

УДК 621.395

АЛГОРИТМ КОМПЛЕКСНОЙ ОЦЕНКИ КАЧЕСТВА РЕЧИ В КАНАЛЕ СВЯЗИ

В. Т. Дмитриев, к.т.н., доцент кафедры РУС РГРТУ; vol77@rambler.ru
Д. С. Константинова, магистрант М РГРТУ, konstantinova-da@yandex.ru

Рассмотрена задача разработки алгоритма комплексной оценки качества речи на выходе канала связи при действии акустических шумов.

Целью работы является разработка алгоритма комплексной оценки качества речи на выходе канала связи, результаты которого максимально близки к результатам субъективной оценки, полученным согласно ГОСТ Р 50840-95. В процессе оценивания качества речи важнейшими критериями являются такие параметры, как разборчивость, узнаваемость, сохранение эмоциональной окраски голоса и другие. В современном мире очень важно передавать речь с наивысшим качеством для комфортного общения и ощущения эмоционального состояния говорящего. Для осуществления поставленной задачи рассмотрен ряд известных алгоритмов объективной оценки качества речевых сигналов и проанализированы результаты их работы. В результате из известных алгоритмов объективной оценки выбрано три, результаты которых оказались наиболее коррелированными с субъективной оценкой, полученной согласно ГОСТ Р 50840-95. На их основе разработан алгоритм комплексной оценки качества речевых сигналов, который позволит автоматизировать процесс оценивания характеристик на выходе систем передачи речи и обеспечит более точную оценку по сравнению с известными алгоритмами.

Ключевые слова: речевой сигнал, акустический шум, субъективная оценка качества, оконечное оборудование, системы передачи данных, алгоритм оценки качества речи.

DOI: 10.21667/1995-4565-2016-57-3-42-47

Введение

Наиболее востребованным видом передаваемой информации в телекоммуникационных системах всегда была и остается речевая информация [1]. Речь является наиболее распространенным способом передачи информации между людьми и содержит информацию не только в содержании, но и в едва уловимых оттенках, паузах, ударениях и прочих элементах, которые сложно описать в виде передаваемого сигнала.

Именно поэтому в системах передачи данных, в частности речи, особое внимание уделяют качеству. К качеству речи в данном случае относится разборчивость (объективная количественная мера, характеризующая способность тракта передавать содержащуюся в речевом сигнале (РС) смысловую информацию в данных конкретных акустических условиях окружающей среды), узнаваемость (отражает информацию о говорящем абоненте, т.е. его индивидуальных особенностях речи) и сохранность всех особенностей речи конкретного человека (эмоциональная окраска, тембр и т.п.).

Под мерой разборчивости понимается выраженное в процентах или долях единицы отно-

шение числа правильно принятых элементов речи (звуков, слогов, слов и фраз) к достаточно большому объему числа переданных. Разборчивость представляет собой семантическое содержание РС.

Качество передачи РС способствует сохранению индивидуальности речи каждого человека, что позволяет создать эффект присутствия и четкого проявления эмоциональной окраски голоса [2]. Вот почему оценка качества РС занимает одно из основных положений в вопросах обработки РС.

Одной из важнейших проблем при проектировании и эксплуатации систем кодирования РС, наряду с устранением избыточности передаваемой информации, является обеспечение высокого качества передачи РС (КПРС).

Современные системы передачи речевой информации обеспечивают при достаточной громкости до 89 % слоговой и до 100 % фразовой разборчивости РС, зачастую за счет потери узнаваемости. Поэтому к КПРС по каналам связи предъявляются требования как по разборчивости, так и по узнаваемости РС.

В настоящее время основные исследования

качества РС осуществляются согласно ГОСТ Р 50840-95 для субъективной оценки КПРС, которая учитывает оба эти критерия.

Субъективная оценка КПРС требует проведения большого количества тестовых испытаний, что приводит к значительным организационным и временным затратам. Кроме того, сильное влияние на результаты оказывают условия проведения испытаний, язык, настроение, возрастная группа экспертов и другие факторы [3]. Все это приводит к низкой повторяемости результатов и высокому разбросу субъективных оценок КПРС.

