

УДК 621.391

А.Ю. СТРЮК, Я.В. ЯНСОНС

Полтавський військовий інститут зв'язи, Україна

МЕТОДИКА ІССЛЕДОВАНИЯ ПОМЕХОУСТОЙЧИВОСТИ БИТ ПАРАМЕТРОВ НИЗКОСКОРОСТНЫХ КОДЕКОВ РЕЧИ

Представлена методика исследования помехоустойчивости бит параметров сжатого сигнала кодеков речи с использованием объективной оценки качества имеющей максимальный коэффициент корреляции с субъективными оценками. Приведены результаты применения методики для низкоскоростного кодека MELP на скорости 2.4 Кбит/с.

помехоустойчивость, сжатие речи, кодирование речи, объективная оценка качества речи, коротковолновые, ультракоротковолновые радиосети

Введение

Одной из основных проблем систем низкоскоростной передачи речи, к которым относятся преимущественно системы специальной и диспетчерско-поездной радиосвязи, является компромисс между качеством и скоростью передачи речевого потока.

Качество речи в свою очередь зависит от уровня помехоустойчивости передаваемого речевого сигнала. Классически задача повышения помехоустойчивости заключается в применении определенного помехоустойчивого кода ко всей совокупности передаваемых данных [1].

Одним из методов дальнейшего повышения помехоустойчивости, при передаче речи, является деление выходных данных с кодера на группы для последующего применения различных по характеристикам помехоустойчивых кодов (с целью снижения общей избыточности). Такое деление основано на субъективной оценке бит параметров передаваемого сигнала. Предлагаемая методика предназначена для автоматизации указанного деления на основе выбранного объективного показателя.

Ограничения методики

Ограничением методики является использование предложенной модели помех [2] в канале связи, которая позволяет акцентировать внимание на рассматриваемом бите параметров передаваемого рече-

вого сигнала. Канал связи – дискретный симметричный без памяти (ДСБП).

Математический аппарат

Оцифрованный речевой сигнал, множество $X = \{x_1, x_2, \dots, x_e\}$, является множеством своих отсчетов x_e . После применения кодирования (кодера источника) речевой сигнал преобразуется в множество $Z = \{A^1, A^2, \dots, A^b\}$ параметров речевого сигнала, $b > 0$, $b \ll e$, которые могут быть как одноэлементными $A^{b'} = \{A_1^{b'}\}$, $b', b'', b''' \in b$, $b', b'', b''' \leq b$ (параметр синхронизации, параметр аперидичности и т.д.), так и многоэлементными $A^k = \{A_1^k, A_2^k, \dots, A_l^k\}$ (линейные спектральные частоты, коэффициенты усиления и т.д.), $0 < k < b$, $l > 0$. Каждый параметр речевого сигнала представляется подмножеством бит $A_1^l = \{a_{1,1}^l, a_{1,2}^l, \dots, a_{1,d}^l\}$, $d > 0$.

Все множество бит параметров может быть разделено на группы $H = \{h_1, h_2, \dots, h_i\}$, $h_i \subset H$, по степени влияния на качество восстановленного сигнала [3], $h_i = \left\{ a_{q,d}^{b'}, a_{q,d}^{b''}, \dots, a_{q,d}^{b'''} \right\}$ – определенная группа бит i характеризующаяся приблизительно одинаковой значимостью, $q', q'', q''' \in q$, $q', q'', q''' \leq q$, $d', d'', d''' \in d$, $d', d'', d''' \leq d$.

Для разделения бит на группы необходимо объективно оценить значимость каждого бита параметров речевого сигнала. Получение объективной оценки M в значениях MOS для каждого бита состоит из трех этапов:

1. Получение множеств $Z'_i = \{A^{1'}, A^{2'}, \dots, A^{b' i}\}$, $i = \overline{1, N}$, N – количество бит в одном кадре речевого сигнала. Множества Z'_i выражаются из множества Z путем наложения на него всей совокупности помех (имитация канала связи). Множество $Z' = \{A^{1'}, A^{2'}, \dots, A^{b' i}\}$ – множество параметров, полученное в результате кодирования и декодирования множества Z без помех в канале.

