

Метод векторного дельта-квантування параметрів мовленнєвого сигналу з прогнозуванням

Ткаченко О. М., Грійо Тукало О. Ф., Дзісь О. В.

Вінницький національний технічний університет

Вінниця, Україна

ant@vstu.vinnica.ua, xxmargox@gmail.com, alexdz47@gmail.com

Анотація

У статті запропоновано метод векторного квантування LSF-параметрів мовленнєвого сигналу з прогнозуванням наступного значення. Основна ідея методу полягає в тому, що замість квантування дійсного вектора LSF кодується різниця між дійсним та прогнозованим значеннями. Такий підхід дозволяє зменшити динамічний діапазон вхідних величин і відповідно похибку квантування. Розроблено ітераційну процедуру побудови кодових книг для реалізації запропонованого методу. Наведено експериментальні результати апробації методу для різних швидкостей кодування мовленнєвого сигналу.

1. Вступ

Передавання та зберігання мовленнєвої інформації є основою функціонування сучасних телекомунікаційних та комп'ютерних систем. Зменшення обсягів даних, необхідних для коректного відновлення цифрового мовленнєвого сигналу, дозволяє суттєво підвищити продуктивність роботи таких систем.

У процесі ущільнення цифрового мовленнєвого сигналу можна виділити дві стадії: параметризацію та квантування. На першій стадії сигнал розбивають на ділянки (фрейми) однакової довжини, як правило, по 20 – 30 мс, та для кожної з них обчислюють деякий вектор параметрів. У сучасних системах ущільнення мовлення стандартом де-факто стало обчислення параметрів на основі моделі лінійного прогнозування десятого порядку.

На другій стадії отриманий вектор параметрів замінюють так званим квантованим вектором – найближчим до вхідного представником репрезентативної сукупності параметрів, що зберігаються в кодових книгах. У роботі [1] було показано

1. Як параметри квантування доцільно використовувати лінійні спектральні частоти (LSF), що обчислюються за значеннями коефіцієнтів лінійного прогнозування (LPC).

2. Оптимальне квантування повного LSF-вектора є обчислювально складним, тому його розбивають на 2 або 3 підвектори.

3. Похибку квантування для i -го фрейму оцінюють за спектральним спотворенням SD_i :

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

де $P_i(f) = 1/|A_i(\exp(j2\pi f/F_s))|^2$, $\hat{P}_i(f) = 1/|\hat{A}_i(\exp(j2\pi f/F_s))|^2$ – спектральні щільності відповідно початкового і квантованого векторів для i -го фрейму,

$A_i(z), \hat{A}_i(z)$ – початковий та квантований LPC-поліноми, що відповідають i -му фрейму,

F_s – частота дискретизації.

4. Для досягнення транспарентності середнє значення похибки квантування SD_{av} по всім фреймам має складати приблизно 1 дБ; кількість фреймів, для яких $SD_{av} > 2\text{дБ}$, не повинна перевищувати 2%; фреймів, для яких $SD_{av} > 4\text{дБ}$ не повинно бути взагалі.

5. Зазначені умови досягаються при розбитті вектора LSF-параметрів на два підвектори та використанні для квантування вхідного вектора 24 бітів на один фрейм.

Подальшому зменшенню обсягу даних для опису спектральної інформації було присвячено цілу низку робіт [2, 3]. Проте задача зменшення бітової швидкості, необхідної для передавання мовленнєвої інформації, за умов збереження транспарентності квантування, а також виконання існуючих обмежень за обсягами пам'яті, складністю обчислень та затримкою кодування, залишається актуальною.

В роботі пропонується метод ущільнення мовленнєвої інформації, який дозволяє зменшити обсяги даних для опису спектральної інформації до 20 бітів на один фрейм при збереженні придатної якості відновленого мовленнєвого сигналу.

2. Метод векторного дельта-квантування з прогнозуванням

Будь-який метод ущільнення базується на використанні надлишковості, що притаманна мовленнєвому сигналу. Кількісною мірою надлишковості слугує коефіцієнт кореляції. У таблиці 1 наведено значення коефіцієнта кореляції між окремими значеннями LSF у межах фрейму.