Оценки объективных алгоритмов КПРС получают с помощью технических средств, что обеспечивает лучшую повторяемость результатов по сравнению с алгоритмами субъективной оценки. Однако методы объективной оценки недостаточно учитывают влияние на РС многих искажений, заметных на слух, и не способны оценивать качество речи так, как это делает человек.

При этом методы объективной оценки КПРС, позволяющие оперативно определить качество речи на выходе декодеров, особенно при влиянии акустических шумов, в большинстве своем менее надежны, чем субъективные процедуры [4]. При этом отдельные критерии оценки качества речи существенно чувствительны к определенным видам шумов.

Таким образом, актуальными являются задача разработки и исследования алгоритмов объективной оценки и разработка алгоритма комплексной оценки качества РС при действии акустических шумов и коррелированной с субъективной оценкой, полученной согласно ГОСТ Р 50840-95.

Теоретическая часть

В телефонных и вещательных каналах мешающее действие шумов определяется наличием в их частотном спектре составляющих, которые наиболее сильно воздействуют на слух человека.

Известно, что не все частоты телефонного тракта одинаково воспроизводятся телефоном или громкоговорителем и воспринимаются ухом человека [5]. Поэтому при определении действия акустических шумов на речевые и музыкальные сигналы необходимо учитывать частотную зависимость чувствительности слуха и частотные характеристики телефона и микрофона.

Доказано, что наибольшая чувствительность системы «телефон-ухо» лежит в области 800 Гц. Это положение легло в основу оценки шумов в каналах связи и вещания, рекомендованной МККТТ [6].

Очень важно рассмотреть влияние акустических шумов, так как при изменении их мощности меняется и само отношение сигнал-шум.

Главным фактором, определяющим качество передачи РС в низкоскоростных цифровых системах связи, является ограниченная пропускная способность цифровых каналов систем связи [7]. Она задает способ преобразования и степень сжатия РС при формировании цифрового потока. Помимо этого при эксплуатации аппаратуры связи, часто возникают второстепенные мешающие факторы, дополнительно снижающие качество речевой связи, такие как:

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

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

Для объективной оценки качества РС могут быть использованы следующие критерии [8]:

- среднеквадратическая ошибка (СКО)

$$SKO = \frac{1}{N} \sum_{i=1}^N (x(i) - y(i))^2, \quad (1)$$

где $x(i)$ и $y(i)$ – i -е отсчеты исходного и синтезированного РС; N – общее количество отсчетов в исследуемом речевом фрагменте;

- критерий отношения сигнала-шум (ОСШ)

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

- анализ спектрально-корреляционных характеристик

$$L = \frac{1}{N_c} \sum_{i=1}^{N_c} \lambda(i) l(i), \quad (3)$$

где $\lambda(i)$ – мера спектральной неравномерности РС на i -м сегменте анализа; $l(i)$ – мера субполосных искажений РС на i -м сегменте анализа; N_c – число сегментов в речевом фрагменте;

- расстояние Итакура-Саито (ISD) и Кепстральное расстояние (CD)

$$ISD = \frac{1}{\pi} \int_0^{\pi} \left(\frac{P(\omega)}{\hat{P}(\omega)} - \log \frac{P(\omega)}{\hat{P}(\omega)} - 1 \right) d\omega; \quad (4)$$

- искажение спектра барков DSD (BarkSpectralDistortion)

$$BSD = \sum_{i=1}^N (S_x^{(k)}(i) - S_y^{(k)}(i))^2, \quad (5)$$

где $S_x^{(k)}$ и $S_y^{(k)}$ – спектры барков k -го сегмента исходного и синтезированного РС в i -й критической полосе; N – общее количество критических полос;

- модифицированные искажения спектра барков MBSD (ModifiedBarkSpectralDistortion)

$$MBSD = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^K M(n,i) D(n,i), \quad (6)$$

где $M(n,i)$ и $D(n,i)$ – i -е параметры искажений уровня ощущения и разницы уровней громкости n -го сегмента анализа; N – число сегментов в речевом фрагменте; K – общее количество критических полос [9];

- критерий LPC – спектральная разница огибающей между чистым входным сигналом и обработанным или поврежденным сигналом

$$d_{LLR}(\bar{a}_p, \bar{a}_c) = \log\left(\frac{\bar{a}_c R_c \bar{a}_p^T}{\bar{a}_p R_c \bar{a}_c^T}\right), \quad (7)$$