Имитация канала связи заключается в расчете множеств Z'_i , в которых i -й бит $\bar{a}_{q',d'}^{b' i}$ определяется из выражения:

$$\bar{a}_{q',d'}^{b' i} = \text{xor} \left(a_{q',d'}^{b' i} \right). \quad (1)$$

2. Декодирование множеств Z' и Z'_i с использованием декодера речи определенного типа, для получения итоговых комбинаций восстановленной речи $X' = \{x'_1, x'_2, \dots, x'_e\}$ и $X'_i = \{x'_{i1}, x'_{i2}, \dots, x'_{ie}\}$, прошедших канал связи.

3. Используя реализацию рекомендации ITU-T P.862 PESQ (Perceptual evaluation of speech quality – оценка восприятия качества речи) [3], вычислить множество значений качества речи

$$M = \left\{ M^1_{a_{q',d'}^{b' i}}, M^2_{a_{q',d'}^{b' i}}, \dots, M^i_{a_{q',d'}^{b' i}} \right\}$$

для каждого из множеств X'_i путем попарного сравнения исходного и итоговых множеств X и X'_i . А также значения M_{\max} – качество речи рассматриваемого кодека, для множеств X и X' .

Далее необходимо расположить все значения $M^i_{a_{q',d'}^{b' i}}$ в возрастающем порядке и осуществить преобразование $M^i_{a_{q',d'}^{b' i}} \rightarrow Mp_j$, получим множество

$Mp = \{Mp_1, Mp_2, \dots, Mp_j\}$ отсортированных по воз-

растанию значений качества речи, так что, $Mp_1 < Mp_2 < \dots < Mp_j$, где $j = \overline{1, N}$ и $i \neq j$.

Нормированные значения y_j для битов $a_{q',d'}^{b' i}$:

$$y_j = \frac{Mp_j}{M_{\max}} \quad (2)$$

образуют множество $y = \{y_1, y_2, \dots, y_j\}$ нормированных значений. Вычисление границ условного разделения нормированных значений битов параметров речевого сигнала:

$$\begin{aligned} \mu_1 &= s \times y_{\max}; \\ \mu_2 &= s^2 + \mu_1; \\ &\vdots \\ \mu_k &= s^2 + \mu_{k-1}; \\ &\vdots \\ \mu_i &= s^2 + \mu_{i-1}, \end{aligned} \quad (3)$$

где $i = \overline{1, n}$, $k \leq n$, n – количество битовых групп; s – коэффициент условного разделения битовых групп, рассчитывается экспериментально.

Соответствие битовых групп битовым областям имеет следующий вид:

$$V = \begin{cases} a_{q,d}^b \in v_1, \text{ при } y_j < \mu_1, \text{ где } j \leftrightarrow a_{q,d}^b; \\ a_{q',d'}^{b'} \in v_2, \text{ при } y_j < \mu_2, \text{ где } j \leftrightarrow a_{q',d'}^{b'}; \\ \vdots \\ a_{q'',d''}^{b''} \in v_k, \text{ при } \mu_{k-1} \leq y_j \leq \mu_k, \text{ где } j \leftrightarrow a_{q'',d''}^{b''}; \\ \vdots \\ a_{q''',d'''}^{b''' } \in v_n, \text{ при } y_j > \mu_n, \text{ где } j \leftrightarrow a_{q''',d'''}^{b''' } \end{cases} \quad (4)$$

где $V \supseteq H$, $v_i \supseteq h_i$, $v_1, v_2, \dots, v_n \in V$, v_n – битовые области? параметр n – количество битовых областей (на практике $n_{\max} = 2 \div 4$) и соответственно битовых групп определяется следующим выражением:

$$V_{\max} > ((v_1 + \Delta V_1) + (v_2 + \Delta V_2) + \dots + (v_n + \Delta V_n)), \quad (5)$$

где V_{\max} – максимально допустимое количество бит для определенного значения пропускной способности канала связи; ΔV_n – избыточность вносимая тем или иным кодом:

$$\Delta V_n = f(R_m), \quad (6)$$

где R_m – вид помехоустойчивого кодирования.

Входные и выходные данные

Таким образом, в соответствии с определенным математическим аппаратом и поставленными задачами, входными данными методики должны быть: K_{fr} – количество кадров речевого сигнала; выборка оцифрованных речевых сигналов X с характеристиками определенными тем или иным кодеком; количество бит в одном кадре речи N ; коэффициент условного разделения битовых групп s – определяется экспериментально; количество областей группирования n_{max} .