Табл. 1. Коефіцієнти кореляції LSF всередині фрейму

№	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	-0,10
2	0,45	1,00	0,57	0,23	0,11	-0,02	0,03	0,06	-0,04	-0,01
3	-0,05	0,57	1,00	0,56	0,32	0,39	0,38	0,38	0,19	0,06
4	-0,07	0,23	0,56	1,00	0,58	0,46	0,44	0,26	0,19	0,03
5	-0,17	0,11	0,32	0,58	1,00	0,62	0,42	0,24	0,01	0,01
6	0,27	-0,02	0,39	0,46	0,62	1,00	0,62	0,45	0,20	0,01
7	-0,28	0,03	0,38	0,44	0,42	0,62	1,00	0,63	0,25	0,22
8	-0,28	0,06	0,38	0,26	0,24	0,45	0,63	1,00	0,58	0,30
9	0,24	0,04	0,19	0,19	0,01	0,20	0,25	0,58	1,00	0,56
10	-0,15	-0,01	0,06	0,03	0,01	0,01	0,22	0,30	0,56	1,00

У таблиці 2 наведено значення коефіцієнта кореляції між відповідними LSF-параметрами для послідовних фреймів.

Табл.2. Коефіцієнти кореляції LSF для сусідніх фреймів

№	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

Як можна побачити з таблиць, спостерігається доволі висока кореляція як між окремими параметрами всередині фрейму, так і між компонентами вектора LSF для сусідніх фреймів. Можливі такі шляхи використання цих двох видів кореляції.

Перехід від скалярного квантування (СК) параметрів до векторного квантування (ВК).

Застосування методів прогнозування з подальшим видаленням прогнозованих значень з метою зменшення динамічного діапазону вхідних величин.

Існує принципова можливість застосування обох підходів для використання як міжфреймової кореляції, так і кореляції всередині фрейму. Проте СК LSF-параметрів з прогнозуванням всередині фрейму, хоча вимагає мінімальних обчислювальних витрат, значно поступається за ефективністю ВК, що виражається у збільшенні спектрального спотворення [Farshad Lahouti]. В той же час квантування LSF-параметрів для послідовних фреймів пов'язано з внесенням додатково затримки, що є неприпустимим для деяких застосувань. Таким чином, доцільним виглядає варіант ВК параметрів LSF для окремих фреймів з подальшим прогнозуванням значень для послідовних фреймів. Саме цей підхід, названий авторами векторним дельта-квантуванням з прогнозуванням (ВДКП) розглядається у даній роботі.

В основу ВК за допомогою кодових книг (КК) покладено просту ідею: замість передачі дійсного значення параметрів вхідного мовленнєвого сигналу, обчислюється та передається індекс вектора з КК, який є найближчим до вхідного. При цьому у КК зберігаються найбільш репрезентативні значення параметрів, отримані в результаті кластеризації тренувальної послідовності вхідних векторів LSF. На відміну від цього, за методом ВДКП у КК зберігаються значення параметрів, отримані як результат кластеризації різниці (дельти) e_i дійсного I_i та прогнозованого значення \tilde{I}_i LSF. За умов вдалого прогнозу динамічний діапазон вхідних величин, що використовуються для створення КК, і, відповідно, похибка кластеризації має зменшитися, що, в свою чергу, призведе до зменшення спектрального спотворення в процесі квантування.

Таким чином, в процесі створення КК вхідна послідовність буде складатися з векторів $e_i = I_i - \tilde{I}_i$. При використанні авторегресійної моделі лінійного прогнозування першого порядку прогнозоване значення $\tilde{I}_i(m)$ m -го компонента LSF для i -го фрейму обчислюється за формулою:

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

де параметри $\alpha(m)$ та $\beta(m)$ обчислюються, виходячи з умови мінімізації середньоквадратичної похибки прогнозування:

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

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

де E, COV, VAR позначають відповідно математичне сподівання, коваріацію та дисперсію,

$\hat{I}_{i-1}(m)$ - квантоване значення m -го компонента вектора LSF на $i-1$ -му фреймі; $\hat{I}_{i-1}(m) = I_{i-1}(m)$.

Процедуру квантування LSF-параметрів за методом ВДКП представлено на рис.1.

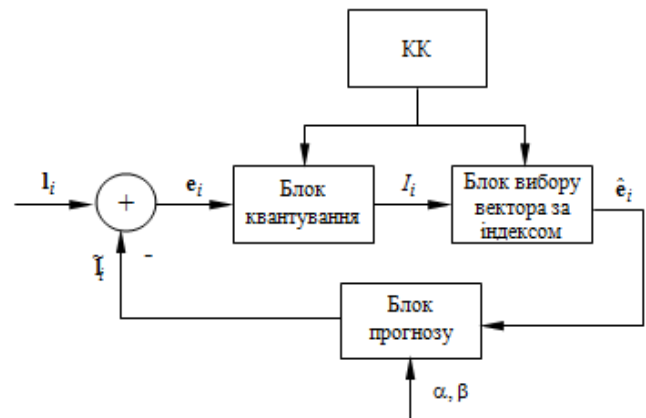


Рис. 1. Структурна схема квантування за методом ВДКП

3. Створення кодових книг за методом ВДКП

Параметри α та β обчислюються для векторів тренувальної послідовності векторів LSF. Це означає неможливість безпосереднього застосування формул (2) та (3) в процесі створення КК, оскільки значення $\hat{I}_{i-1}(m)$ обчислюються лише під час квантування, яке, в свою чергу, може відбуватися лише за наявності КК.