где a_c - LPC – вектор чистого РС, a_p - обработанный (расширенный) РС, R_c - автокорреляционная матрица исходного РС;

- взвешенный спектральный склон WSS

$$d_{WSS} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W(j,m) (S_c(j,m) - S_p(j,m))^2}{\sum_{j=1}^K W(j,m)}, \quad (8)$$

где $S_c(j,m), S_p(j,m)$ – спектральные склоны j -й полосы частот в рамках m чистых и переработанных РС соответственно;

- метод AI-ST делит сигнал на короткие (30 мс) сегменты данных, вычисляет AI для каждого сегмента и в среднем сегментарные значения AI над всеми кадрами:

$$AI - ST = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W(j,m) T(j,m)}{\sum_{j=1}^K W(j,m)}, \quad (9)$$

где M – общее число сегментов данных в сигнале, $W(j,m)$ – весовая функция (функция группы значений, размещенных на j -й полосе частот);

- квадрат величины когерентности MSC:

$$MSC(\omega) = \frac{\left| \sum_{m=1}^M X_m(\omega) Y_m^*(\omega) \right|^2}{\sum_{m=1}^M |X_m(\omega)|^2 \sum_{m=1}^M |Y_m(\omega)|^2}, \quad (10)$$

где * означает комплексное сопряжение,

$X_m(\omega)$ и $Y_m(\omega)$ – обозначены спектры БПФ сигналов $x(t)$ и $y(t)$ соответственно, вычисляемых в m -м сегменте данных.

Описание эксперимента

Объектом исследования в настоящей работе является оценка качества речи при смеси РС и акустических шумов с различным отношением сигнал/шум, которые подвергаются обработке в кодеках РС.

Исходные РС подвергались воздействию на них различных акустических шумов [10]. Данное наложение вносит искажения в РС в каждой из областей, которые его характеризуют: временной области, частотной области и области изменения спектральной динамики. В работе использовались такие акустические шумы, как звук стиральной машинки, чайника, поезда, двигателя, самолета и т.п.

Как известно [8], акустические шумы делятся на узкополосные, широкополосные и импульсные. К узкополосным шумам в работе относятся звук двигателя и самолета; к широкополосным шумам – звук чайника и стиральной машинки; к импульсным – звук поезда. Далее полученная смесь РС и акустического шума оценивалась с помощью выбранных критериев объективной оценки качества [10].

Для исследования выбраны наиболее часто используемые в настоящее время кодеки: LBRAMR 1, 1.2, 2, 2.4 кбит/с; ICELP 4.8, 6, 8 кбит/с; G.726 16, 24, 32, 40 кбит/с; G.722 48, 56, 64 кбит/с и другие (всего более 20 кодеков со скоростями передачи от 1 до 64 кбит/с). Все они используются в современных телекоммуникационных системах и обеспечивают различное качество восстановленного РС.

Эксперимент проведен по схеме, представленной на рисунке 1, где записанная на микрофоне M согласно рекомендациям ГОСТ Р 50840-95 фраза поступает на сумматор, где на РС накладывается АШ с определенным уровнем, чтобы обеспечить требуемое ОСШ (-10, -5, 0, 5, 10, 15, 20 дБ). Полученная смесь поступает на вход исследуемого кодера, где происходит кодирование смеси РС и АШ. Затем сигнал восстанавливается в декодере и поступает на динамик D для прослушивания бригадой auditors, а также на блок объективной оценки (ОБ).

Смесь РС и акустического шума, полученная после декодирования проходит оценку качества сначала по алгоритму субъективной оценки КПРС согласно ГОСТ Р50840-95, а затем оценку с помощью трех критериев объективной оценки качества по одному для каждой области изменения параметров РС в блоке объективной оценки.

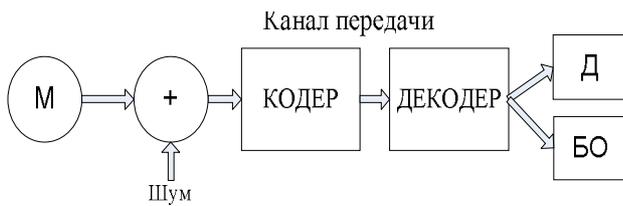


Рисунок 1 – Структурная схема получения субъективной и объективной оценки качества РС

Далее все результаты от трех выбранных критериев объективной оценки качества РС подвергаются нормировке и суммируются с заранее определенными коэффициентами для каждого вида акустического шума.

