

А. Н. Ткаченко, к. т. н., доц.

МЕТОД ВЕКТОРНОГО ДЕЛЬТА-КВАНТОВАНИЯ ПАРАМЕТРОВ РЕЧЕВОГО СИГНАЛА

В статье предложен метод векторного квантования LSF-параметров речевого сигнала с прогнозированием последующего значения. Основная идея метода состоит в том, что вместо квантования действительного вектора LSF кодируется разница между действительным и спрогнозированным значениями. Такой подход позволяет уменьшить динамический диапазон входных величин и соответственно ошибку квантования. Разработана итерационная процедура создания кодовых книг для реализации предложенного метода. Приведены экспериментальные результаты апробации метода для различных скоростей кодирования речевого сигнала.

Ключевые слова: речевой сигнал, параметризация, кодовые книги, раздельное и многоэтапное векторное квантование, вектора LSF-параметров, ошибка кластеризации, коэффициент корреляции, спектральное искажение.

Введение

Замеченная Г. Муром закономерность возрастания производительности процессоров вдвое каждые два года действительно подтверждена практикой последних 40 лет [1]. Приблизительно такими же темпами возрастали и объемы мультимедийной информации, передающейся и хранящейся в компьютерных системах. Однако наличие фундаментальных ограничений, связанных с атомарной природой вещества и скоростью распространения света, заставили в 2007 г. Г. Мура признать, что закон вскоре перестанет действовать. В то же время аналогичных ограничений на возрастание объемов информации на сегодняшний день не обнаружено.

Таким образом, существует проблема недостаточной производительности компьютерных систем, предназначенных для обработки мультимедийной, в частности, голосовой информации. Уменьшение объемов данных, необходимых для корректного воспроизведения цифрового речевого сигнала, позволяет существенно повысить эффективность работы таких систем.

На сегодняшний день самую высокую степень сжатия обеспечивают параметрические методы с последующим квантованием полученных параметров. На этапе параметризации сигнал разбивают на кадры (фреймы) одинаковой длины, как правило, по 10 – 30 мс, и для каждого из них вычисляют некоторый вектор параметров. В современных системах сжатия речевых сигналов стандартом де-факто стало вычисление параметров на основе модели линейного прогнозирования десятого порядка. На этапе квантования полученный вектор параметров заменяют так называемым квантованным вектором – ближайшим ко входному вектору представителю репрезентативной совокупности параметров, хранящихся в кодовых книгах (КК). В работе [2] было показано:

1. В качестве параметров квантования целесообразно использовать линейные спектральные частоты (LSF), которые вычисляют по значениям коэффициентов линейного прогнозирования (LPC).

2. Оптимальное квантование полного LSF-вектора является сложным в вычислительном отношении, поэтому его разбивают на два или три подвектора, каждый из которых квантуется раздельно. Альтернативным вариантом служит многоэтапное квантование полного вектора при помощи КК меньшего размера, при котором на каждом последующем этапе квантуется ошибка, оставшаяся после квантования на предыдущем этапе.

3. Ошибку квантования для i -го фрейма оценивают по спектральному искажению SD_i :

$$SD_i^2 = \frac{1}{F_s} \int_0^{F_s} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df,$$

где $P_i(f) = 1/|A_i(\exp(j2\pi f / F_s))|^2$, $\hat{P}_i(f) = 1/|\hat{A}_i(\exp(j2\pi f / F_s))|^2$ – спектральные плотности начального и квантованного векторов для i -го фрейма, $A_i(z)$, $\hat{A}_i(z)$ – начальный и квантованный LPC-полиномы, соответствующие i -му фрейму, F_s – частота дискретизации.

4. Для выполнения условий транспарентности среднее значение ошибки квантования SD_{av} по всем фреймам должно составлять примерно 1 дБ; число фреймов, для которых $SD > 2\text{дБ}$, не должно превышать 2%; фреймов, для которых $SD > 4\text{дБ}$, не должно быть вообще.

5. Указанных условий достигают при разбиении вектора LSF-параметров на два подвектора и использовании для квантования входного вектора 24 бита на один фрейм.

