

ИССЛЕДОВАНИЕ ВЕКТОРНОГО КВАНТОВАНИЯ РАСЩЕПЛЕННОГО МНОЖЕСТВА LSF-КОЭФФИЦИЕНТОВ В ПРИЛОЖЕНИИ К КОДИРОВАНИЮ ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ РЕЧЕВОГО СИГНАЛА В ВОКОДЕРНЫХ СИСТЕМАХ

Павловец А.Н., Петровский А.А., Савицкий А.

Белорусский государственный университет информатики и радиоэлектроники
220027, Минск, ул. П.Бровки, 6 (Беларусь), E-mail: palex@it.org.by

Введение

Векторное квантование широко применяется в низкоскоростных вокодерах [1], так как по сравнению со скалярным квантованием дает значительный выигрыш в объеме кода, не ухудшая качество. Исследования, рассматриваемые в данной статье, выполнялись в рамках работ по реализации низкоскоростного кодера речи, основанного на раздельном представлении периодической (тональной) и аperiodической (шумовой) компонент речевого сигнала [2, 3].

1. Алгоритм работы кодера

Анализ речевого сигнала при разделении его на периодическую и аperiodическую составляющие, происходит следующим образом: 1) на сегменте речи определяется частота основного тона; 2) вычисляется ДПФ, согласованное с частотой основного тона, результат – огибающая спектра периодической составляющей; 3) синтезированная периодическая составляющая вычитается из исходного сигнала, результат – шумовая составляющая; 4) над обоими компонентами проводится LP-анализ (10-го порядка – над периодической и 5-го порядка над аperiodической), после чего производится трансформация из LPC-коэффициентов в линейные спектральные пары LSF; 5) параметры, характеризующие речевой сегмент (частота основного тона F_0 , LSF-коэффициенты периодической и аperiodической компонент, коэффициенты усиления периодической и аperiodической компонент) подвергаются квантованию.

2. Векторное квантование линейных спектральных пар

Важным свойством LSF является локализованность их спектральной чувствительности (изменение одного LSF повлияет на форму спектра только в его окрестности). Данное свойство делает представление LPC-коэффициентов в форме LSF особенно удобным для векторного квантования с расщеплением (Split Vector Quantization – SVQ) [4]. Суть SVQ состоит в расщеплении входного вектора на субвекторы с последующим их независимым квантованием, что позволяет снизить сложность и требования к памяти квантователя.

Одним из преимуществ подхода SVQ является возможность простой модификации распределения бит между кодовыми субкнигами. Это свойство применяется в том случае, если какой-либо субвектор требует более точного квантования, чем остальные. Основными вопросами при разработке SVQ-квантователя являются способ расщепления входного вектора и распределение бит между субвекторами.

3. Постановка эксперимента и результаты

Для исследования векторного квантования с расщеплением воспользуемся методикой, схожей с [4, 5]. Исходные данные были взяты из базы данных TIMIT [6]. Для тренировки было отобрано 462 файла, которые подверглись низкочастотной фильтрации и децимации для получения речевого материала с частотой дискретизации 8000 Гц. После выделения по алгоритму кодера периодической компоненты были получены обучающие множества для квантования (33200 вектора LSF).

Оценка качества при исследовании проводилась с применением объективного параметра – логарифмического искажения спектра [4]:

$$sd = \sqrt{\frac{1}{\pi} \int_0^{\pi} [10 \log_{10} S(\omega) - 10 \log_{10} S'(\omega)]^2 d\omega}, \quad (1)$$

где $S(\omega)$ и $S'(\omega)$ – спектры мощности, соответствующие оригинальному синтезирующему LP-фильтру и модифицированному синтезирующему LP-фильтру.

В [4] были экспериментально установлены следующие условия, которые приводят к отсутствию слышимых искажений из-за спектральных несоответствий: среднее значение искажения спектра не превышает 1 дБ; количество сегментов, на которых sd принимает значения от 2 до 4 дБ, не превышает 2%; отсутствие сегментов со значением sd более 4 дБ.