Выходными данными являются: границы условного разделения битовых групп μ_i ; распределение бит параметров речевого сигнала $a_{q',d}^{b'}$ по областям группирования.

Алгоритм реализации

Алгоритм реализации методики (рис. 1) состоит из следующих действий:

1. Задать K_{fr} – объем выборки (кадров речевого сигнала), множество X оцифрованного сигнала, количество бит в речевом кадре N , коэффициент условного разделения битовых групп s , количество областей группирования n_{max} .
2. Используя кодер речи вычислить множество Z параметров речевого сигнала.
3. Рассчитать множество Z' . Рассчитать N множеств Z'_i , используя имитацию помех в канале связи.
4. Вычислить значения качества речи $M_{i,fr}$ и $M_{max,fr}$ для количества кадров K_{fr} .
5. Вычислить средние арифметические значения качества речи $M_{i,av}$ и $M_{max,av}$.
6. Преобразовать значения $M_{i,av}$ в Mp_j .
7. Рассчитать нормированные значения y_j .
8. Вычислить значения границ условного разделения битовых групп μ_i , и области группирования бит V .
9. Определить битовые группы H .

Пример использования методики

Рассмотрим использование методики на примере кодека речи типа MELP (Mixed Excitation Linear Prediction – линейное предсказание со смешанным возбуждением) на скорости 2,4 Кбит/с.

Тестовыми речевыми сигналами для методики выбраны речевые базы ARCTIC. Выборка K_{fr} составляет 50 тысяч кадров англоязычной речи (мужской, женской, детской).

Данные кадры перекодируются со следующими характеристиками отсчетов, с учетом требований входного сигнала для кодека MELP [5] – 8 кГц, 16 бит/отсчет – множество X .

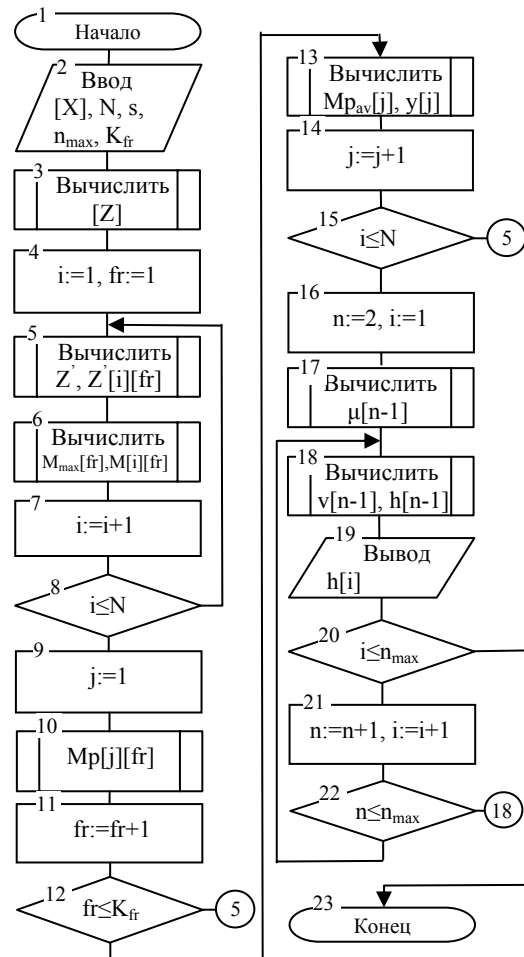


Рис. 1. Алгоритм распределения бит параметров по областям группирования

Используя реализацию кодера MELP (Texas Instruments, Inc., 1996) на языке программирования C++ вычислим множество параметров речевого сигнала Z табл. 1 [5].

Рассматриваемый кодер различает огласованные и неогласованные кадры. В неогласованных кадрах часть бит защищена кодом Хаффмана в пределах 54 бит за счет освободившихся 13 бит при передаче в сравнении с огласованными [5, 6]. Рассмотрение помехоустойчивости бит представляющих код Хафф-

мана являється нецелесообразным. Имитируя канал связи, сначала без ошибок, далее с ошибками вычисляются множества $Z'[fr]$ и $Z_i'[fr]$ для всей выборки речевого сигнала, где $i = \overline{1,54}$, $fr = \overline{1,50000}$.