В итоге получаем комплексную оценку РС и результаты ее сравнения с алгоритмом субъективной оценки согласно ГОСТ Р 50840-95.

Комплексный алгоритм оценки качества РС

Полученная объективная комплексная оценка сравнивается с результатами субъективной оценки согласно ГОСТ Р 50840-95 и объективной оценкой, полученной с помощью алгоритма PESQ.

В качестве алгоритмов объективной оценки качества РС для разрабатываемого блока оценки качества РС выбраны наиболее распространенные и коррелированные с субъективными методами оценки критерии: SKO, MBSD и ESC.

Все полученные результаты от каждого из выбранных критериев имеют различный числовой порядок, поэтому для приведения их к единой пятибалльной оценке введены коэффициенты k_1, k_2, k_3 .

С помощью весовых коэффициентов α, β, γ появляется возможность, используя группу критериев объективной оценки качества РС, получить результаты, близкие к результатам субъективной оценки качества РС.

Таким образом, весь процесс оценивания становится автоматическим, достаточно быстрым и универсальным.

Используя три выбранных критерия оценки КЛРС: SKO, MBSD, ESC и коэффициенты α, β, γ , единую оценку КЛРС можно получить по формуле:

$$K = \alpha SKO k_1 + \beta MBSD k_2 + \gamma ESC k_3. \quad (11)$$

Весовые коэффициенты α, β, γ при проведении экспериментальных исследований изменялись в пределах $[0, 1 \dots 1]$ с шагом 0,1. При этом необходимо выполнение условия нормировки коэффициентов:

$$\alpha + \beta + \gamma = 1. \quad (12)$$

Универсальный блок оценки РС работает по схеме, изображенной на рисунке 2.

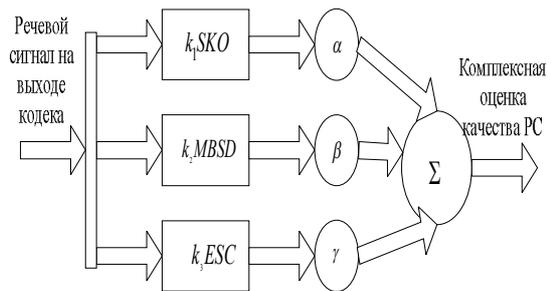


Рисунок 2 – Структурная схема блока оценки КЛРС

Результаты экспериментальных исследований

Результаты, полученные блоком оценки по предложенному алгоритму, имеют высокий коэффициент корреляции порядка 0,9 и выше с результатами алгоритма ГОСТ Р 50840-95 по сравнению с известными методами оценки КЛРС.

На рисунке 3 показаны графики разности оценок между ГОСТ Р 50840-95, PESQ и выбранными алгоритмами объективной оценки КЛРС, где $\Delta_1 = |ГОСТ - X|$, а $\Delta_2 = |ГОСТ - PESQ|$, где X – оценка качества согласно выбранному критерию объективной оценки (SKO – рисунок 3, а; ESC – рисунок 3, б; MBSD – рисунок 3, в).