Оценка качества проводилась с помощью тестовой базы данных, в которую вошли 168 файлов из базы данных TIMIT, над которыми были проведены те же манипуляции, что и с тренировочным набором. Объем тестового множества для оценки качества квантования составил 12182 вектора. Отметим, что тренировка и тестирование проводились на множествах векторов, полученных из разных речевых источников, со следующими вариантами расщеплений: 3-4-3, 3-3-4, 4-3-3, 4-4-2, 4-6, 5-5, 6-4. Разрядность трехкомпонентных книг варьировалась от 6 до 9, двухкомпонентных – от 9 до 12. Распределение бит для кодирования было равномерным. Начальная инициализация проводилась по алгоритму [7]. Идея его заключается в том, что тренировочные вектора, наиболее далекие друг от друга, с большей вероятностью принадлежат к разным классам. Для тренировки использовался обобщенный алгоритм Ллойда [8]. В

качестве параметра минимизации применялось взвешенное Эвклидово расстояние (2), что позволяет учесть особенности человеческого слуха [4]:

$$d(f, \hat{f}) = \sum_{i=1}^p [c_i w_i (f_i - \hat{f}_i)]^2, \quad (2)$$

где c_i и w_i – веса для f_i -го LSF-коэффициента, \hat{f}_i – квантованный LSF-коэффициент. Для LP-фильтра 10-го порядка фиксированные веса c_i определяются следующим образом:

$$c_i = \begin{cases} 1.0, & 1 \leq i \leq 8 \\ 0.8, & i = 9 \\ 0.4, & i = 10 \end{cases} \quad (3)$$

Эффективная схема для вычисления переменных весов w_i была предложена в [9]:

$$w_i = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i}, \quad (4)$$

где f_i – LSF-коэффициенты (в радианах), $f_0=0, f_{p+1}=\pi$.

Результаты тестирования сведены в таблицу 1 (представлены варианты для общего количества бит от 21 до 27, как обеспечивавшие лучшее качество квантования).

Таблица 1. Оценка качества для различных вариантов расщепления исходного вектора

Общее число бит	Вариант расщепления	SD (дБ)	Отклонения	
			2 – 4 дБ, %	> 4 дБ, %
21	4-3-3	1.4248	8.2335	0.057462
21	4-4-2	1.6766	20.399	0.18059
21	3-3-4	1.3172	5.7133	0.065671
21	3-4-3	1.5379	14.891	0.14776
22	4-6	1.3276	7.9462	0.0082088
22	5-5	1.2629	4.6955	0.024626
22	6-4	1.3683	7.1663	0.041044
24	4-3-3	1.1783	2.6761	0.016418
24	4-4-2	1.4353	9.4648	0.041044
24	3-3-4	1.0941	2.2738	0.016418
24	3-4-3	1.305	6.7723	0.032835
24	4-6	1.2349	5.2783	0.016418
24	5-5	1.1731	3.0537	0.0082088
24	6-4	1.2775	5.1798	0.024626
27	4-3-3	0.99107	1.1328	0.0082088
27	4-4-2	1.2298	3.9074	0.024626
27	3-3-4	0.91127	0.98506	0.016418
27	3-4-3	1.109	3.2014	0.024626

Очевидно, что квантование, близкое к прозрачному, может быть получено при емкости около 24 бит / вектор, причем лучшими вариантами расщеплений являются 3-3-4 и 4-3-3. Для исследования влияния распределения бит между кодовыми книгами на качество квантования было проведено тестирование на том же тестовом наборе по тем же параметрам, рассматривались трехкомпонентные книги. Лучшие варианты распределения бит для различных вариантов расщепления приведены в таблице 2 (рассматривалось распределение бит в общей емкости кода, 23 – 24 бит / вектор).