Таблица 2

Значения качества речи для бит параметров кодека MELP 2,4 Кбит/с

Биты параметров	$M_{i,av}$	y_j
LSF(1)-1	1,526443	0,663671
LSF(1)-2	1,411112	0,613527
LSF(1)-3	1,352556	0,588068
LSF(1)-4	1,337278	0,581425
LSF(1)-5	1,269168	0,551812
LSF(1)-6	1,14911	0,499613
LSF(1)-7	1,019611	0,443309
LSF(2)-1	1,707667	0,742464
LSF(2)-2	1,647943	0,716497
LSF(2)-3	1,603889	0,697343
LSF(2)-4	1,516277	0,659251
LSF(2)-5	1,500557	0,652416
LSF(2)-6	1,488056	0,646981
LSF(3)-1	1,833056	0,796981
LSF(3)-2	1,820721	0,791618
LSF(3)-3	1,756223	0,763575
LSF(3)-4	1,751057	0,761329
LSF(3)-5	1,737222	0,755314
LSF(3)-6	1,624389	0,706256
LSF(4)-1	1,865611	0,811135
LSF(4)-2	1,843611	0,80157
LSF(4)-3	1,772888	0,770821
LSF(4)-4	1,761667	0,765942
LSF(4)-5	1,760834	0,76558
LSF(4)-6	1,75939	0,764952
FM – 1	1,900221	0,826183
FM – 2	1,853223	0,805749
FM – 3	1,852701	0,805522
FM – 4	1,801167	0,783116
FM – 5	1,800445	0,782802
FM – 6	1,799888	0,78256
FM – 7	1,798722	0,782053
FM – 8	1,788167	0,777464
G(1) – 1	1,873166	0,81442
G(1) – 2	1,832944	0,796932
G(1) – 3	1,785778	0,776425
G(2) – 1	1,7229	0,749087
G(2) – 2	1,569168	0,682247
G(2) – 3	1,482334	0,644493
G(2) – 4	1,047666	0,455507
G(2) – 5	0,694722	0,302053
P – 1	1,67127	0,726639
P – 2	1,525167	0,663116
P – 3	1,401112	0,609179
P – 4	1,391889	0,605169
P – 5	1,367389	0,594517
P – 6	1,341779	0,583382
P – 7	1,292722	0,562053
BP – 1	1,820289	0,79143
BP – 2	1,647778	0,716425
BP – 3	1,571167	0,683116
BP – 4	1,571089	0,683082
AF	1,548445	0,673237
SYNC	1,72339	0,7493
	$M_{\max, av} = 2,3$	

Таблица 1

Распределение бит в кадре кодека MELP

Параметр	Количество бит огласованной речи	Количество бит неогласованной речи
LSF(1)	7	7
LSF(2)	6	6
LSF(3)	6	6
LSF(4)	6	6
FM – Амплитуды Фурье	8	-
G(1) – Уровень речи	4	4
G(2) – Уровень речи	4	4
P – Основной тон и тон/шум	7	7
BP – Огласованность в полосах	4	-
AF – Аперриодичность	1	-
SYNC – Синхронизация	1	1
Общее число бит	54	41

Применяя реализацию декодера MELP 2,4 Кбит/с, определяем множества $X'[fr]$ и $X_i'[fr]$. Расчет значений качества речи $M_{\max}[fr]$ и $M_i[fr]$ для каждого бита производится с использованием реализации действующей рекомендации МСЕ-Т Р.862 на языке программирования С++.

Рассчитав среднеарифметические значения $M_{\max,av}$ и $M_{i,av}$ (табл. 2) расположим $M_{i,av}$ в возрастающем порядке (табл. 3 и 4) – множество Mp_j где $j = \overline{1,54}$ так, что $Mp_1 < Mp_2 < \dots < Mp_j$.

Для табл. 2–4, параметры $X-1, X-2, \dots, X-n$ определяют следующее:

- X – параметр кодера;
- $1, 2, \dots, n$ – номер бита параметра, где 1 – наименее значащий бит.

Необходимо отметить, что 24 наиболее важных бита, с точки зрения восстановления речевого сигнала, параметров кодека MELP [5, 6] размещаются в начале полученной последовательности (табл. 3 и 4) с использованием объективного показателя, что подтверждает достоверность исследований и полученных результатов.