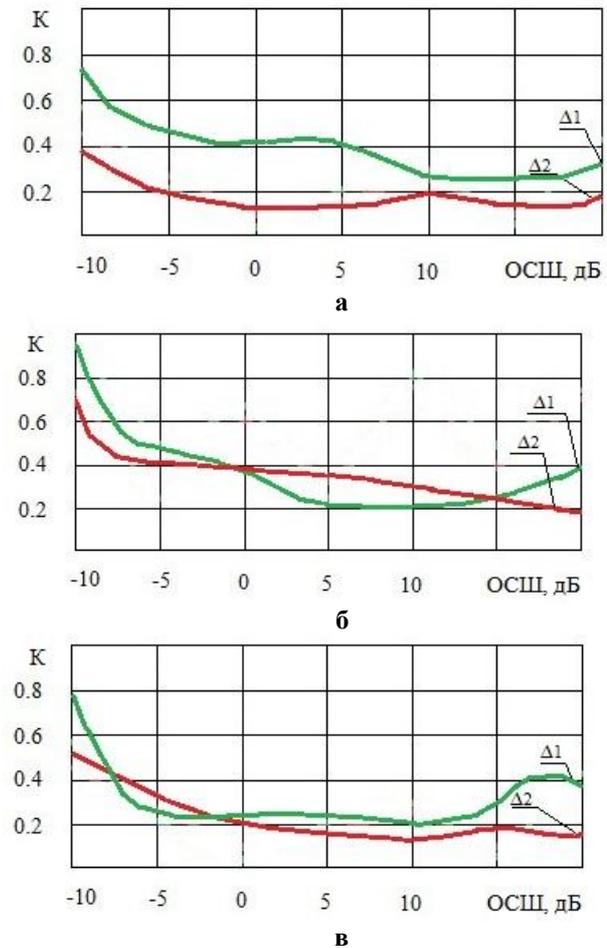


Рисунок 3 – График разности оценок между выбранным алгоритмом объективной оценки, алгоритмом ГОСТ Р 50840-95 и PESQ

С помощью выбранных алгоритмов оценки КПРС получена комплексная оценка смеси РС и акустического шума.

На рисунке 4 представлена разница между результатами комплексной оценки и результатами алгоритмов ГОСТ Р 50840-95 и PESQ.

Таким образом, как следует из анализа зависимостей, оценка качества смеси РС и акустического шума каждым из алгоритмов объективной оценки обеспечивает результаты, недостаточно коррелированные с субъективной оценкой.

Показано, что применение комплексного алгоритма обеспечивает более точную оценку по сравнению с субъективной оценкой по ГОСТ Р 50840-95, которая не превышает 0,5 баллов.

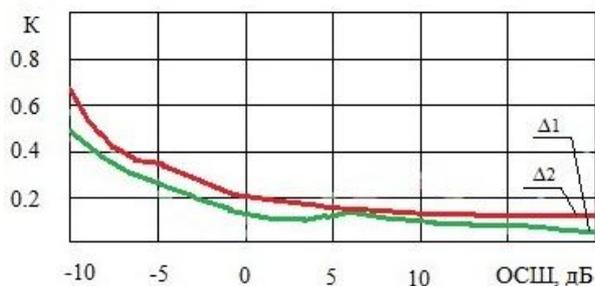


Рисунок 4 – График разности оценок между комплексным алгоритмом, PESQ и ГОСТ Р 50840-95

Заключение

В работе предложен один из вариантов комплексного алгоритма оценки качества РС на выходе канала связи при действии акустических шумов, который позволит в автоматическом режиме производить субъективную оценку качества РС и получить разность оценки с ГОСТ Р 50840-95 не более 0,5 баллов, что меньше, чем в случае использования одного из распространенных алгоритмов объективной оценки качества PESQ.

Работа поддержана грантом Российского фонда фундаментальных исследований (номер договора №16-47-620544\16 от 11.05.2016) и Правительством Рязанской области.

Библиографический список

1. 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.
2. Кириллов С. Н., Дмитриев В. Т., Картавенко Я. О., Ежов С. А., Ватугин В. М. Оценка качества речевых сигналов в спутниковых каналах связи // Труды IV Всероссийской научно-технической конференции «Актуальные проблемы ракетно-космического приборостроения и информационных технологий». М., 2012. - С.275 – 281.
3. ГОСТ Р 50840-95. Передача речи по трактам связи. - М.: Госстандарт России, 1995. – С 180.
4. Цыбулин М. К., Бочаров М. О. Анализ методов оценки качества передачи речевой информации по каналам связи различной структуры // Электро-связь. – 2008. – № 11. – С. 46-48.
5. Ромашкин Ю. Н., Кириллов С. 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.
6. Корячко В. П., Шибанов В. А., Ижванов Ю. Л., Шибанов А. П. Оценка времени передачи файла с учетом старения информации // Информационные технологии. 2010. – № 10. – С. 40-45.
7. Поляков А. Н. Об одном из способов решения задачи определения оптимальных управляющих параметров системы низкоскоростной компрессии речевой информации // Телекоммуникации. – 2008. – № 3. – С. 15-18.
8. Шелухин О. И., Лукьянцев Н. Ф. Цифровая обработка и передача речи. – М.: Радио и связь, 2000. – С. 456.
9. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации. – 2-е изд., перераб. и доп. – М.: Связь, 1971. – С 255.
10. Кириллов С. Н., Дмитриев В. Т., Янак А. Ф. Изучение влияния акустических шумов на первичные кодеки речевых сигналов // Современные технологии в науке и образовании СТНО-2016: Сб. трудов международной научно-технической и научно-методической конференции в 4 т. Т.1 / под. общ. ред. О.В. Миловзорова. – Рязань: Рязан. гос. Радио-техн. ун-т, 2016. - С. 121-124.

