

С.Н. КИРИЛЛОВ, В.Т. ДМИТРИЕВ  
**КОМПЛЕКСНЫЙ АЛГОРИТМ ОБЪЕКТИВНОЙ ОЦЕНКИ  
КАЧЕСТВА ДЕКОДИРОВАННОГО РЕЧЕВОГО СИГНАЛА ПРИ  
ДЕЙСТВИИ АКУСТИЧЕСКИХ ПОМЕХ**

---

*Кириллов С.Н., Дмитриев В.Т. Комплексный алгоритм объективной оценки качества декодированного речевого сигнала при действии акустических помех.*

**Аннотация.** Рассматриваются алгоритмы объективной оценки качества речи, основанные на измерении динамических и статических характеристик речевых сигналов на выходе кодека источника. Обоснована функциональная схема проведения экспериментальных исследований. Приведены результаты анализа корреляции объективной и субъективной оценки качества речи. Предложены модификации показателя объективной оценки качества на основе корреляции возбуждения спектра MESC и модификации показателя на основе вычисления функции ощущения спектральной динамики MFOSD. Предложен алгоритм формирования кривых регрессии, позволяющий выполнить преобразование объективной оценки к шкале субъективной оценки качества речи.

На основании использования наиболее точных модификаций показателей оценки качества речи для восстановленных речевых сигналов предложен комплексный алгоритм объективной аппаратной оценки качества речи при воздействии на микрофон широкополосных и низкочастотных стационарных и нестационарных акустических помех. Показано, что применение комплексного алгоритма позволяет получить объективную оценку качества речи согласно ГОСТ Р 50840-95 со средней ошибкой не более 0,35 балла при отношениях сигнал-шум от 30 дБ до -10 дБ.

**Ключевые слова:** оценка качества речи, низкоскоростное кодирование, канал связи, акустическая помеха.

---

**1. Введение.** В телекоммуникационных системах существует оперативная необходимость в объективной оценке качества речи для речевого сигнала (РС) на выходе кодека источника. Особенно это важно при функционировании системы в сложной быстро изменяющейся акустической помеховой обстановке. Кроме того, на качество речи оказывают существенное влияние искажения, возникающие в первичном кодере РС.

Одним из наиболее распространенных видов информации, передаваемых в телекоммуникационных системах, является речевая информация. Она занимает значительную часть трафика, передаваемого по современным сетям связи. Зачастую для уменьшения избыточности передаваемой по каналам связи речевой информации применяются вокодеры и полувокодеры, которые искажают исходный РС, кроме того на качество РС значительное влияние оказывают различные акустические помехи (АП).

В настоящее время субъективная оценка качества речи для восстановленного РС осуществляется в соответствии с методами,

описанными в ГОСТ Р 50840-95 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» и ГОСТ Р 51061-97 «Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений» [1, 2]. Данные методы требуют проведения значительного количества тестовых испытаний и создания голосовых баз данных большого объема, а также во многом зависят от профессионализма привлеченных экспертов. Кроме того, полученные оценки могут служить лишь первым приближением и не отражают изменения качества речи для восстановленного РС при действии АП [3] и искажений в кодеках РС.

Как показано в [4-6] существующие алгоритмы объективной аппаратной оценки качества восстановленного РС не обладают необходимой точностью и не могут в полной мере учитывать искажения, вносимые кодеками источника при действии АП.

В известных работах [7-21] не проведены подробные исследования алгоритмов объективной оценки качества восстановленного РС на выходе декодера РС при действии различных АП.

Поэтому необходимы исследования и обоснование комплексного алгоритма объективной оценки качества речи для декодированного РС, который учитывает наилучшие свойства известных алгоритмов оценки качества речи [22-25], а также изменение качества речи для восстановленного РС при действии АП на входе кодека источника.

Кодеки источника имеют разную чувствительность к действию АП, поэтому целесообразно производить совместную оценку качества речи с учетом АП и искажений РС в кодере и декодере источника.

Цель данной работы — исследование и обоснование комплексного алгоритма объективной оценки качества речи для декодированного РС при действии широкополосных и низкочастотных, стационарных и нестационарных АП на входе микрофона, а также искажений в кодеке источника.

**2. Анализ известных алгоритмов объективной оценки качества речи.** Существует ряд алгоритмов объективной оценки качества речи, основанных на измерении статических и динамических характеристик РС (например, коэффициента передачи, частотной и амплитудной характеристик, динамического диапазона и других) [6-19].

К этим алгоритмам относят показатели оценки качества речи во временной области (на основе среднеквадратического отклонения (СКО), отношения сигнал-шум (ОСШ), сегментного отношения сигнал-шум (ОСШ<sub>сег</sub>), а также в частотной области (на основе расстояния Итакура-

Саито (Itakura-Saito Distance — ISD), показателя искажения спектра барков (Bark Spectral Distortion — BSD), модифицированной оценки искажения спектра барков (Modified Bark Spectral Distortion — MBSD), расхождения COSH (COSH Distance — COSH), искажения фазового спектра (Spectral Phase Distortion — SP), корреляции возбуждения спектра (Excitation Spectral Correlation — ESC), искажения спектра (Spectral Distortion — SD)), кроме того используются показатели оценки на основе методов психоакустического анализа (Perceptual Evaluation of Speech Quality — PESQ) [7] и ряд других [5-19].

В [4, 5] дан подробный обзор показателей оценки качества речи во временной и частотной области. Проанализированы такие показатели, как ОСШ, ОСШ<sub>ср</sub>, ISD, COSH, SP, ISD, ESC. В [5] отдельно рассмотрены пять различных моделей психоакустической предварительной обработки для оценки качества речи в низкоскоростных кодеках.

В работе [7] рассмотрены различные алгоритмы оценки качества речи при передаче речевых пакетов, которые используются при исследовании надежности сети NGN. В частности, рассмотрен показатель «рассчитываемого планируемого параметра ухудшения» ICPIF (Calculated Planning Impairment Factor).

В работе [8] в качестве меры искажения РС принимается спектральное отклонение SD, а также модификация данного показателя MBSD. В дальнейшем для оценки качества речи предложен показатель при котором используется значимость критических полос. В [9] обоснована дальнейшая модификация показателя оценки качества речи — EMBSD. В работе [10] проведен анализ основных показателей оценки качества речи и их классификация как показателей на основе восприятия, а также других показателей во временной и частотной области. Приведены девятнадцать показателей качества речи для восстановленного РС.

В работе [11] исследованы семь показателей качества речи на выходе двенадцати различных систем кодирования, а также без кодирования при воздействии аддитивного белого шума на входе системы. Для отдельных отношений сигнал-шум и кодеков получены значения корреляции объективных и субъективных оценок. В результате эксперимента показано [11], что наилучшие характеристики обеспечивают алгоритм на основе автоматического распознавания речи ASR и алгоритм PESQ.