Таблица 3

Отсортированные значения бит параметров кодека MELP 2,4 Кбит/с на основе объективного показателя для огласованного кадра

№ п/п	Биты параметров	№ п/п	Биты параметров	№ п/п	Биты параметров
1	G(2)-5	19	LSF(1)-1	37	LSF(4)-4
2	LSF(1)-7	20	AF	38	LSF(4)-3
3	G(2)-4	21	G(2)-2	39	G(1)-3
4	LSF(1)-6	22	BP-4	40	FM-8
5	LSF(1)-5	23	BP-3	41	FM-7
6	P-7	24	LSF(2)-3	42	FM-6
7	LSF(1)-4	25	LSF(3)-6	43	FM-5
8	P-6	26	BP-2	44	FM-4
9	LSF(1)-3	27	LSF(2)-2	45	BP-1
10	P-5	28	P-1	46	LSF(3)-2
11	P-4	29	LSF(2)-1	47	G(1)-2
12	P-3	30	G(2)-1	48	LSF(3)-1
13	LSF(1)-2	31	SYNC	49	LSF(4)-2
14	G(2)-3	32	LSF(3)-5	50	FM-3
15	LSF(2)-6	33	LSF(3)-4	51	FM-2
16	LSF(2)-5	34	LSF(3)-3	52	LSF(4)-1
17	LSF(2)-4	35	LSF(4)-6	53	G(1)-1
18	P-2	36	LSF(4)-5	54	FM-1

Таблица 4

Отсортированные значения бит параметров кодека MELP 2,4 Кбит/с на основе объективного показателя для неогласованного кадра

№ п/п	Биты параметров	№ п/п	Биты параметров	№ п/п	Биты параметров
1	G(2) -5	15	LSF(2) -6	29	LSF(3)-4
2	LSF(1) -7	16	LSF(2) -5	30	LSF(3)-3
3	G(2) -4	17	LSF(2) -4	31	LSF(4) -6
4	LSF(1) -6	18	P-2	32	LSF(4) -5
5	LSF(1) -5	19	LSF(1) -1	33	LSF(4)-4
6	P-7	20	G(2) -2	34	LSF(4)-3
7	LSF(1) -4	21	LSF(2)-3	35	G(1)-3
8	P-6	22	LSF(3) -6	36	LSF(3)-2
9	LSF(1) -3	23	LSF(2)-2	37	G(1) -2
10	P-5	24	P-1	38	LSF(3) -1
11	P -4	25	LSF(2) -1	39	LSF(4)-2
12	P-3	26	G(2)-1	40	LSF(4) -1
13	LSF(1) -2	27	SYNC	41	G(1)-1
14	G(2) -3	28	LSF(3) -5		

Таблица 5

Распределение бит параметров по областям группирования

Область группирования бит	Количество бит
v_1	4
v_2	27
v_3	23

Заключение

Использование предложенной методики позволит снизить канальную скорость передачи речи при сохранении качества речи в допустимых пределах или же повысить помехоустойчивость речевых сигналов при неизменной скорости передачи, за счет решения задачи объективного распределения бит параметров речевого сигнала по областям группирования в зависимости от степени влияния на качество восстановленного сигнала.

Литература

1. Маркел Дж.Д., Грей А.Х. Линейное предсказание речи. – М.: Связь, 1980. – 308 с.
2. Стрюк А.Ю., Янсонс Я.В. Методика оценки помехоустойчивости параметров сжатого сигнала кодеков речи // Радиоэлектронные и компьютерные системы. – 2006. – № 5. – С. 198.
3. ITU-T P.862 (2001). Methods for objective and subjective assessment of quality (Perceptual evaluation of speech quality (PESQ)).
4. Chu, Wai C. Speech coding algorithms: Foundation and evolution of standardized coders. – John Wiley & Sons, Inc., Hoboken, New Jersey, 2003. – 558 p.
5. McCree A., Truong K., George E.B., Barnwell T.P., Viswanathan V. A 2,4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard // Proceedings of ICASSP, IEEE. – 1996. – P. 200-203.
6. Supplee L., Cohn R., Collura J., McGree A. A 2,4 kbit/s MELP- New U.S. Federal Standard // Proceedings of ICASSP, IEEE. – 1997. – P. 1591-1594.

Поступила в редакцию 5.02.2006

Рецензент: д-р техн. наук, проф. В.М. Илюшко, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.