UDC 621. 395

ALGORITHM OF SPEECH QUALITY INTEGRATED ASSESSMENT IN THE COMMUNICATION CHANNEL

V. T. Dmitriev, Ph.D., assistant professor of RUS, RSREU; vol77@rambler.ru

D. S. Konstantinova, master student, gr.419, RSREU, konstantinova-da@yandex.ru

The task of developing an algorithm of integrated assessment of speech quality at the output of communication channel by the action of acoustic noise is considered.

The aim is to develop an algorithm of complex estimation of speech quality at the output of communication channel, the results of which as much as possible are close to the results of subjective evaluation, prepared according to GOST 50840-95. In the process of evaluating the quality of speech such parameters as clarity, awareness, preservation of emotional color to vote are the most important criteria. In modern world it is very important to transmit the highest quality for comfortable communication and feelings of speaker's emotional state. As a result from all known algorithms of objective evaluation three algorithms were chosen, the results of which turned out to be mostly correlating with subjective evaluation prepared in accordance with GOST 50840-95. On their basis the algorithm of complex evaluation speech signals quality which will allow to automate the process of evaluating the characteristics of output speech transmission systems and to provide more accurate evaluation compared to known algorithms.

Keywords: *speech, acoustic noise, subjective evaluation of quality of terminal equipment, data transmission systems, voice quality assessment algorithm.*

DOI: 10.21667/1995-4565-2016-57-3-42-47

References

1. **Kirillov S. N., Dmitriev V. T., Kartavenko Y. O.** Objective quality assessment algorithm decoded speech signal based on changes in the spectral dynamics critical bands of the spectrum Bulletin of RSREU 2011 no.3 (37) pp.3-7.
2. **Kirillov S. N., Dmitriev V. T., Kartavenko Y. O., Ezhov S. A., Vatutin V. M.** Evaluation of the quality of voice signals in satellite communication channels. Proceedings of the IV All-Russian scientific-technical conference "Actual problems of rocket and Space Instrumentation and Information Technology" M.2012. pp. 275 - 281.
3. GOST 50840-95. Voice over communication paths. M.: State Standard of Russia, 1995. - 180 p.
4. **Tsybulin M. K., Bocharov M. O.** Analysis of the quality of assessment methods for the transmission of voice information communication channels with different structures. Telecommunications. - 2008. - no.11. - pp. 46-48
5. **Romachkin Y. N., Kirillov S. N., Kartavenko Ya. O., Dmitriev V. T.** Analytical evaluation of speech quality at the output of low-rate coding systems under the influence of acoustic noise. Speech Technology 2012 no.4. FROM. - pp. 16-23.
6. **Koryachko V. P., Shibanov V. A., Izhvanov Y. L., Shibanov A. P.** Evaluation file transmission time based on the aging of information. Information technologies. 2010. no. 10. - pp. 40-45.
7. **Polyakov A. N.** On one of the ways of solving the problem of determining optimal control parameters of the system low-rate voice data compression. Telecommunications. - 2008. - no.3. - pp. 15-18.
8. **Shelukhin O. I., Lukyantsev N. F.** Digital processing and transmission of speech. - M.: Radio and Communications, 2000. - 456 p.
9. **Zwicker E., Feldkeller R.** Ear as the receiver of information - 2 nd ed., Revised. & Accessories - M.: Communications, 1971. - 255 p.
10. **Kirillov S. N., Dmitriev V. T., Yanak A. F.** Study of the effect of acoustic noise in speech codecs primary signals. Modern technologies in science and education STNO 2016. Proceedings of the international scientific-technical and scientific-methodical conference in 4 volumes. Volume 1 under. Society. Ed. OV Milovzorova. - Ryazan: Ryazan. State. Radi. University Press, 2016 – pp. 121-124.