В работе [12] рассмотрены методические вопросы по оценке качества синтезированной речи, показаны основные проблемы и предложены возможные решения. В [13] приведен алгоритм оценки

интегрального значения отношения сигнал-шум и его значений в частотных полосах для определения качества фонограмм. В [14] рассмотрены методы оценки качества при потере пакетов. Проведен корреляционный анализ объективных и субъективных показателей оценки качества. В [15] предложен новый алгоритм на основе восприятия речи POLQA, который разработан взамен более простого PESQ и обеспечивает более точную оценку искажений РС. Другой перспективный алгоритм оценки качества речи на основе модели восприятия ViSQOL рассмотрен в [16]. В [17] приведен обзор инструментальных моделей для оценки качества речи на основе перцепционных и когнитивных характеристик слуховой системы человека. В [18] также рассмотрены методы и критерии психоакустической оценки качества речи в сетях связи на основе алгоритмов методы PESQ, POLQA, E-model и другие. Показана возможность применения данных алгоритмов для вычисления разборчивости речи на выходе современных сетей связи. В [19] проведено сравнение алгоритмов PESQ и POLQA и показаны преимущества последнего алгоритма в современных сетях связи.

Наиболее простые из них оценивают изменение одного из параметров РС в частотной или временной областях. Более сложные производят исследование качества речи и осуществляют комплексный анализ искажений РС и его отдельных характеристик в виде [4-10, 13, 14]:

– показатель отношения сигнал-шум [4, 5, 13]:

$$\text{ОСШ} = 10 \lg \frac{\sum_{n=1}^{N_d} x^2(n)}{\sum_{n=1}^{N_d} (x(n) - y(n))^2}, \quad (1)$$

где  $x(n)$  и  $y(n)$  —  $n$ -е отсчеты исходного и декодированного РС,  $N_d$  — количество отсчетов дискретного РС;

– показатель сегментного отношения сигнал-шум ( $\text{ОСШ}_{\text{сег}}$ ) [5, 13]:

$$\text{ОСШ}_{\text{сег}} = \frac{10}{M} \sum_{m=1}^M \lg \sum_{n=1}^{N_{\text{seg}}} \left( \frac{x^2(n, m)}{(x(n, m) - y(n, m))^2} \right), \quad (2)$$

где  $x(n, m)$  и  $y(n, m)$  —  $n$ -е отсчеты на  $m$ -м сегменте исходного и декодированного РС,  $N_{\text{seg}}$  — количество отсчетов в сегменте РС,  $M$  — число сегментов РС;

– показатель расстояния Итакура-Саито (Itakura-Saito Distance — ISD) [5, 6, 9]:

$$\text{ISD} = \frac{1}{N_d} \sum_{n=1}^{N_d} \left( \frac{|X(f_n)|}{|Y(f_n)|} - \log \frac{|X(f_n)|}{|Y(f_n)|} - 1 \right), \quad (3)$$

где  $X(f_n)$  и  $Y(f_n)$  — спектры исходного РС и декодированного РС;

– показатель искажения спектра барков (Bark Spectral Distortion — BSD) [5, 10]:

$$\text{BSD} = \sum_{g=1}^G \sum_{n=1}^{N_{b,g}} (|X(f_n, g)| - |Y(f_n, g)|)^2, \quad (4)$$

где  $X(f_n, g)$  и  $Y(f_n, g)$  —  $g$ -ая критическая полоса спектров исходного и декодированного РС,  $N_{b,g}$  — количество отсчетов в  $g$ -ой критической полосе,  $G$  — общее количество критических полос, для сигнала в полосе частот 300 — 3400 Гц  $G=18$ ;

– показатель модифицированной оценки искажения спектра барков (Modified Bark Spectral Distortion — MBSD) [8-9]

$$\text{MBSD} = \frac{1}{M} \sum_{m=1}^M \left[ P \sum_{g=1}^G \sum_{n=1}^{N_{b,g}} |X(f_n, m, g) - Y(f_n, m, g)|^2 \right], \quad (5)$$

где  $X(f_n, m, g)$  и  $Y(f_n, m, g)$  —  $g$ -ая критическая полоса спектров  $m$ -го сегмента исходного и декодированного РС,  $P$  — мера восприятия искажений (в случае, когда искажения ощутимы  $P=1$ , в других случаях  $P=0$ );

– показатель расстояния COSH (COSH Distance) [5, 11]

$$\text{COSH} = \frac{1}{N_d} \sum_{n=1}^{N_d} \left[ \frac{1}{2} \left( \frac{|Y(f_n)|}{|X(f_n)|} + \frac{|X(f_n)|}{|Y(f_n)|} - 1 \right) \right]; \quad (6)$$

– показатель искажения фазового спектра (Spectral Phase Distortion – SP) [5, 10]:

$$\text{SP} = \frac{1}{N_d} \sum_{n=1}^{N_d} (\theta_x(f_n) - \theta_y(f_n))^2, \quad (7)$$

где  $\theta_x(f_n)$  и  $\theta_y(f_n)$  — фазовые спектры исходного и декодированного РС,  $N_d$  — количество отсчетов дискретного РС;

– показатель корреляции возбуждения спектра (Excitation Spectral Correlation — ESC) [5, 12]:

$$ESC = \frac{\left( \sum_{n=1}^{N_d} |X(f_n)| \cdot |Y(f_n)| \right)^2}{\sum_{n=1}^{N_d} |X(f_n)|^2 \sum_{n=1}^{N_d} |Y(f_n)|^2}; \quad (8)$$

– показатель искажения спектра (Spectral Distortion — SD) [5, 10]:

$$SD = \frac{10}{N_d} \sum_{n=1}^{N_d} \left[ \log_{10} |X(f_n)| - \log_{10} |Y(f_n)| \right]. \quad (9)$$

Также для оценки качества речи может использоваться фонетическая функция [20], предложенная А.А. Пироговым. Физический смысл фонетической функции заключается в оценке изменения спектральной динамики речи.

Согласно [21], фонемы отличаются друг от друга, прежде всего, изменениями спектрального распределения. Фонетическая функция (функция ощущения спектральной динамики — FOSD) имеет вид [20, 21]:

$$\begin{aligned} P_H(\omega, t) &= P_2(\omega, t) - P_1(\omega, t) = \\ &= \int_0^{\infty} \exp\left(\frac{-\tau}{T}\right) \cdot \lg\left(\frac{S_2(\omega, t)}{S_2(\omega, t-\tau)}\right) d\tau - \int_0^{\infty} \exp\left(\frac{-\tau}{T}\right) \cdot \lg\left(\frac{S_1(\omega, t)}{S_1(\omega, t-\tau)}\right) d\tau, \end{aligned} \quad (10)$$

где  $S_1(\omega, t)$  — мгновенный спектр исходного РС,  $S_2(\omega, t)$  — мгновенный спектр декодированного РС,  $P_2(\omega, t) - P_1(\omega, t)$  — разность FOSD декодированного и исходного РС,  $T$  — длина анализируемого кадра РС,  $\tau$  — временной сдвиг РС.

**3. Функциональная схема проведения экспериментальных исследований** алгоритмов объективной оценки качества речи приведена на рисунке 1.

На рисунке 1 используются следующие обозначения: М — микрофон, ПФ — полосовой фильтр, ПК — программный конвертор, К — ключ, БФП — блок формирования помехи, БУ —

блок управления, КС — канал связи, БООК — блок объективной оценки качества речи, ОО, СО — объективная и субъективная оценки качества речи соответственно.