Дальнейшему уменьшению объема данных для описания спектральной информации был посвящен несколько работ [3, 4]. Однако задача уменьшения битовой скорости, необходимой для передачи речевой информации, при условии сохранения транспарентности квантования, а также выполнения существующих ограничений по объемам памяти, сложности вычислений и задержки кодирования, остается актуальной.

В работе предложен метод сжатия речевой информации, позволяющий уменьшить объемы данных для описания спектральной информации до 20 бит на один фрейм при сохранении приемлемого качества воспроизводимого речевого сигнала.

Метод векторного дельта-квантования с прогнозированием

Любой метод сжатия основан на использовании избыточности, присущей речевому сигналу. Эта избыточность сохраняется и после перехода к параметрическому описанию сигнала. Количественной мерой избыточности служит коэффициент корреляции. В таблице 1 приведены значения коэффициента корреляции между отдельными параметрами LSF в пределах фрейма, полученные для тренировочной последовательности из 90000 векторов.

Таблица 1

Коэффициенты корреляции LSF в пределах фрейма

№	1	2	3	4	5	6	7	8	9	10
1	1,00	0,45	-0,05	-0,07	-0,17	-0,27	-0,28	-0,28	-0,24	-
2		1,00	0,57	0,23	0,11	-0,02	0,03	0,06	-0,04	-
3			1,00	0,56	0,32	0,39	0,38	0,38	0,19	-
4				1,00	0,58	0,46	0,44	0,26	0,19	-
5					1,00	0,62	0,42	0,24	0,01	-
6						1,00	0,62	0,45	0,20	-
7							1,00	0,63	0,25	-
8								1,00	0,58	-
9									1,00	-
10										1,00

В таблице 2 приведены значения коэффициента корреляции между соответствующими

LSF-параметрами для последовательных фреймов.

Таблица 2

Коэффициенты корреляции LSF для соседних фреймов

№	1	2	3	4	5	6	7	8	9	10
	0,68	0,73	0,76	0,81	0,86	0,84	0,82	0,81	0,76	0,75

Как видно из таблиц, наблюдается довольно высокая корреляция как между отдельными параметрами в пределах фрейма, так и между компонентами вектора LSF для соседних фреймов. Возможны следующие варианты использования этих двух видов корреляции:

- переход от скалярного квантования (СК) параметров к векторному квантованию (ВК);
- применение методов прогнозирования с последующим удалением прогнозируемых значений с целью уменьшения динамического диапазона входных величин.

Существует принципиальная возможность применения данных подходов для использования как межфреймовой корреляции, так и корреляции в пределах фрейма. Однако СК LSF-параметров с прогнозированием в пределах фрейма, хотя и требует минимальных вычислительных затрат, значительно уступает по эффективности ВК, что выражается в увеличении спектрального искажения [3]. В то же время квантование LSF-параметров для последовательных фреймов связано с внесением дополнительной задержки, что недопустимо для некоторых приложений. Таким образом, целесообразным представляется вариант ВК параметров LSF для отдельных фреймов с последующим прогнозированием значений для последующих фреймов. Именно этот подход, названный автором векторным дельта-квантованием с прогнозированием (ВДКП), предложен в данной работе.

В основе ВК при помощи КК лежит простая идея: вместо передачи действительного значения параметров входного речевого сигнала вычисляется и передается индекс ближайшего ко входному вектору из КК. При этом в КК хранятся наиболее репрезентативные значения параметров, полученные в результате кластеризации тренировочной последовательности входных векторов LSF. В отличие от этого, по методу ВДКП в КК хранятся значения параметров, полученные в результате кластеризации разницы (дельты) e_i действительного l_i и спрогнозированного значения $\tilde{l}_i$ LSF. При условии удачного прогноза, динамический диапазон входных величин, использующихся для создания КК, и соответственно, ошибка кластеризации должны уменьшиться, что, в свою очередь, приведет к уменьшению спектрального искажения в процессе квантования.

