

ВЕКТОРНОЕ КВАНТОВАНИЕ КОЭФФИЦИЕНТОВ УСИЛЕНИЯ В CELP-КОДЕРЕ РЕЧИ

Дыранов Ю. В.^{*}, Костров В. В., Власов С. Ю.

^{*}Муромский завод радиоизмерительных приборов
602200, г. Муром Владимирской обл., Карачаровское шоссе, 2
Тел. (09234) 3–34–36, Факс (09234) 2–16–16
Муромский институт (филиал) Владимирского государственного университета
602200, г. Муром Владимирской обл., ул. Орловская, 23, кафедра «Радиотехника»
Тел. (09234) 3–54–19 Факс (09234) 2–28–85

Реферат. Представлены результаты экспериментального исследования CELP-кодеров с векторным квантованием коэффициентов усиления. Обсуждаются результаты оценки качества речи при использовании Диагностического Критерия Приемлемости для кодовых книг размером 64 и 128 4-мерных векторов, полученных по обучающей выборке, сформированной на основе 8-минутного речевого материала, включающего в себя речь от 12 дикторов (5 женщин и 7 мужчин).

Разработанные за последние двадцать лет методы кодирования обеспечивают хорошее качество (разборчивость, натуральность звучания, повышенную возможность опознавания говорящего) при передаче речи в цифровой форме по узкополосным каналам связи на скоростях от 16 до 4,8 кбит/с. На практике нашли широкое применение кодеры с линейным предсказанием при многоимпульсном возбуждении (Multi Pulse Linear Predictive Coding — MPLPC) и при возбуждении от кода (Code Excited Linear Prediction — CELP) [1].

Наиболее совершенным алгоритмом (с точки зрения качества) является алгоритм CELP-кодера на 4,8 кбит/с (рис.1), принятый в качестве федерального стандарта США FS-1016 [2].

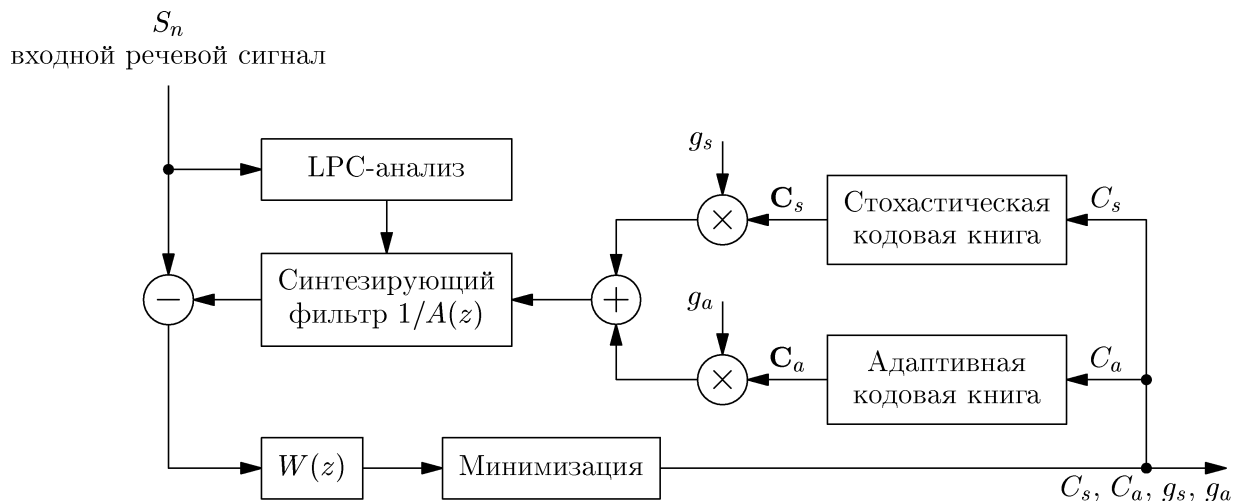


Рисунок 1— Структурная схема кодирования CELP-кодера FS-1016

В CELP-кодере FS-1016 речевой сигнал S_n разделяется на кадры длительностью в 30 мс. В каждом кадре с использованием алгоритма линейного предсказания (LPC) определяются параметры синтезирующего фильтра $1/A(z)$, после чего методом анализа через синтез находятся параметры сигнала возбуждения, минимизирующие взвешенный сигнал ошибки. Сигнал возбуждения представляется наборами индексов (C_s, C_a) векторов (C_s, C_a) , извлекаемых из стохастической и адаптивной кодовых книг, а также наборами соответствующих им коэффициентов усиления (g_s, g_a) . При кодировании сигнала возбуждения кадр разбивается на 4 подкадра по 7,5 мс. В каждом подкадре кодируются и передаются индексы C_s (9 бит на индекс), C_a (8 бит в четных и 6 бит в нечетных подкадрах на индекс), коэффициенты усиления g_s и g_a (по 5 бит на каждый коэффициент). В целом кадр кодируется 144 битами, из которых 40 бит (28,8%) отводятся на кодирование коэффициентов усиления с использованием скалярного квантования.