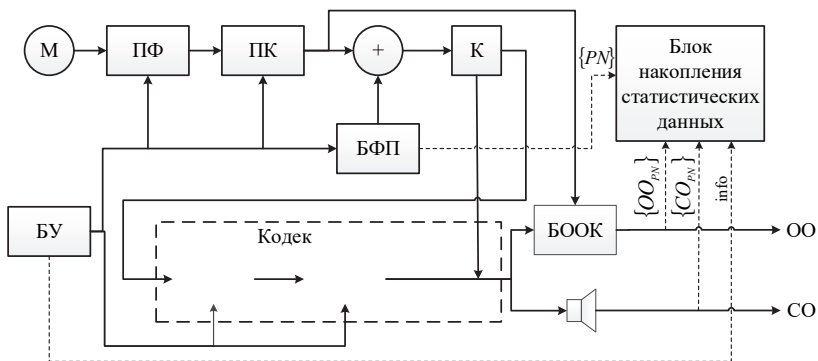


Рис.1. Функциональная схема проведения экспериментальных исследований

Регистрируемый микрофоном М тестовый РС с частотой дискретизации 44,1 кГц и разрядностью квантования 16 бит поступает на ПФ с полосой пропускания 0,3-3,4 кГц и далее преобразуется программным конвертором ПК с частотой дискретизации 8 кГц или 16 кГц согласно требованиям кодека. Блок БФП осуществляет формирование АП с учетом характеристик микрофона и ПФ с определенным уровнем, чтобы обеспечить требуемое ОСШ (-10, 0, 10, 20, 30 дБ). Выбранная в блоке БУ АП из блока БФП суммируется с РС с выхода ПК, и полученная аддитивная смесь РС и АП поступают на вход блока Кодер. При ОСШ равном 30 дБ АП мало заметны на слух и их влиянием можно пренебречь.

В блоках Кодер и Декодер происходит кодирование и декодирование РС. Декодированный РС поступает на динамик, где осуществляется СО качества речи по критериям разборчивости речи и узнаваемости голоса, а также на вход блока объективной оценки качества речи БООК, где осуществляется ОО качества речи по предложенным выше показателям. Блок БУ осуществляет выбор режимов работы таких управляемых блоков, как ПФ, ПК, БФП, Кодер и Декодер.

Ключ К имеет два режима работы: при его размыкании на вход блока БООК и динамик подается исходный РС с АП, а при его замыкании смесь РС с АП подается на вход кодера. Это позволяет отдельно оценивать искажения РС, обусловленные АП, и искажениями в кодеке, а также их совместное влияние.

В Блоке накопления статистических данных осуществляется запись и хранение условий и результатов эксперимента: параметров

помехи из БФП — PN, информацию о выбранной паре Кодер и Декодер — info, объективную оценку из БООК —  $OO_{PN}$ , а также субъективные оценки экспертов —  $CO_{PN}$ .

Для проверки эффективности алгоритмов объективной оценки качества был собран массив PC — около 40 различных записей для каждого из 40 дикторов, полученных согласно ГОСТ Р 50840-95 [1], так и фрагментов спонтанной речи.

Запись PC осуществлялась в специальном помещении кабинетного типа (с размерами 5,7\*2,9\*3 м и временем реверберации порядка 350 мс) при наличии естественного фонового шума слабого уровня. Для записи использовался профессиональный диктофон Olympus LS-10 (Linear PCM recorder), обеспечивающий возможность записи PC в формате WAV со следующими параметрами: частота дискретизации — 44,1 кГц; разрядность квантования — 16 бит; тип кодирования — импульсно-кодовая модуляция (ИКМ). Диктофон был установлен на расстоянии 0,5 м перед диктором на уровне лица. Общее количество незашумленных записанных фраз составило 1600.

В связке блоков Кодер-Декодер (рисунки 1) использовались 24 вида кодеков (LBRAMR, MMBE, G723.1, ICELP, G729a, G728i, AMR, GSM, G726, Vorbis OGG, MPEG 1.2.8) со скоростями передачи 1...64 кбит/с.

Проводились исследования оценки качества речи при отсутствии блока Кодек, то есть при воздействии АП на PC на входе микрофона.

В дальнейшем аналогичные исследования проводились при подаче PC и АП на вход кодека.

АП по-разному влияют на характеристики первичных кодеков, поэтому важна совместная оценка влияния АП и искажений, возникающих в кодеках источника на качество речи.

Исходя из [3] и других источников [22-24] целесообразно проводить анализ качества речи для следующих четырех классов АП, которые формировались в БФП в пределах частот 0,3–3,4 кГц:

- широкополосная стационарная, в качестве которой использовался белый гауссовский шум (БГШ);
- низкочастотная стационарная в виде БГШ на выходе фильтра низких частот, амплитудно-частотная характеристика которого имеет частоту среза 2 кГц и наклон 9 дБ/октаву в сторону высоких частот;
- низкочастотная нестационарная, представляющая собой нарезку реализаций шума транспортного потока;
- широкополосная нестационарная в виде записи музыки.

Восстановленный PC поступает на акустические колонки наряду с незашумленным PC с выхода ПФ, который служил в качестве образца, и в соответствии с ГОСТ Р 50840-95 производилась оценка



качества восстановленной на выходе декодера речи в баллах по пятибальной шкале десятью аудиторами.

Данные исследования были проведены для всех исследуемых кодеков, для десяти фраз, рекомендованных ГОСТ Р 50840-95 и тридцати фрагментов спонтанной речи, сорока дикторов при действии каждого вида АП с различными уровнями ОСШ (30, 20, 10, 0, -10 дБ), после чего было проведено усреднение по всем дикторам и аудиторам. Полученные зависимости субъективного качества речи К от ОСШ сравнивались с аналогичными зависимостями объективных оценок качества полученных в БООК и сохраненных в Блоке накопления статических данных.

**4. Исследования корреляции объективной и субъективной оценки качества речи.** В качестве меры эффективности объективного алгоритма использовалась степень корреляции между полученными объективными и субъективной оценкой качества речи, согласно выражению [11]:

$$R_{corr} = \frac{\left( \sum_{k=1}^K (X_1(k) - \overline{X_1})(X_2(k) - \overline{X_2}) \right)}{\sqrt{\sum_{k=1}^K (X_1(k) - \overline{X_1})^2 \sum_{k=1}^K (X_2(k) - \overline{X_2})^2}}, \quad (11)$$

где  $X_1(k)$ ,  $X_2(k)$  —  $k$ -ые значения массивов объективных и субъективных оценок качества речи,  $\overline{X_1}$ ,  $\overline{X_2}$  — средние значения массивов объективных и субъективных оценок качества речи,  $k$  — номер массива реализаций.

В таблице 1 приведены усредненные значения корреляций объективных и субъективных оценок качества речи (темным цветом в таблице выделены максимальные коэффициенты корреляции для различных алгоритмов оценки качества речи и кодеков) при действии различных видов АП в диапазоне отношений сигнал–шум от 30 до -10 дБ для кодеков ММВЕ при скорости передачи 1,2 кбит/с, ICELP при скорости передачи 4,8 кбит/с, G723.1 со скоростью передачи 5,3 кбит/с, G729а со скоростью передачи 8 кбит/с. Аналогичные исследования проведены для других вышперечисленных кодеков РС как при наличии АП, так и при их отсутствии.