Таким образом, в процессе создания КК входная последовательность будет состоять из векторов $e_i = l_i - \tilde{l}_i$. При использовании авторегрессионной модели линейного прогнозирования первого порядка спрогнозированное значение $\tilde{l}_i(m)$ m -го компонента LSF для i -го фрейма вычисляют по формуле:

$$\tilde{l}_i(m) = \alpha(m)\hat{l}_{i-1}(m) + \beta(m), \quad (1)$$

где параметры $\alpha(m)$ и $\beta(m)$ вычисляют, исходя из условия минимизации среднеквадратической ошибки прогнозирования:

$$\alpha(m) = \frac{COV(l_i(m), \hat{l}_{i-1}(m))}{VAR(l_i(m))}, \quad (2)$$

$$\beta(m) = E(l_i(m)) - \alpha(m)E(\hat{l}_{i-1}(m)), \quad (3)$$

где E , COV , VAR обозначают соответственно математическое ожидание, ковариацию и

дисперсию, $\hat{l}_{i-1}(m)$ – квантованное значение m -го компонента вектора LSF на $i-1$ -м фрейме; $\hat{l}_{i-1}(m) = l_{i-1}(m)$.

Процедура квантования LSF-параметров по методу ВДКП представлена на рис. 1.

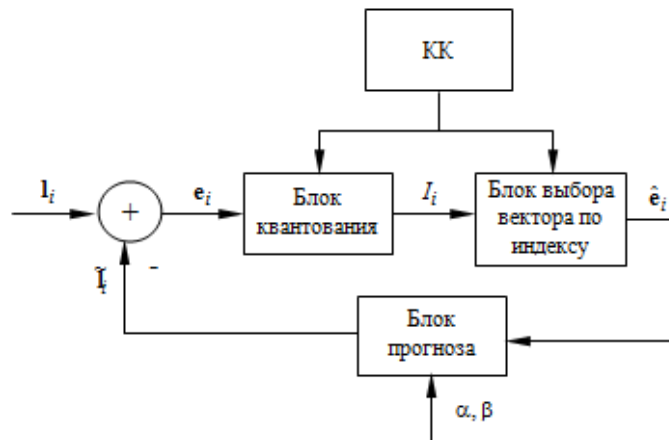


Рис. 1. Структурная схема квантования по методу ВДКП

Создание кодовых книг по методу ВДКП

Параметры α и β вычисляют для тренировочной последовательности векторов LSF. Это означает невозможность непосредственного применения формул (2) и (3) в процессе создания КК, поскольку значения $\hat{l}_{i-1}(m)$ вычисляют только в процессе квантования, которое, в свою очередь, может происходить лишь при наличии КК.

Таким образом, практическое применение ВДКП требует реализации итерационной процедуры создания КК.

Шаг 1. Для $i = 1, 2, \dots, N$, $m = 1, 2, \dots, M$ присвоить $\hat{l}_{i-1}(m) = l_{i-1}(m)$.

Шаг 2. По формулам (2) и (3) вычислить параметры α, β .

Шаг 3. По формуле (1) вычислить прогнозируемые значения параметров LSF $\tilde{l}_i$.

Шаг 4. Вычислить значения векторов ошибки $\mathbf{e}_i = \mathbf{l}_i - \tilde{\mathbf{l}}_i$.

Шаг 5. Для полученных векторов ошибки $\mathbf{e}_i$ по методу k-средних создать КК $\mathbf{Y}$.

Шаг 6. Используя параметры α, β и созданную КК $\mathbf{Y}$, провести квантование векторов тренировочной последовательности $\mathbf{l}_i$ и получить квантованные значения векторов $\hat{\mathbf{l}}_i$.

Шаг 7. Вычислить ошибку квантования. Если она изменилась на достаточно малую величину по сравнению с предыдущей итерацией, закончить процедуру. Если нет, возвратиться к шагу 2.