Целью настоящей работы является экспериментальное исследование CELP-кодера с векторным квантованием коэффициентов усиления. В частности, обсуждаются результаты применения векторного квантования при раздельном кодировании коэффициентов усиления g_s, g_a .

При использовании векторного квантования для каждого из коэффициентов усиления g_s и g_a производилось объединение четырех значений, полученных для подкадров одного кадра, в один четырехмерный вектор. В результате этого для каждого кадра формировались два вектора коэффициентов усиления: $\{g_{s_1}, g_{s_2}, g_{s_3}, g_{s_4}\} = \mathbf{g}_s$ и $\{g_{a_1}, g_{a_2}, g_{a_3}, g_{a_4}\} = \mathbf{g}_a$, для квантования которых использовались раздельные кодовые книги. Формирование кодовых книг выполнялось на основе обучающей выборки размером 16000 векторов, с использованием которой для каждого из векторов \mathbf{g}_s и \mathbf{g}_a были построены по две кодовые книги размером 64 и 128 эталонных векторов (длина кодового слова 6 и 7 бит соответственно). При таких размерах кодовых книг количество бит, отводимых на кодирование коэффициентов усиления, сокращается соответственно на 28 и 26 бит на кадр по сравнению со стандартным алгоритмом FS-1016.

Обучающая выборка формировалась в результате обработки CELP-кодером FS-1016 речевого материала от двенадцати дикторов (5 женщин и 7 мужчин) общей продолжительностью 8 минут. Для построения кодовых книг использовался алгоритм К-средних с начальными условиями, полученными методом кластеризации при неравномерной дихотомии [3]. Определение качества звучания производилось с использованием Диагностической Меры Приемлемости (Diagnostic Acceptability Measure — DAM) путем прослушивания 12 фонетически сбалансированных 6-словых предложений, произносимых дикторами (мужчина и женщина), не участвовавшими в формировании обучающей выборки. Качество звучания оценивалось бригадой из десяти слушателей. По результатам оценки вычислялась средняя оценка мнений (процент предпочтений).

Результаты испытания кодеров с векторным квантованием коэффициентов усиления при различных размерах кодовых книг сведены в таблицу 1.

Таблица 1 — Результаты испытания кодеров

Предпочтения	Кодовые книги размером 64 вектора		Кодовые книги размером 128 векторов	
	Кол-во голосов	Процент предпочтения	Кол-во голосов	Процент предпочтения
Предпочитают CELP-кодер FS-1016	3	30%	2	20%
Не видят различий	7	70%	7	70%
Предпочитают модернизированный кодер	0	0%	1	10%
Итого	10	100%	10	100%

Анализируя полученные результаты, можно отметить, что при оценке качества по методу DAM большинство экспертов не находят различий в качестве звучания CELP-кодера FS-1016 и модернизированного кодера, использующего векторное квантование коэффициентов усиления (70–80% экспертов не видят различий в качестве звучания, и только 20–30% отдают предпочтение CELP-кодеру FS-1016).

С точки зрения размера кодовой книги незначительные преимущества имеет кодовая книга большей размерности (80% предпочтений отданы кодовой книге размером 128 векторов и 70% — кодовой книге размером в 64 вектора).

Таким образом, использование векторного квантования коэффициента усиления позволяет без ущерба качеству звучания понизить скорость в CELP-кодере с 4,8 кбит/с до 3,87–3,93 кбит/с (за счет сокращения на 28–26 бит расходов на кодирование сигнала возбуждения на каждом кадре длительностью 30 мс).

Библиография

1. Trancoso I. M., Marques J. S. *Speech Communication*, 1990, v. 9, № 5.
2. Federal Standard 1016. *Telecommunications: Analog to Digital Conversion of Radio Voice by 4,800 bit/second Code Excited Linear Prediction (CELP)*. February 14, 1991.
3. Дж. Макхоул. Векторное квантование при кодировании речи. ТИИЭР, т. 73, № 11, ноябрь 1985.

VECTOR QUANTIZATION OF GAIN COEFFICIENTS IN THE CELP VOCODER

Dyranov Yu. V. *, Kostrov V. V., Vlasov S. Yu.