Как видно, из анализа полученных данных, наибольший коэффициент корреляции обеспечивают показатели оценки ESC и FOSD.

Таблица 1. Коэффициенты корреляции показателей объективной и субъективной оценки качества речи

Алгоритмы \ Кодеки	MMBE	ICELP	G723.1	G729a
Скорость, кбит/с	1,2	4,8	5,3	8
СКО	0,7	0,97	0,97	0,90
ОСШsegm	0,72	0,96	0,98	0,96
ESC	0,82	0,99	0,89	0,89
SD	0,67	0,98	0,97	0,97
SP	0,26	0,07	0,57	0,29
ISD	0,62	0,94	0,71	0,21
COSH	0,56	0,92	0,6	0,15
BSD	0,89	0,91	0,81	0,85
MBSD	0,90	0,96	0,93	0,96
FOSD	0,93	0,88	0,84	0,95

Из алгоритмов оценки качества речи во временной области, которые применяются главным образом для кодеков формы со скоростью передачи РС 32...64 кбит/с наибольший коэффициент корреляции с субъективной оценкой обеспечивал показатель оценки качества на основе отношения сигнал-шум сегментное (signal-to-noise ratio, сокр. ОСШ). Для других методов кодирования сильнее коррелированы результаты объективных оценок при использовании более сложных алгоритмов.

**5. Модификация показателя ESC.** Среди методов объективной оценки качества речи для алгоритмов осуществляющих обработку в частотной области выделяется показатель корреляции возбуждения спектра ESC. Рассмотрим возможную модификацию данного показателя в целях более точного его соответствия субъективным оценкам РС.

В случае наличия исходного и декодированного РС предложен показатель MESC [22, 23], осуществляющий вычисление оценки качества речи в пределах критических полос спектра с учетом коэффициентов их значимости:

$$MESC = \frac{1}{G} \sum_{g=1}^G \left( \frac{\left( \sum_{n=1}^{N_{b,g}} \beta_g |X(f_n, g)| \cdot |Y(f_n, g)| \right)^2}{\left( \sum_{n=1}^{N_{b,g}} \beta_g |X(f_n, g)|^2 \sum_{n=1}^{N_{b,g}} \beta_g |Y(f_n, g)|^2 \right)} \right), \quad (12)$$

где  $\beta_g$  — коэффициент «значимости» спектральных составляющих  $g$ -ой критической полосы,  $X(f_n, g)$  и  $Y(f_n, g)$  —  $g$ -ая критическая

полоса спектров исходного и декодированного РС,  $N_{b,g}$  — количество отсчетов в  $g$ -ой критической полосе,  $G$  — общее количество критических полос.

Другая возможная модификация показателя ESC при отсутствии исходного РС имеет вид:

$$MESOC2 = \frac{1}{G} \sum_{g=1}^G \left( \frac{\left( \sum_{n=1}^{N_{b,g}} \beta_g |X_0(f_n, g)| \cdot |Y(f_n, g)| \right)^2}{\left( \sum_{n=1}^{N_{b,g}} \beta_g |X_0(f_n, g)|^2 \sum_{n=1}^{N_{b,g}} \beta_g |Y(f_n, g)|^2 \right)} \right), \quad (13)$$

где  $X_0(f_n)$  — модель усредненного спектра РС.

Согласно [5, 6], усредненный спектр РС определяется в соответствии со следующим выражением:

$$X_0(f_n) = 4\sigma^2 \cdot (f_n/f_0) / \left[ 1 + (f_n/f_0)^2 \right]^2, \quad (14)$$

где  $\sigma^2$  — дисперсия выборок РС,  $f_n$  —  $n$ -ая спектральная составляющая РС в пределах от 0,1 до 8,0 кГц,  $f_0$  — частота основного тона речи, равна 400 Гц.

Коэффициенты «значимости» критической полосы  $\beta_g$ , выделенной согласно рекомендациям [25], оценивались как отношение энергии РС, сосредоточенной в критической полосе  $E_g$  к энергии  $E_{X_0}$  усредненного спектра РС (14), в виде:

$$\beta_g = \{E_g / E_{X_0}\}. \quad (15)$$

В соответствии с выражением (15), для каждой  $g$ -ой из критических полос, которые определялись в соответствии с [25], были получены следующие коэффициенты «значимости»  $\beta_g$  (таблица 2).

Таблица 2. Коэффициенты значимости критических полос

$g$	$\beta_g$	$g$	$\beta_g$	$g$	$\beta_g$
1	0,0049	7	0,0890	13	0,0409
2	0,0371	8	0,0775	14	0,0355
3	0,0685	9	0,0653	15	0,0317
4	0,0837	10	0,0611	16	0,0282
5	0,0926	11	0,0521	17	0,0254
6	0,0908	12	0,0455	18	0,0235

При экспериментальных исследованиях [22, 23] было показано, что показатель MESC обеспечивает коэффициент корреляции для кодека ММВЕ — 0,9, для кодека ICELP — 0,99, для кодека G723.1 — 0,98 и для кодека G729a — 0,96.

Таким образом, предложенная модификация показателя ESC увеличивает значения коэффициента корреляции в среднем на 0,08 при действии всех рассмотренных видов АП.

**6. Модификация показателя FOSD.** В случае наличия исходного и декодированного РС предложен показатель MFOSD (модификация показателя оценки FOSD (10)), осуществляющий вычисление изменения спектральной динамики критических полос спектров соседних сегментов РС по формуле:

$$\begin{aligned}
 MFOSD &= \sum_{m=2}^M \sum_{g=1}^G \beta_g \sum_{n=1}^{N_{b,g}} P_2(f_n, g, N_{seg} m) = \\
 &= \sum_{m=2}^M \sum_{g=1}^G \beta_g \sum_{n=1}^{N_{b,g}} \left( \frac{|Y(f_n, g, N_{seg} m)|}{|Y(f_n, g, N_{seg} (m-1))|} - \frac{|X(f_n, g, N_{seg} m)|}{|X(f_n, g, N_{seg} (m-1))|} \right), \quad (16)
 \end{aligned}$$

где  $M$  — количество сегментов РС,  $P_2(f_n, g, m)$  — показатель FOSD для  $g$ -ой критической полосы спектра сегмента  $m$  с длиной  $N_{seg}$ ,  $\beta_g$  — коэффициент «значимости» спектральных составляющих  $g$ -ой критической полосы,  $X(f_n, g, m)$  и  $Y(f_n, g, m)$  —  $g$ -ая критическая полоса спектра сегмента  $m$  с длиной  $N_{seg}$  исходного и декодированного РС. При экспериментальных исследованиях показано, что при отсутствии исходного РС остается возможным определение разборчивости речи, при этом модернизация показателя FOSD имеет вид:

$$\begin{aligned}
 MFOSD &= \sum_{m=2}^M \sum_{g=1}^G \beta_g \sum_{n=1}^{N_{b,g}} P_2(f_n, g, N_{seg} m) = \\
 &= \sum_{m=2}^M \sum_{g=1}^G \beta_g \sum_{n=1}^{N_{b,g}} \frac{|Y(f_n, g, N_{seg} m)|}{|Y(f_n, g, N_{seg} (m-1))|}. \quad (17)
 \end{aligned}$$

Показано, что при экспериментальных исследованиях [22, 23] показатель MFOSD обеспечивает коэффициент корреляции для кодека ММВЕ — 0,95, для кодека ICELP — 0,89, для кодека G723.1 — 0,88 и для кодека G729a — 0,97.

