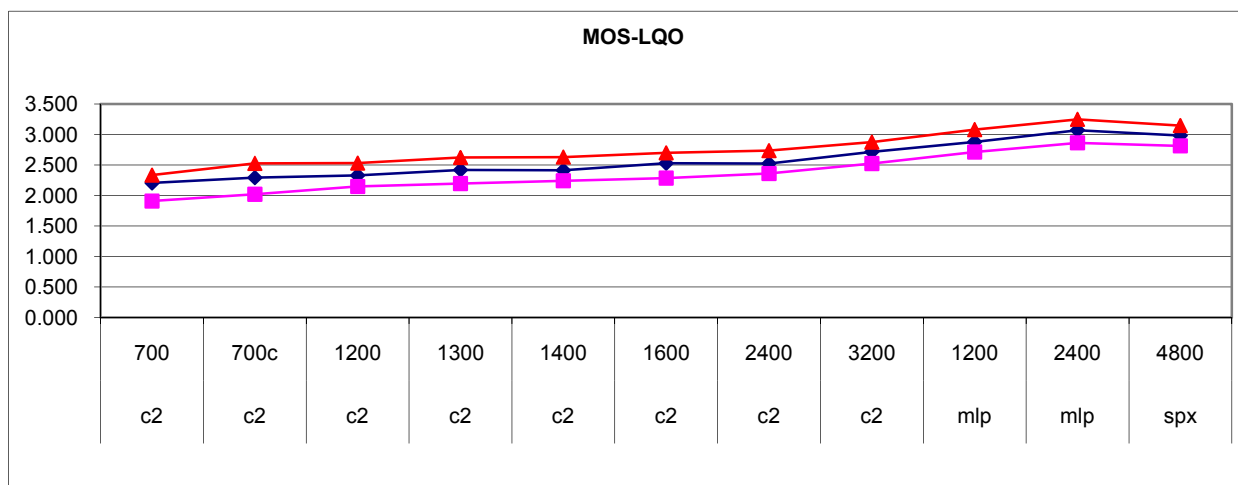
Оценка проводилась согласно стандарту  $PESQ$ . В таблице и рис. 3 приведены результаты тестирования качества вокодеров: средние ( $avg$ ), минимальные ( $min$ ) и максимальные ( $max$ ) оценки  $Raw\ MOS$  и  $MOS-LQO$ . Тестирование проводилось для вокодеров  $Codec2$  ( $c2$ ),  $MELPe$  ( $mlp$ ) и  $Speex$  ( $spx$ ) на битовых скоростях от 700 до 4800 бит/с. На рис. 3 на основе данных таблицы построены графики ( $min$ ,  $avg$ ,  $max$ ): а)  $Raw\ MOS$ ; б)  $MOS-LQO$ . Для тестирования качества вокодеров использовались аудиофайлы артикуляционных таблиц количеством 20 записей ( $wav$ , 8000 КГц, 16 бит, моно).

### Сравнение качества вокодеров

Вокодер	Битовая скорость, бит/с	avg Raw MOS	avg MOS-LQO	min Raw MOS	min MOS-LQO	max Raw MOS	max MOS-LQO
c2	700	2,555	2,206	2,300	1,909	2,656	2,334
c2	700с	2,621	2,292	2,402	2,020	2,796	2,526
c2	1200	2,651	2,329	2,509	2,147	2,799	2,530
c2	1300	2,718	2,419	2,549	2,195	2,863	2,622
c2	1400	2,713	2,413	2,585	2,241	2,868	2,629
c2	1600	2,796	2,528	2,618	2,285	2,916	2,698
c2	2400	2,794	2,524	2,675	2,360	2,941	2,735
c2	3200	2,927	2,715	2,794	2,523	3,035	2,874
mlp	1200	3,038	2,879	2,925	2,712	3,172	3,079
mlp	2400	3,166	3,070	3,027	2,862	3,286	3,248
spx	4800	3,108	2,983	2,993	2,812	3,217	3,146



a



б

Рис. 3. Результаты тестирования качества вокодеров: а – Raw MOS; б – MOS-LQO

### Выводы

1. Показана эффективность применения объективных методов оценки качества речи.

2. В результате анализа экспериментальных данных выделим в качестве перспективного для дальнейших разработок вокодера *MELPe*, обеспечивающего на битовых скоростях 1200 и 2400 бит/с оценку качества *MOS* соответственно 2,9...3,2 и 3,0...3,3.

3. Вокодер *Codec2* показал результат согласно оценке качества *MOS* ниже, даже на скорости 3200 бит/с, по сравнению с вокодером *MELPe*.

4. Вокодер *Speex*, являющийся свободной реализацией алгоритма *CELP*, показал при тестировании оценку качества *MOS* 3,0...3,2 на битовой скорости 4800 бит/с, что почти совпадает с качеством вокодера *MELPe* на битовой скорости 2400 бит/с.

5. Необходимо помнить, что стандарт *P.862* может иметь ошибки при оценке, связанные

с особенностями реализации конкретных вокодеров, и окончательные выводы о качестве вокодеров нужно принимать на основе экспертной оценки.

### Библиографические ссылки

1. Тестирование цифровых микросхем и программирование стендового оборудования «Formula 2k» для измерения параметров / А. Н. Копысов, Р. А. Хатбуллин, В. В. Хворенков, Ф. М. Ермаков, К. А. Зырянов // Интеллектуальные системы в производстве. 2017. Т. 15, № 4. С. 29–34. DOI 10.22213/2410-9304-2017-4-29-34.