Таким образом, процедура создания КК состоит из трех этапов: вычисления параметров прогноза, квантования, а также непосредственно кластеризации. Создание начальной КК целесообразно производить по усовершенствованному методу k-средних [5]. На последующих итерациях можно применять классический алгоритм k-средних, используя имеющиеся векторы $\mathbf{Y}$ для инициализации центроидов. Схематически процедура создания КК представлена на рис. 2:

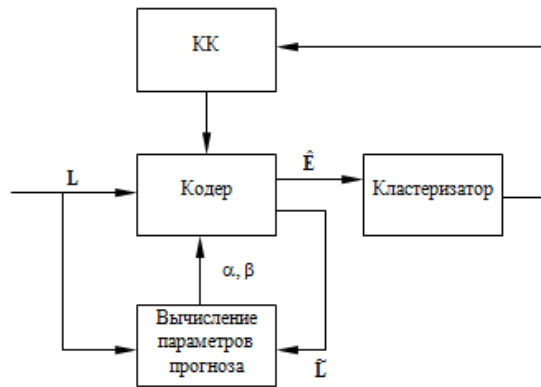


Рис. 2. Структурная схема создания КК для метода ДВКП

Рассмотренную выше процедуру создания КК можно использовать как для отдельного, так и для многоэтапного ВК. В последнем случае процедура создания КК может потребовать дополнительных итераций, обусловленных необходимостью согласования между собой отдельных частей КК [6]. Однако повышение производительности, выражающееся в уменьшении спектрального искажения в процессе квантования, является достаточным обоснованием дополнительных временных затрат на стадии создания КК, которая выполняется один раз на подготовительном этапе.

Экспериментальные результаты

Для экспериментального исследования разработанного метода была использована общедоступная часть англоязычного акустического корпуса TIMIT [7]. Тренировочная последовательность состояла из 90000 векторов LSF, полученных на основе модели линейного прогнозирования десятого порядка. Тестовая последовательность состояла из 15000 векторов LSF, отличающихся от векторов тренировочной последовательности. Длина фрейма составляла 20 мс. Расстояние измерялось по взвешенной Евклидовой метрике с использованием весов разрядов, вычисленных по спектральной чувствительности [8].

В качестве основы для сравнения была выбрана схема отдельного векторного квантования (РВК) с разбиением 10-мерного вектора LSF на два подвектора размерностью пять. Такое же разбиение использовалось при отдельном векторном квантовании с прогнозированием (РВДКП). При использовании многоэтапного квантования процесс кодирования происходил в два этапа, размеры обеих КК были одинаковы. Моделировали два варианта поиска: последовательный, при котором для поиска на втором этапе использовался один ближайший вектор, полученный на первом этапе, и поиск по дереву, при котором на втором этапе использовали N_1 векторов, полученных на первом этапе. Соответствующие схемы обозначены как МЭВДКП с ПП и МЭВДКП с ДП.

В таблицах 3 – 5 приведены результаты, полученные при использовании для квантования входного вектора 24, 22 и 20 бит на один фрейм соответственно. Производительность квантования оценивали по спектральному искажению.

Таблица 3

Производительность квантования спектральной информации 24 битами на фрейм

Схема квантования	SD, дБ	% превышений	
		SD > 2 дБ	SD > 4 дБ
РВК	1,18	1,96	0,00
РВДКП	0,93	1,45	0,00
МЭВДКП с ПП	0,94	1,45	0,00
МЭВДКП с ДП, $N_1 = 10$	0,92	1,27	0,00
МЭВДКП с ДП, $N_1 = 30$	0,90	1,12	0,00

Таблица 4

Производительность квантования спектральной информации 22 битами на фрейм

Схема квантования	SD, дБ	% превышений	
		SD > 2 дБ	SD > 4 дБ
РВК	1,34	6,28	0,06
РВДКП	1,06	3,56	0,03
МЭВДКП с ПП	1,06	3,57	0,03
МЭВДКП с ДП, $N_1 = 10$	1,01	2,23	0,01
МЭВДКП с ДП, $N_1 = 30$	0,98	2,04	0,00

Таблица 5

Производительность квантования спектральной информации 20 битами на фрейм

Схема квантования	SD, дБ	% превышений	
		SD > 2 дБ	SD > 4 дБ
РВК	1,51	15,23	0,16
РВДКП	1,22	7,11	0,10
МЭВДКП с ПП	1,23	7,13	0,11
МЭВДКП с ДП, $N_1 = 10$	1,10	3,63	0,05
МЭВДКП с ДП, $N_1 = 30$	1,08	3,42	0,03

Как видно из таблиц, условия транспарентности (см. п. 4 введения) выполняются при использовании для описания спектральной информации 20 бит на один фрейм сигнала.