*Murom radio-measuring instruments plant
2 Karacharovskoye Highway, Murom, Vladimir region, 602200, Russia
Murom Institute (branch) of the Vladimir State University
23 Orlovskaya Street, Murom, Vladimir region, 602200, Russia

Abstract. The results of experimental evaluation of CELP vocoders with vector quantization of gain coefficients are presented. The result of speech quality evaluation using the Diagnostic Acceptability Measure for codebooks of 64 and 128 4-dimensional vectors, generated from the learning set from 8-minute speech material from 12 speakers (5 female and 7 male), are discussed.

Speech coding methods developed during last twenty years provide good quality (legibility, natural sound, improved possibility for speaker recognition) when transmitting digitally encoded speech via narrow-band channels at speeds of 16 to 4.8 kbit/s. In practice Multi Pulse Linear Predictive Coding (MPLPC) and Code Excited Linear Prediction (CELP) are widely used [1].

The most perfect algorithm (in the sense of quality) is the CELP algorithm at 4.8 kbit/s, accepted as the USA federal standard FS-1016 [2].

In the FS-1016 CELP coder the speech signal S_n is divided into frames of 30 ms. For each frame, the parameters of the synthesis filter $1/A(z)$ are determined by the LPC analysis. The excitation signal is represented by the set of indices (C_s, C_a) of codebook vectors (C_s, C_a) from the stochastic codebook and the adaptive codebook, and by corresponding gain coefficients (g_s, g_a) . During the excitation signal encoding the frame is divided into four subframes having duration of 7.5 ms each. For each subframe, the following parameters are encoded: C_s (9 bits), C_a (8 bits in even subframes, 6 bits in odd subframes), g_s and g_a (5 bits each). The whole frame is encoded using 144 bits, from which 40 bits (28.8%) are used for gain coefficient encoding using scalar quantization.

The aim of this work is the experimental research of the CELP coder with vector quantization of gain coefficients. In particular the results of using separate vector quantization of gain coefficients g_s and g_a are discussed.

For each of the gain coefficients g_s and g_a , four values corresponding to subframes of one frame were combined into a single four-dimensional vector, giving two vectors $\{g_{s_1}, g_{s_2}, g_{s_3}, g_{s_4}\} = \mathbf{g}_s$ and $\{g_{a_1}, g_{a_2}, g_{a_3}, g_{a_4}\} = \mathbf{g}_a$ per frame. Separate codebooks were built for \mathbf{g}_s and \mathbf{g}_a from the learning set of 16000 vectors. There are two variants of each codebook with 64 and 128 vectors, which allows decreasing the number of bits for encoding of gain coefficients to 28 and 26 bits per frame, respectively.

The learning set was formed by processing the speech material from twelve speakers (5 female and 7 male), with the whole duration of 8 minutes, using the standard FS-1016 CELP coder. The K-means algorithm with the initial condition from the clustering with non-homogeneous dichotomy [3] was used. The speech quality was measured using the Diagnostic Acceptability Measure (DAM) through listening to 12 phonetically balanced 6-syllable sentences pronounced by speakers (male and female) which did not take part in building the learning set. The speech quality was estimated by ten listeners. From the results of listening the mean opinion score (preference percentage) was calculated.

The results of evaluating the coders with vector quantization of gain coefficients with different codebook sizes are presented in Table 1.

One can notice that most of the experts do not notice any difference in the speech quality of the FS-1016 CELP coder and the new coder with vector quantization of gain coefficients (70%–80% of the experts do not notice any difference, and only 20%–30% prefer the FS-1016 CELP coder). The bigger codebook has a slight advantage in quality (80% of the experts treat the coder with 128 vector codebook as not being worse than FS-1016 CELP; 70% think so about the new coder with 64 vector codebook).

Table 1: Coder Evaluation Results

Preference	64 vector codebook		128 vector codebook	
	Number of votes	Percentage	Number of votes	Percentage
Prefer FS-1016 CELP coder	3	30%	2	20%
No preference	7	70%	7	70%
Prefer the new coder	0	0%	1	10%
Total	10	100%	10	100%

In conclusion, vector quantization of gain coefficients allows decreasing the CELP coder bit rate from 4.8 kbit/s to 3.87–3.93 kbit/s without noticeable loss of quality (by decreasing the number of bits used for encoding the excitation signal by 28–26 bits per frame).

Bibliography

1. I. M. Trancoso, I. J. Marques. Speech Communication, 1990, v. 9, No. 5.
2. Federal Standard 1016. Telecommunications: Analog to Digital Conversion of Radio Voice by 4,800 bit/second Code Excited Linear Prediction (CELP). February 14, 1991.
3. Дж. Макхоул. Векторное квантование при кодировании речи. ТИИЭР, т. 73, № 11, ноябрь 1985.