2. ГОСТ 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости.

3. ITU-T Rec. P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Available at: <http://www.itu.int/rec/T-REC-P.862> (accessed 01.11.2018).

4. Там же.
5. ITU-T Rec. P.800: Methods for subjective determination of transmission quality. Available at: <http://www.itu.int/rec/T-REC-P.800> (accessed 01.11.2018).
6. Audio File Format Specifications. WAVE or RIFF WAVE sound file. Available at: <http://www-mmmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html> (accessed 01.11.2018).
7. MELPe – Enhanced Mixed-Excitation Linear Predictive Vocoder. Available at: <http://melpe.org/> (accessed 01.11.2018).
8. Standard: NATO – STANAG 4591. The 600 bit/s, 1200 bit/s and 2400 bit/s NATO interoperable narrow band voice coder. Available at: <https://standards.globalspec.com/std/1664099/nato-stanag-4591> (accessed 01.11.2018).
9. Speex: A Free Codec For Free Speech. Available at: <https://www.speex.org/> (accessed 01.11.2018).
10. Standard: ISO/IEC 14496-3. Information technology – Coding of audio-visual objects. Part 3: Audio amendment 4: New levels for AAC profiles technical corrigendum 1. Available at: <https://standards.globalspec.com/std/9907734/iso-iec-14496-3> (accessed 01.11.2018).
11. Codec2. Available at: [http://www.rowetel.com/?page\\_id=452](http://www.rowetel.com/?page_id=452) (accessed 01.11.2018).
2. *Peredacha rechi po traktam svyazi. Metody otsenki kachestva, razborchivosti i uznavaemosti*, GOST 50840-95 [Transmission of speech through communication paths. Methods for assessing quality, intelligibility and recognizability]. (in Russ.).
3. ITU-T Rec. P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Available at: <http://www.itu.int/rec/T-REC-P.862> (accessed 01.11.2018).
4. Ibid.
5. ITU-T Rec. P.800: Methods for subjective determination of transmission quality. Available at: <http://www.itu.int/rec/T-REC-P.800> (accessed 01.11.2018).
6. Audio File Format Specifications. WAVE or RIFF WAVE sound file. Available at: <http://www-mmmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html> (accessed 01.11.2018).
7. MELPe – Enhanced Mixed-Excitation Linear Predictive Vocoder. Available at: <http://melpe.org/> (accessed 01.11.2018).
8. Standard: NATO – STANAG 4591. The 600 bit/s, 1200 bit/s and 2400 bit/s NATO interoperable narrow band voice coder. Available at: <https://standards.globalspec.com/std/1664099/nato-stanag-4591> (accessed 01.11.2018).
9. Speex: A Free Codec For Free Speech. Available at: <https://www.speex.org/> (accessed 01.11.2018).
10. Standard: ISO/IEC 14496-3. Information technology – Coding of audio-visual objects. Part 3: Audio amendment 4: New levels for AAC profiles technical corrigendum 1. Available at: <https://standards.globalspec.com/std/9907734/iso-iec-14496-3> (accessed 01.11.2018).
11. Codec2. Available at: [http://www.rowetel.com/?page\\_id=452](http://www.rowetel.com/?page_id=452) (accessed 01.11.2018).

### References

1. Kopysov A.N., Khatbullin R.A., Khvorenkov V.V., Ermakov F.M., Zyryanov K.A. [Testing of Digital Microcircuits and Programming of Bench Equipment “Formula 2k” for Measuring Parameters]. *Intellektual'nye sistemy v proizvodstve*, 2017, vol. 15, no. 4, pp. 29-34 (in Russ.). DOI 10.22213/2410-9304-2017-4-29-34.

\*\*\*

### Using the Standard P.862 to Compare the Quality of Low-Bitrate Vcoders

*A. V. Korobeynikov*, PhD in Engineering, Associate Professor, Kalashnikov ISTU, Izhevsk, Russia  
*M. A. Boyarshinov*, PhD in Engineering, Associate Professor, Kalashnikov ISTU, Izhevsk, Russia  
*A. I. Nistyuk*, DSc in Engineering, Professor, Kalashnikov ISTU, Izhevsk, Russia  
*V. N. Emelianov*, PhD in Engineering, Kalashnikov ISTU, Izhevsk, Russia

*The objective methods for assessing the speech signal quality are considered: 1) Perceptual Evaluation of Speech (PESQ, ITU-T Rec. P.862), 2) Listening Quality Objective (LQO, ITU-T Rec. P.800.1). A brief description and work schemes of the PESQ technique and formulas for converting Raw MOS estimates to MOS-LQO and back are given. Low-bitrate vocoders were chosen for testing: 1) MELPe, 2) Speex, 3) Codec2. Testing of vocoders was performed on bit speeds from 700 to 4800 bps. For testing we used audio files of articulation tables with 20 records (wav, 8000 KHz, 16 bit, mono). As a result of testing, tables and graphs for Raw MOS and MOS-LQO estimates of the selected vocoders were built. When analyzing the experiments results, the conclusion is made about the effectiveness of objective methods of speech quality assessment. MELPe was identified as a promising vocoder for further development, providing at bit rates of 1200 and 2400 bps MOS quality assessment respectively 2.9...3.2 and 3.0...3.3. Speex vocoder showed comparable with MELPe evaluation results at a higher bitrate (4800 bps). Codec2 vocoder showed lower evaluation results than MELPe.*

**Keywords:** speech compression, low-bitrate vocoders, quality assessment, standard P.862.

Получено: 12.11.18