Следует отметить, что вычислительная сложность метода ВДКП практически совпадает с РВК. Лишь при использовании поиска по дереву количество операций увеличивалось в N_1 раз, что обусловлено особенностями схемы поиска, а не метода ВДКП. Возможность уменьшения вычислительной сложности поиска ближайшего вектора рассмотрена в [9].

На рис. 3 представлена ошибка квантования как функция от номера итерации в процессе создания КК. Ошибку вычисляют как значение расстояния между входным вектором и ближайшим вектором КК и усредняют по всем фреймам. Как можно видеть, уже на пятой итерации ошибка квантования изменяется лишь на 0,7%.



Рис. 3. Зависимость ошибки квантования от номера итерации

Выводы

Разработанный метод ВДКП позволяет повысить производительность квантования параметров речевого сигнала благодаря уменьшению динамического диапазона входных

векторов, использующихся для создания КК и, соответственно, уменьшению ошибки кластеризации. Это достигается за счет усложнения подготовительного этапа, который требует итерационной процедуры создания КК. Экспериментальная проверка разработанного метода показала, что условия транспарентности выполняются для двухэтапного ВДКП с поиском по дереву при использовании для квантования 20 бит на один фрейм речевого сигнала.

СПИСОК ЛИТЕРАТУРЫ

1. Родоначалник [Электронный ресурс] / Ю. Ревич // Компьютерра-Онлайн. – 2012. – Режим доступа до журн.: <http://www.computerra.ru/xterra/homo/21659/>.
2. Paliwal K. K. Efficient vector quantization of LPC parameters at 24 bits/frame / K. K. Paliwal, B. S. Atal. // IEEE Transaction on Speech and Audio Processing. – 1993. – No. 2, vol. 1. – P. 3 – 14.
3. Lahouti F. Quantization of LSF parameters using a trellis modeling / F. Lahouti, A. K. Khandani // IEEE Transaction Speech and Audio Processing. – Sep. 2003. – No. 5, vol. 11. – P. 400 – 412.
4. Hai Le Vu. Efficient Distance Measure for Quantization of LSF and Its Karhunen–Loeve Transformed Parameters / Hai Le Vu and Laszlo Lois // IEEE Transactions on speech and audio processing. – Nov. 2000. – No. 6, vol. 8. – P. 744 – 746.
5. Ткаченко О. М. Метод кластеризації на основі послідовного запуску k-середніх з обчисленням відстаней до активних центроїдів / О.М. Ткаченко, Н. О. Біліченко, О. Ф. Грійо Тукало, О. В. Дзісь // Реєстрація, зберігання і обробка даних. – 2012. – № 1, Т. 14. – С. 25 – 34.
6. Chu W. C. Speech Coding Algorithms: Foundation and Evolution of Standardized Coders / Wai C. Chu // NY.: John Wiley & Sons, Inc. – 2003 – 558 p.
7. Fisher William M. The DARPA Speech Recognition Research Database: Specifications and Status / William M. Fisher, George R. Doddington, Kathleen M. Goudie-Marshall // Proceedings of DARPA Workshop on Speech Recognition. – Feb. 1986. – P. 93 – 99.
8. Gardner W. R. Theoretical analysis of the high-rate vector quantization of LPC parameters / W. R. Gardner, B. D. Rao // IEEE Transaction Speech and Audio Processing. – Sep. 1995. – Vol. 3. – P. 367 – 381.
9. Ткаченко О. М. Двоетапна стратегія пошуку у векторних кодових книгах для ущільнення мовлення / О. М. Ткаченко, О. Ф. Грійо Тукало // Вісник Вінницького політехнічного інституту. – 2011. – № 6. – С. 194 – 201.

Ткаченко Александр Николаевич – к. т. н., доцент кафедры вычислительной техники, тел. 59-84-13, e-mail: ant@vstu.vinnica.ua.

Винницкий национальный технический университет.