Таким чином, практичне застосування ВДКП потребує реалізації ітераційної процедури створення КК.

Крок 1. Для $i = 1, 2, \dots, N$, $m = 1, 2, \dots, M$ присвоїти

$$\hat{I}_{i-1}(m) = I_{i-1}(m),$$

Крок 2. За формулами (2) та (3) обчислити параметри α, β .

Крок 3. За формулою (1) обчислити прогнозовані значення параметрів LSF \tilde{I}_i .

Крок 4. Обчислити значення векторів похибки $e_i = I_i - \tilde{I}_i$.

Крок 5. Для отриманих векторів похибки e_i за методом k-середніх створити кодову книгу γ .

Крок 6. Використовуючи параметри α, β та створену кодову книгу γ провести квантування тренувальної

послідовності векторів \mathbf{l}_i та отримати квантовані значення векторів параметрів $\hat{\mathbf{l}}_i$.

Крок 7. Обчислити похибку квантування. Якщо вона змінилася на достатньо малу величину порівняно з попередньою ітерацією, завершити процедуру. Якщо ні, повернутися до кроку 2.

Таким чином, процедура створення КК складається з трьох частин: обчислення параметрів прогнозу, квантування, а також безпосередньо кластеризації. Створення початкової КК доцільно використовувати за вдосконаленим методом к-середніх [4]. На наступних ітераціях можна застосувати класичний алгоритм к-середніх, використовуючи наявні вектори \mathbf{y} для ініціалізації центрів. Схематично процедуру створення КК представлено на рис.2

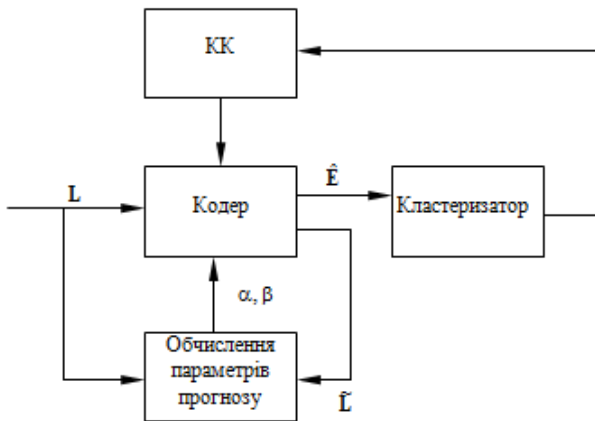


Рис. 2. Структурна схема створення КК для методу ДВКП

Розглянуту вище процедуру створення КК можна використовувати як для роздільного (split), так і для багатоетапного (multistage) ВК. В останньому випадку процедура створення КК може потребувати додаткових ітерацій, зумовлених необхідністю узгодження між собою окремих частин КК [5]. Проте підвищення продуктивності, що виражається у зменшенні спектрального спотворення в процесі квантування, є достатнім обґрунтуванням додаткових часових витрат на стадії створення КК, яка виконується один раз на підготовчому етапі.

4. Експериментальні результати

Для експериментального дослідження розробленого методу було використано загальнодоступну частину англомовного акустичного корпусу ТІМІТ. Тренувальна послідовність складалася з 90000 векторів LSF, отриманих на основі моделі лінійного прогнозування 10-го порядку. Тестова послідовність складалася з 15000 векторів LSF, що відрізнялися від векторів тренувальної послідовності. Довжина фрейму складала 20 мс.

Вимірювання відстані відбувалося за зваженою Евклідовою метрикою з використанням ваг розрядів, обчислених за спектральною чутливістю [6].