Таким образом, предложенная модификация показателя FOSD увеличивает значение коэффициента корреляции в среднем на 0,03 при действии всех рассмотренных видов АП.

**7. Формирование кривых регрессии.** Немаловажным элементом получения оценки качества речи является приведение шкалы объективной оценки полученных данных к шкале субъективной оценки, более удобной для восприятия человеком.

Для этих целей формируется так называемая «кривая регрессии» — массив коэффициентов полинома для преобразования шкалы [22]. На рисунке 2 показана схема получения коэффициентов кривой регрессии на этапе накопления статистических данных.

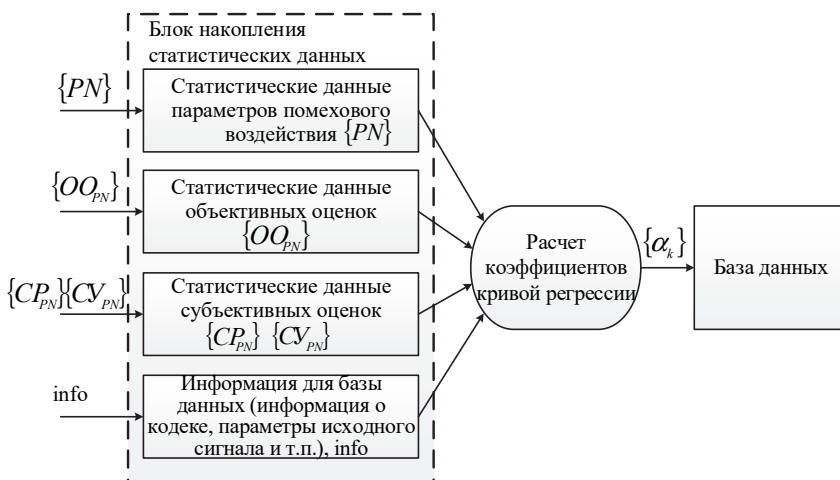


Рис. 2. Схема получения коэффициентов кривой регрессии на этапе накопления статистических данных

Исходными статистическими данными в пределах исследования одного кодека являются:

- массив информации о значении переменного параметра помехового воздействия  $\{PN\}$ , из блока БФП (рисунок 2);
- массив объективной оценки  $\{OO_{PN}\}$ , соответствующий массиву  $\{PN\}$ , из блока БООК (рисунок 2);
- массивы субъективных оценки разборчивости речи  $\{CP_{PN}\}$  соответствующие массиву  $\{PN\}$ , при проведении субъективного оценивания качества речи (рисунок 2);
- информация для учета в базе данных  $info$ , из блока БУ (рисунок 2).

Преобразование шкалы задается в виде полинома:

$$CO' = \sum_{k=1}^K \alpha_k \cdot OO^{k-1}, \quad (18)$$

где  $\alpha_k$  —  $k$ -ый коэффициент кривой регрессии,  $OO$  — значение объективной оценки качества речи,  $K$  — порядок кривой регрессии. Набор коэффициентов  $\alpha_k$  передается для хранения в базу данных с меткой `info` для учета конкретных параметров исследования (наименование и параметры кодека, параметры РС и т.п.).

Объективная оценка, полученная с блока БООК и массив коэффициентов  $\{\alpha_k\}$  с блока базы данных, позволяют преобразовать объективную оценку качества речи к шкале субъективной оценки (рисунок 3).

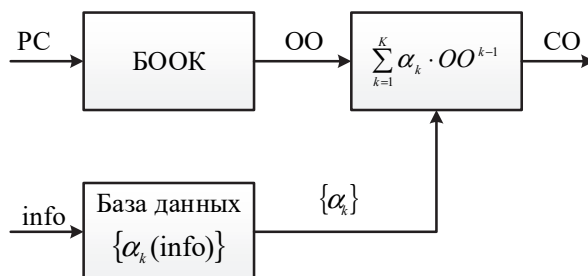


Рис. 3. Преобразование объективной оценки качества речи к шкале субъективной оценки

**8. Комплексный алгоритм оценки качества речи на выходе кодеков РС.** Для оценки качества речи на выходе различных кодеков источника, как при отсутствии, так и при наличии только АП, были выбраны показатели, обеспечивающие наиболее достоверную оценку в частотной и временной областях, а также в области изменения спектральной динамики: показателя MESC, который обеспечивает равномерно высокую корреляцию среди всех алгоритмов оценки в частотной области; отношение сигнал-шум сегментное, обеспечивающее высокую точность для основных кодеков формы, и модификация показателя FOSD (MFOSD) — в области спектральной динамики [22].

В качестве критерия для выбора конкретных показателей являлись близость полученных объективных оценок к субъективной, возможность оценки сигнала во временной и спектральной области, а

также в области спектральной динамики. Кроме того, учитывалась простота реализации представленных методов.

Оценка качества речи проводилась на основе трех показателей с учетом воздействия на каждый из них АП:

$$k = \alpha k_1 SNR_{seg} + \beta k_2 MESC + \gamma k_3 MFOSD, \quad (19)$$

где  $k_1, k_2, k_3$  — нормировочные коэффициенты. Коэффициенты  $k_1, k_2, k_3$  необходимы для нормировок вклада методов объективной оценки, так как эти оценки имеют различный порядок и изначально несопоставимы,  $\alpha, \beta, \gamma$  — коэффициенты, с помощью которых осуществляется учет различного вклада каждого из показателей оценки качества речи  $SNR_{seg}$ , MESC и MFOSD при действии различных АП, для которых выполняется условие  $\alpha + \beta + \gamma = 1$ . Структурная схема предложенного алгоритма изображена на рисунке 4.