Таблица 2. Оценка качества для различного распределения бит между кодовыми книгами

SD (дБ)	Отклонения		Общее число бит	Вариант расщепления / распределение бит
	2 – 4 дБ, %	> 4 дБ, %		
1.0766	1.5925	0.0164	24	3-3-4 / 6-9-9
1.0954	2.0276	0.0082	24	4-3-3 / 9-9-6
1.1903	4.0305	0.0328	24	3-4-3 / 8-9-7
1.2432	4.0716	0.0328	24	4-4-2 / 9-9-6
1.1543	2.2164	0.0164	23	3-3-4 / 6-8-9
1.1746	2.7828	0.0082	23	4-3-3 / 8-9-6
1.235	4.4246	0.0328	23	3-4-3 / 7-9-7
1.315	5.2044	0.0164	23	4-4-2 / 8-9-6

Перераспределение бит заметно улучшило качество квантования для всех вариантов расщеплений. Наиболее приемлемым для использования в практических целях является вариант расщепления 3-3-4 с распределением бит между кодовыми книгами 6-8-9, всего 23 бит / вектор.

На рисунке 1 (а и б) приведены спектрограммы оригинального и квантованного сигналов (периодической составляющей). Кодирование происходило по методу SVQ, вариант 3-3-4, распределение бит между кодовыми книгами – 6-8-9.

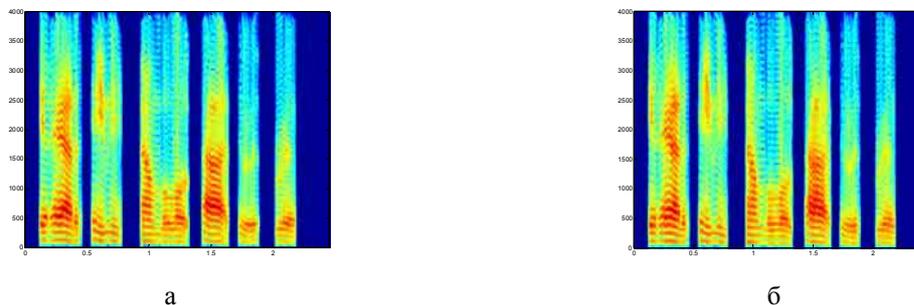


Рис. 1. Спектрограмма периодической составляющей речевого сигнала: а) до квантования; б) после квантования.

Выводы

На основании исследований выбрана схема квантования, построены кодовые книги. Огибающая спектра периодической составляющей представляется с использованием 23 бит / вектор. Дальнейшая работа будет направлена на использование принципов психоакустики для снижения объема кода.

Литература

1. Gersho, A. and Gray, R. Vector quantization and signal compression. Boston, Kluwer Academic Publishers, 1992.
2. Petrovsky, A., Zubricki, P., Savicki, A. “Tonal and noise components separation based on a pitch synchronous DFT analyzer as a speech coding method”, in Proceedings European Conference on Circuit Theory and Design, (Cracow, Poland), vol. 3, pp. 169 – 172, Sept. 2003.
3. 4. Петровский А.А., Серков В.В. Низкоскоростной вокодер с моделью речеобразования «гармоники + шум». Цифровая обработка сигналов, №2, 2002. – с.2-12.
4. Palival, K.K. and Atal, B.S. “Efficient vector quantization of LPC parameters at 24 bits/frame”, IEEE Transactions on Acoustics, Speech and Signal Processing., vol.1, № 1, pp. 3 – 14, Jan. 1993.
5. Sinervo, U., Nurminen, J., Heikkinen, A. and Saarinen, J. “Evaluation of split and multistage techniques in LSF quantization”, in Proceedings IEEE Nordic Signal Processing Symposium, 2001
6. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus, Department of Commerce, NIST, Springfield, Virginia, Oct. 1990.
7. Katsavounidis, I., Kuo, C. - C.J., Zhang, Z. “A new initialization technique for Generalized Lloyd Iteration”, IEEE Signal Processing Letters, vol. 1, № 10, pp. 144 – 146, Oct. 1994.
8. Linde, Y., Buzo, A., and Gray, R.M. “An algorithm for vector quantizer design”, IEEE Transactions on Communications, vol. COM-28, pp. 84 – 95, Jan. 1980.
9. Laroia, R., Phamdo, N., Farvardin, N. “Robust and efficient quantization of speech LSP parameters using structured vector quantizers”, in Proceedings IEEE International Conference on Acoustics, Speech, Signal Processing, (Toronto, Canada), pp. 641 – 644, May 1991.