За базу для порівняння було обрано схему роздільного векторного квантування (РВК) з розбиттям

10-вимірному вектора LSF на два підвектори розмірності п'ять. Таке ж розбиття використовувалося при роздільному векторному квантуванні з прогнозуванням (РВДКП). При використанні багатоетапного квантування процес кодування відбувався у два етапи, розміри обох КК були однакові. Моделювалися два варіанти пошуку: послідовний, коли для пошуку на другому етапі використовувався лише один найкращий вектор, отриманий на першому етапі, і пошук по дереву, коли для пошуку на другому етапі використовувалися N_1 векторів, отриманих на першому етапі. Відповідні схеми позначено як РВДКП з ПП та РВДКП з ДП.

У таблицях 3 – 5 наведено результати, отримані при використанні для квантування вхідного вектора 24, 22 та 20 бітів на один фрейм відповідно. Продуктивність квантування оцінювалася за спектральним спотворенням.

Табл.3. Продуктивність квантування спектральної інформації 24 бітами на фрейм

Схема квантування	SD, дБ	% викидів	
		2 – 4	>4
РВК	1.18	1.96	0.00
РВДКП	0.93	1.45	0.00
БЕВДКП з ПП	0.94	1.45	0.00
БЕВДКП з ДП, $N_1 = 10$	0.92	1.27	0.00
БЕВДКП з ДП, $N_1 = 30$	0.90	1.12	0.00

Табл.4. Продуктивність квантування спектральної інформації 22 бітами на фрейм

Схема квантування	SD, дБ	% викидів	
		2 – 4	>4
РВК	1.34	6.28	0.06
РВДКП	1.06	3.56	0.03
БЕВДКП з ПП	1.06	3.57	0.03
БЕВДКП з ДП, $N_1 = 10$	1.01	2.23	0.01
БЕВДКП з ДП, $N_1 = 30$	0.98	2.04	0.00

Табл.5. Продуктивність квантування спектральної інформації 20 бітами на фрейм

Схема квантування	SD, дБ	% викидів	
		2 – 4	>4
РВК	1.51	15.23	0.16
РВДКП	1.22	7.11	0.10
БЕВДКП з ПП	1.23	7.13	0.11
БЕВДКП з ДП, $N_1 = 10$	1.10	3.63	0.05
БЕВДКП з ДП, $N_1 = 30$	1.08	3.42	0.03

Як можна побачити з таблиць, умови прозорості виконуються при використанні для опису спектральної інформації 20 бітів на один фрейм мовленнєвого сигналу.

Слід зазначити, що обчислювальна складність квантування за методом ВДКП практично збігається з РВК. Лише при застосуванні пошуку по дереву кількість операцій збільшувалася в N_1 разів, що зумовлено особливостями схеми пошуку, а не методу ВДКП.

Можливість зменшення обчислювальної складності пошуку найближчого вектора розглянуто в [7].

На рис. 3 представлено похибку квантування як функцію від номера ітерації в процесі створення КК. Похибка обчислювалася як значення відстані між вхідним вектором і найближчим вектором КК та усереднювалася по всім фреймам. Як можна побачити, вже на п'ятій ітерації похибка квантування змінюється лише на 0.7%.

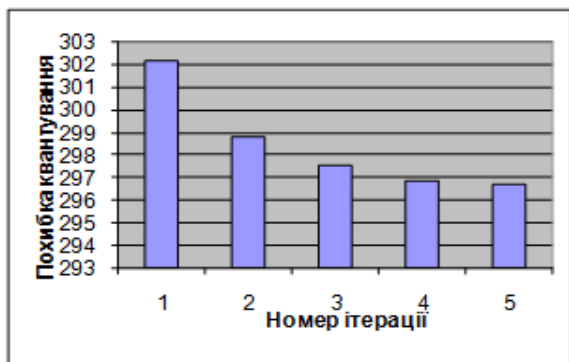


Рис. 3. Залежність похибки квантування від номеру ітерації

5. Висновки

Розроблений метод ВДКП дозволяє підвищити продуктивність квантування параметрів мовленнєвого сигналу завдяки зменшенню динамічного діапазону вхідних векторів, що використовуються для створення КК і, відповідно, зменшенню похибки кластеризації. Це досягається за рахунок ускладнення підготовчого етапу, який потребує ітераційної процедури створення КК. Експериментальна перевірка розробленого методу показала, що умови транспарентності виконуються для двохетапного ВДКП з пошуком по дереву при використанні 20 бітів на один фрейм для опису спектральної інформації.

6. Список літератури

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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 // *Регстрація, зберігання і обробка даних*. — 2012. — № 1, Т. 14. — С. 25-34.

- [5] Chu W. C. *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders* / Wai C. Chu // NY.: John Wiley & Sons, Inc. — 2003 — 558 p. — ISBN 0-471-37312-5.
- [6] 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.
- [7] 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.