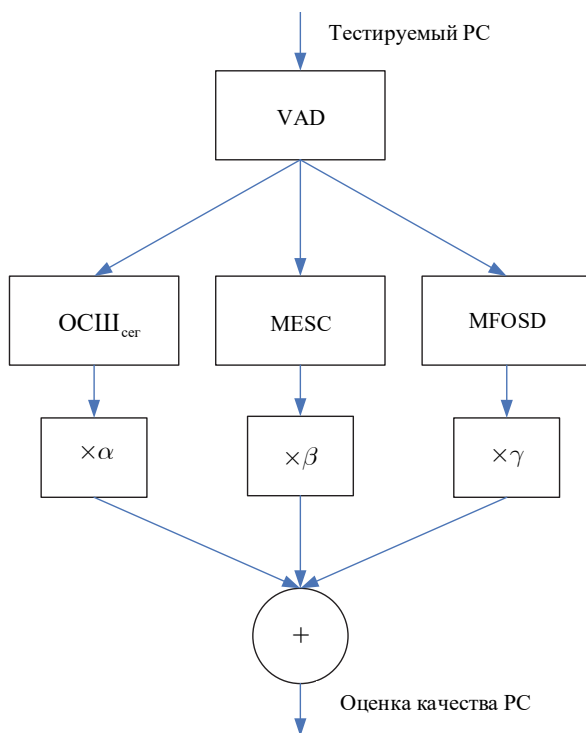


Рис. 4. Структурная схема комплексного алгоритма объективной оценки качества речи

Согласно показной на структурной схеме комплексного алгоритма объективной оценки качества речи, тестируемый сигнал сначала подается на схему VAD, в которой производится отброс пауз РС, для оценки только содержательной части сообщения. Затем РС без пауз подается параллельно на три схемы оценки — где происходит оценка качества речи по каждому из показателей  $SNR_{seg}$ , MESCC и MFOSD. Затем вклады этих методов в общую оценку суммируются с учетом их коэффициентов  $\alpha, \beta, \gamma$ . Набор коэффициентов определяется для каждого вида АП.

**9. Результаты экспериментальных исследований комплексного алгоритма оценки качества речи при действии АП.** В соответствии с приведенной на рисунке 4 структурной схемой комплексного алгоритма объективной оценки качества речи проведен ряд экспериментальных исследований.

Осуществлен выбор значений коэффициентов  $\alpha, \beta, \gamma$  (с шагом 0,1) при которых коэффициент корреляции между объективной оценкой предложенного комплексного алгоритма и субъективной оценкой качества речи достигал максимальных значений. При широкополосной квазистационарной помехе значения коэффициентов должны быть  $\alpha = \beta = 0,2, \gamma = 0,6$ ; при широкополосной нестационарной помехе —  $\alpha = 0,1, \beta = 0,9, \gamma = 0,1$ ; при низкочастотной квазистационарной помехе —  $\alpha = 0,1, \beta = \gamma = 0,45$ ; при низкочастотной нестационарной помехе —  $\alpha = 0,25, \beta = 0,5, \gamma = 0,25$ . Также можно определить коэффициенты  $\alpha, \beta, \gamma$  для каждого показателя кодирования РС.

При использовании комплексного алгоритма объективной оценки качества речи для декодированного РС при действии АП достаточно определить тип действующей помехи и вид кодера РС, чтобы получить необходимые значения коэффициентов  $\alpha, \beta, \gamma$ .

На основе полученных коэффициентов при условии что известен тип действующей АП, возможно оценить объективное качество РС и сравнить его с субъективными оценками, полученными согласно ГОСТ Р 50840-95 [1].

Разница  $\Delta$  между субъективной и объективной оценкой, полученной с помощью комплексного алгоритма оценки качества речи, приведенная к пятибальной шкале оценки и усредненная для всех видов АП и кодеков РС в зависимости от отношения сигнал-шум ОСШ показана на рисунке 5.

Как видно из приведенной зависимости ошибка оценки качества речи согласно ГОСТ Р50840-95 не превышает 0,35 балла при отношении сигнал-шум от 30 дБ до -10 дБ.



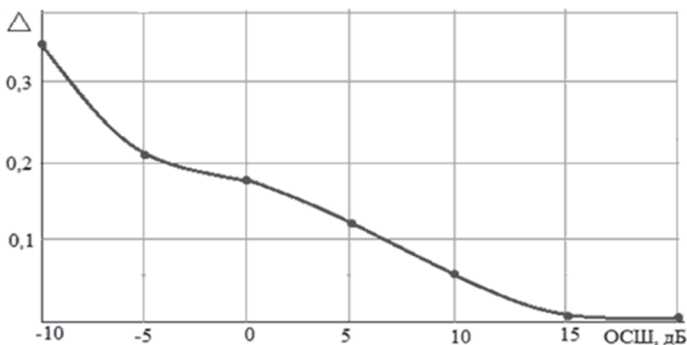


Рис. 5. Зависимость разницы между субъективной и объективной оценкой качества речи полученной с помощью комплексного алгоритма от отношения сигнал-шум

Экспериментальные исследования воздействия АП на качество речи проводились при отсутствии и наличии блока КОДЕК (рисунок 1). При отсутствии блока КОДЕК влияние АП сказывалось больше во временной и частотной областях, поэтому наибольший вес имели показатели  $SNR_{seg}$  и MESC. Полученные оценки качества речи на выходе кодеков РС в зависимости от отношения сигнал-шум показаны в работе [22].

Аналогичные исследования проводились при наличии блока КОДЕК. Результаты проведенных исследований показали:

- возможность использования комплексного алгоритма объективной оценки качества по формуле (19). Показано, что ошибка оценки качества речи согласно ГОСТ Р50840-95 не превышает 0,35 балла при отношении сигнал-шум от 30 дБ до -10 дБ,

- возможность адаптации предложенного комплексного алгоритма объективной оценки качества к известным видам АП и алгоритмам первичного кодирования РС. Показано, что при выборе коэффициентов, соответствующих определенному виду АП или виду первичного кодирования ошибка снижается более чем на 20%;

- уменьшение ошибки оценки качества речи при воздействии АП в случае отсутствия кодека РС до 50%;

- наименьшие искажения в тракт передачи вносят кодеки формы на основе ИКМ и АДИКМ со скоростью передачи не менее 32 кбит/с, а также кодеки на основе алгоритма CELP;

- наибольшие искажения в тракт передачи вносят низкоскоростные кодеки LBRAMR, ММВЕ при скоростях передачи 2 кбит/с и ниже;

- наиболее чувствительными к действию АП являются кодеки LBRAMR, ММВЕ со скоростями передачи ниже 2,4 кбит/с;

– наименее чувствительные к действию АП кодеки на основе алгоритма CELP и высокоскоростные кодеки формы.

**10. Заключение.** Рассматриваются методы объективной оценки качества речи, основанные на измерении динамических и статических характеристик РС на выходе кодека источника. Обоснована функциональная схема проведения экспериментальных исследований. Приведены результаты анализа корреляции объективной и субъективной оценки качества речи. Предложены модификации показателя корреляции возбуждения спектра MESC и модификация показателя функции ощущения спектральной динамики MFOSD. Предложен алгоритм формирования кривых регрессии, позволяющий выполнить преобразование объективной оценки к шкале субъективной оценки качества речи.

На основании использования наиболее точных методов оценки качества речи восстановленных РС предложен комплексный алгоритм объективной аппаратурной оценки качества речи согласно ГОСТ Р 50840-95 при воздействии на микрофон широкополосных и низкочастотных стационарных и нестационарных АП. Показано, что применение комплексного алгоритма позволяет получить объективную оценку качества речи согласно ГОСТ Р 50840-95 со средней ошибкой не более 0,35 балла при отношениях сигнал-шум от 30 дБ до -10 дБ как при отсутствии так и при наличии кодека РС.

### Литература

1. ГОСТ Р 50840-95. Передача речи по трактам связи // М.: Госстандарт России. 1995. 202 с.
2. ГОСТ Р 51061-97 «Системы низкоскоростной передачи речи по цифровым каналам // М.: Госстандарт России. 1997. 24 с.
3. *Кропотов Ю.А., Ермолаев В.А.* Моделирование и методы исследований акустических сигналов, шумов и помех в системах телекоммуникаций: монография // М. Берлин: Директ-Медиа. 2016. 251 с.
4. *Цыбулин М.К., Бочаров М.О.* Анализ методов оценки качества передачи речевой информации по каналам связи различной структуры // Электросвязь. 2008. №11. С. 46–48.
5. *Шелухин О.И., Лукьянцев Н.Ф.* Цифровая обработка и передача речи / под ред. О.И. Шелухина // М.: Радио и связь. 2000. 456 с.
6. *Афанасьев А.А., Ильюшин М.В.* Использование психоакустической модели слуха при разработке вокодеров с линейным предсказанием // Цифровая обработка сигналов. 2011. № 1. С. 49–52.
7. *Мурадова А. А.* Методы оценки качества передачи речевых пакетов при исследовании надежности сети NGN // Молодой ученый. 2013. № 10. С. 162–168.
8. *Павловец А.Н., Петровский А.А.* Использование закономерностей психоакустики в процедуре квантования параметров гармонической модели речевого сигнала // Речевые технологии. 2008. № 4. С. 55–60.
9. *Yang W.* Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based On Audible Distortion And Cognition Model // Temple University. 1999. 164 p.
10. *Ozer H., Avciabas I., Sankur B., Memon N.* Steganalysis of audio based on audio quality metrics, in: SPIE Electronic Imaging Conf. on Security and Watermarking of Multimedia Contents // Proceedings of SPIE. 2003. vol. 5020. pp. 55–66.

11. *Liu W.M. et al.* Assessment of objective quality measures for speech intelligibility estimation // 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006). 2006. vol. 1. pp. 699–702.
12. *Соломенник А.И. и др.* Оценка качества синтезированной речи: проблемы и решения // Изв. Вузов. Приборостроение. 2013. Т. 56. № 2. С. 38–42.
13. *Столбов М.Б.* Алгоритм оценки отношения сигнал/шум речевых сигналов // Научно-технический вестник информационных технологий, механики, оптики. 2012. № 6(82). С. 67–72.
14. *Lee M.-K., Kang H.-G.,* Speech quality estimation of voice over internet protocol codec using a packet loss impairment model // J. Acoust. Soc. Am. 2013. vol. 134(5). pp. 438–444.
15. *Beerends J.G. et al.* Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement part I—Temporal alignment // Journal of the Audio Engineering Society. 2013. vol. 61(6). pp. 366–384.
16. *Hines A. et al.* Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013. pp. 3697–3701.
17. *Möller S, Heusdens R,* Objective estimation of speech quality for communication systems // Proc. of the IEEE. 2013. vol. 101(9). pp. 1955–1967.
18. *Терехов А.Н., Рысин Ю.С.* Алгоритм определения факторов, влияющих на качество восприятия телефонных услуг связи // Электросвязь. 2016. № 3. С. 65–68.
19. *Poča P., Melvin H., Hines A.* An analysis of the impact of playout delay adjustments introduced by voip jitter buffers on listening speech quality // Acta Acustica united with Acustica. 2015. no. 101(3). pp. 616–631.
20. *Пирогов А.А.* Вокодерная телефония // М.: Связь. 1974. 536 с.
21. *Соболев В.Н.* Информационные технологии в синтетической акустике // М.: ИРИАС. 2007. 360 с.
22. *Кириллов С.Н., Ромашкин Ю.Н., Картавенко Я.О., Дмитриев В.Т.* Аналитическая оценка качества речи на выходе систем низкоскоростного кодирования при воздействии акустических помех // Речевые технологии. 2012. № 4. С. 16–23.
23. *Кириллов С.Н., Дмитриев В.Т., Картавенко Я.О.* Алгоритм объективной оценки качества декодированного речевого сигнала на основе изменения спектральной динамики критических полос спектра // Вестник РГРТУ. 2011. № 3(37). С. 3–7.
24. *Илюшин М.В., Дмитриев В.Т., Басов О.О., Тарусов В.А.* Качество передачи речи и его оценка: учебное пособие /под общ. ред. С. Н. Кириллова // Орел: Академия ФСО России. 2015. 104 с.
25. *Цешкер Э., Фельдкеллер Р.* Ухо как приемник информации: 2-е изд., перераб. и доп. // М.: Связь. 1971. 255 с.

**Кириллов Сергей Николаевич** — д-р техн. наук, профессор, заведующий кафедрой радиоуправления и связи факультета радиотехники и телекоммуникаций, ФГБОУ ВО "Рязанский государственный радиотехнический университет" (РГРТУ). Область научных интересов: теория многокритериального синтеза сложных дискретных сигналов и устройств их обработки при действии мешающих факторов, обработка речевых сигналов. Число научных публикаций — 650. kirillov.rsreu@gmail.com; ул. Гагарина, 59/1, 390005, Рязань; р.т.: +7(4912)46-03-62, Факс: +7(4912)92-22-15.

**Дмитриев Владимир Тимурович** — к-т техн. наук, доцент, доцент кафедры радиоуправления и связи факультета радиотехники и телекоммуникаций, ФГБОУ ВО "Рязанский государственный радиотехнический университет" (РГРТУ). Область научных интересов: методы и алгоритмы обработки речевых сигналов, адаптивные системы обработки. Число научных публикаций — 200. vol77@rambler.ru; ул. Гагарина, 59/1, 390005, Рязань; р.т.: +7(4912)46-03-62, Факс: +7(4912)92-22-15.

**Поддержка исследований.** Работа выполнена при поддержке Российского фонда фундаментальных исследований (№16-47-620544\16) и Правительства Рязанской области.

S.N. KIRILLOV, V.T. DMITRIEV  
**A COMPLEX ALGORITHM FOR OBJECTIVE EVALUATION OF  
 THE DECODED SPEECH SIGNAL QUALITY UNDER THE  
 ACTION OF ACOUSTIC INTERFERENCE**

---

*Kirillov S.N., Dmitriev V.T. A Complex Algorithm for Objective Evaluation of the Decoded Speech Signal Quality Under the Action of Acoustic Interference.*

**Abstract.** The paper presents algorithms for objective evaluation of speech quality are considered, based on the measurement of dynamic and static characteristics of speech signals at the source codec output. The functional scheme of carrying out of experimental researches is proved. The results of the analysis of the correlation of objective and subjective evaluation of speech quality are given. Modifications of the objective quality assessment are proposed on the basis of the correlation of the excitation of the MESC spectrum and modification of the exponent on the basis of the calculation of the sensation function of the spectral dynamics of MFOSD. The algorithm of regression curve formation is proposed, which allows to perform the transformation of objective evaluation to the scale of subjective evaluation of speech quality.

Based on the use of the most accurate modifications of the speech quality assessment indicators for reconstructed speech signals, a complex algorithm for objective hardware evaluation of the speech quality is proposed when the broadband and low-frequency stationary and nonstationary acoustic waves are applied to the microphone. It is shown that the use of a complex algorithm makes it possible to obtain an objective evaluation of the quality of speech according to GOST R 50840-95 with an average error of no more than 0.35 points for signal-to-noise ratios of 30 dB to -10 dB.

**Keywords:** speech quality estimation, low-speed coding, communication channel, acoustic interference.

---

**Kirillov Sergey Nikolaevich** — Ph.D., Dr. Sci., professor, head of department of radio control and communication department of radio engineering and telecommunications faculty, Ryazan State Radioengineering University. Research interests: theory of multi-criterial synthesis of complex discrete signals and devices for their processing under the influence of interfering factors; processing of speech signals. The number of publications — 650. kirillov.rsreu@gmail.com; 59/1, Gagarina street, 390005, Rjazan', Russia; office phone: +7(4912)46-03-62, Fax: +7(4912)92-22-15.

**Dmitriev Vladimir Timurovich** — Ph.D., associate professor, associate professor of radio control and communication department of radio engineering and telecommunications faculty, Ryazan State Radioengineering University. Research interests: methods and algorithms for processing speech signals, adaptive processing systems. The number of publications — 200. vol77@rambler.ru; 59/1, Gagarina street, 390005, Rjazan', Russia; office phone: +7(4912)46-03-62, Fax: +7(4912)92-22-15.

**Acknowledgement.** This research is supported by RFBR (grant 16-47-620544 \ 16) and the Government of the Ryazan Region.

### References

1. GOST R 50840-95. Peredacha rechi po traktam svyazi [State Standard R 50840-95. Transmission of speech through communication paths]. M.: Gosstandart Rossii, 1995. 202 p. (In Russ.).

2. GOST R 51061-97 Sistemy nizkoskorostnoy peredachi rechi po tsifrovym kanalams [State Standard R 51061-97. Systems for low-speed speech transmission over digital channels']. M.: Gosstandart Rossii. 1997. 24 p. (In Russ.).
3. Kropotov Yu.A., Ermolaev V.A. *Modelirovaniye i metody issledovaniy akusticheskih signalov, shumov i pomekh v sistemakh telekommunikatsiy: monografiya* [Modeling and methods for studying acoustic signals, noise and interference in telecommunication systems: monograph]. M.: Berlin: Direct Media, 2016. 251 p. (In Russ.).
4. Tsybulin M.K., Bocharov M.O. [Analysis of methods for assessing the quality of voice information transmission through communication channels of different structures]. *Elektrosvyaz' – Telecommunications*. 2008. no. 11. pp. 46-48. (In Russ.).
5. Shelukhin O.I., Lukyantsev N.F. *Tsifrovaya obrabotka i peredacha rechi. Pod. red. O.I. Shelukhina* [Digital processing and voice transmission. Edited by O.I. Shelukhin]. M.: Radio and Communication. 2000. 456 p. (In Russ.).
6. Afanasyev AA, Ilyushin MV [Use of the psychoacoustic hearing model in the development of vocoders with linear prediction]. *Tsifrovaya obrabotka signalov. – Digital signal processing*. 2011. vol. 1. pp. 49–52.
7. Muradova A.A. [Methods for assessing the quality of voice packets transmission when investigating the reliability of the NGN network]. *Molodoy ucheniy – Young Scientist*. 2013. vol. 10. pp. 162–168. (In Russ.).
8. Pavlovets A.N., Petrovsky A.A. [Use of patterns of psychoacoustics in the procedure for quantizing the parameters of a harmonic model of a speech signal]. *Rechevyye tekhnologii – Speech technologies*. 2008. vol. 4. pp. 55–60. (In Russ.).
9. Yang W. Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based On Audible Distortion And Cognition Model. Temple University. 1999. 164 p.
10. Ozer H., Avcibas I., Sankur B., Memon N. Steganalysis of audio based on audio quality metrics, in: SPIE Electronic Imaging Conf. on Security and Watermarking of Multimedia Contents. Proceedings of SPIE. 2003. vol. 5020. pp. 55–66.
11. Liu W.M. et al. Assessment of objective quality measures for speech intelligibility estimation. 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2006. vol. 1. pp. 699–702.
12. Solomennik A.I. et al. Evaluation of the quality of synthesized speech: problems and solutions. *Izv. High schools. Priborostroenie – Journal of Instrument Engineering* 2013. vol. 56. no. 2. pp. 38–42 (In Russ.).
13. Stolbov M. [Algorithm for estimating the signal-to-noise ratio of speech signals]. *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki, optiki – Scientific-technical bulletin of information technologies, mechanics, optics*. 2012. vol. 6(82). pp. 67–72. (In Russ.).
14. Lee M.-K., Kang H.-G., Speech quality estimation of voice over internet protocol codec using a packet loss impairment model. *J. Acoust. Soc. Am*. 2013. vol. 134(5). pp. 438–444.
15. Beerends J.G. et al. Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement part I—Temporal alignment. *Journal of the Audio Engineering Society*. 2013. vol. 61(6). pp. 366–384.
16. Hines A. et al. Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013. pp. 3697–3701.
17. Möller S, Heusdens R, Objective estimation of speech quality for communication systems. Proc. of the IEEE. 2013. vol. 101(9). pp. 1955–1967.
18. Terekhov A.N., Rysin Yu.S. [Algorithm for determining factors affecting the quality of perception of telephone communication services]. *Elektrosvyaz' – Telecommunications*. 2016. vol. 3. pp. 65–68. (In Russ.).

19. Počta P., Melvin H., Hines A. An analysis of the impact of playout delay adjustments introduced by voip jitter buffers on listening speech quality. *Acta Acustica united with Acustica*. 2015. no. 101(3). pp. 616–631.
20. Pirogov A.A. *Vokodernaya telefoniya* [Vocoderal telephony]. Moscow: Communications. 1974. 536 p. (In Russ.).
21. Sobolev V.N. *Informatsionnyye tekhnologii v sinteticheskoy akustike* [Information technology in synthetic acoustics]. Moscow: IRI-AS. 2007. 360 p. (In Russ.).
22. Romashkin Yu.N., Kirillov SN, Kartavenko Ya.O., Dmitriev V.T. [Analytical evaluation of speech quality at the output of low-speed coding systems under the influence of acoustic noise]. *Rechevyye tekhnologii – Speech technologies* 2012. vol. 4. pp. 16–23. (In Russ.).
23. Kirillov S.N., Dmitriev V.T., Kartavenko Ya. [Algorithm of objective evaluation of the quality of the decoded speech signal on the basis of a change in the spectral dynamics of critical bands of the spectrum]. *Vestnik RGRU – Bulletin of the RGRU*. 2011. vol. 3(37). pp. 3–7. (In Russ.).
24. Ilyushin M.V., Dmitriev V.T., Basov O.O., Tarusov V.A. *Kachestvo peredachi rechi i yego otsenka: uchebnoe posobie. Pod red. S N. Kirillova* [Speech quality and its evaluation: tutorial. Edited by S.N. Kirillov]. Orel: Academy of the FSO of Russia, 2015. 104 p. (In Russ.).
25. Zwicker E., Feldtkeller R. *Das Ohr als Nachrichtenempfänger*. Hirzel-Verlag. 1967. 232 p. (Russ. ed.: Zwicker E., Feldkeller R. *Ear as a receiver of information. 2 nd ed., rev. and add.* Moscow: Communications. 1971. 255 p.).