Дельта-ущільнення мовленнєвих сигналів із застосуванням методу визначення голосової активності

О. М. Ткаченко, к.т.н., доц., О. Д. Феферман, С. В. Хрущак

Вінницький національний технічний університет
ant@vstu.vinnica.ua, sergey.khruschak@gmail.com

Анотація

При передачі мовлення цифровими каналами одним з найбільш важливих питань є обсяг інформації, який доведеться передавати за одиницю часу для забезпечення якісного голосового зв'язку. Ущільнення мовлення зменшує обсяги даних, що передаються, та апаратні витрати.

У даній статті розглянуто процес ущільнення мовленнєвих сигналів за допомогою вокодера LPC та запропоновано низку підходів для подальшого зменшення обсягів даних. Розроблено метод дельта-ущільнення з прогнозуванням наступного індексу квантованого значення в кодовій книзі та вдосконалено метод визначення голосової активності, описаний у рекомендації G.729, Annex B. Проведено експериментальне дослідження запропонованих методів.

Вступ

Переважна більшість алгоритмів ущільнення мовлення використовують модель лінійного прогнозування для опису та кодування сигналу, відповідно до якої спектральна інформація про сигнал описується 10-ма коефіцієнтами лінійного прогнозування (LPC). Для більш ефективного квантування та подальшої інтерполяції коефіцієнти LPC, як правило, перетворюють на лінійні спектральні пари (LSP) та квантують їх за допомогою кодових книг.

Коефіцієнти LSP, отримані в результаті кодування мови за методом лінійного прогнозування, мають залежності між елементами як у середині фрейму, так і між сусідніми фреймами, що свідчить про високий рівень надлишковості при використанні даних коефіцієнтів [1].

Врахування кореляції між коефіцієнтами дозволяє зменшити обсяг даних, необхідних для опису параметрів мовленнєвого сигналу, і таким чином зменшити вимоги до пропускної здатності каналу передачі інформації. У даному дослідженні розглядається параметрична система ущільнення мовлення, яка враховувує кореляцію між коефіцієнтами всередині фрейму сигналу та між послідовними фреймами й таким чином дозволяє досягти більшого ущільнення.

Подальшим етапом зменшення обсягів даних, які необхідно передавати в канал зв’язку, є використання дискретної передачі (Discontinuous transmission - DTX) - вимкнення передавача або передача комфортного шуму (Comfort noise - CN) на приймальну сторону під час пауз у розмові. Такий підхід, наприклад, використовується у мережі GSM. Основною задачєю при цьому є визначення голосової активності (Voice Activity Detection - VAD). Більшість підходів базуються на енергетичних порогах, визначенні періоду основного тону, спектральному аналізі, частоті переходів через нуль або комбінації кількох методів. Складність задачі полягає у тому, що постійну високу точність визначення вокалізованості фрейму сигналу неможливо досягнути, оскільки більшість алгоритмів базується на певному пороговому значенні, яке є фіксованим або обраховується з сигналу на шумових сегментах. Так, наприклад, при використанні для VAD методу найменших квадратів шумові сегменти використовуються у фільтрі з кінцевою імпульсною характеристикою для його тренування. Але у випадках, коли фоновий шум не є стаціонарним або коли мовленнєвий сигнал є переважно вокалізованим, для досягнення високої точності алгоритму VAD необхідно постійно корегувати порогове значення незалежно від типу сегменту. Для цього необхідний метод, який міг би ефективно адаптуватись до зміни фонового шуму.

У даній роботі розглядається метод визначення голосової активності, запропонований міжнародною асоціацією ITU-T - G.729 Annex B [3], який розроблявся для мультимедіа та IP-телефонії. Даний метод запезпечує адаптацію до високого рівня фонового шуму та використовує дані, які отримуються в процесі кодування, що зменшує його обчислювальну складність. Але незважаючи на це він має низьку швидкодію, порівняно з іншими стандартами (ETSI AMR, статистичні методи), особливо при низькому співвідношенні сигнал/шум. Тому запропоновано деякі зміни до даного методу, для покращення його швидкодії та проведено його дослідження з використанням запропонованих змін.

1. Врахування кореляції між коефіцієнтами

Врахування кореляції між коефіцієнтами одного фрейму відбувається за рахунок використання векторного квантування коефіцієнтів LSP. При векторному квантуванні набір коефіцієнтів LSP розглядається як один вектор, для якого відбувається пошук найближчого квантованого значення у спеціальній таблиці - векторній кодовій книзі [4].

Для врахування кореляції між послідовними фреймами в [4] запропоновано метод впорядкування векторів згідно з відношенням мажорування, що дозволяє здійснити перехід до передачі різниці між індексами (дельти) послідовних фреймів. Недоліком запропонованого методу є значна затримка (100 мс), необхідна для досягнення якісного сигналу, при зменшенні обсягу інформації для опису параметрів мовленнєвого сигналу з 24 до 20 бітів, що неприпустимо в деяких системах зв’язку. З огляду на це запропоновано метод дельта-ущільнення з прогнозуванням наступного відліку, який дозволяє отримати меншу затримку (20 мс), при меншій точності передачі індексу.

Для зменшення вікна дельти між індексами сусідніх фреймів при досить малій затримці, порівняно з методом

запропонованим в [5], пропонується прогнозування індексу фрейму на основі попередніх значень. Прогнозування відбувається шляхом створення екстраполюючої функції для індексів кількох фреймів на приймальній та передавальній стороні. При ущільненні на приймальну сторону також передається дельта між прогнозованим та реальним значеннями. Ілюстрацію запропонованого методу представлено на рисунку 1.

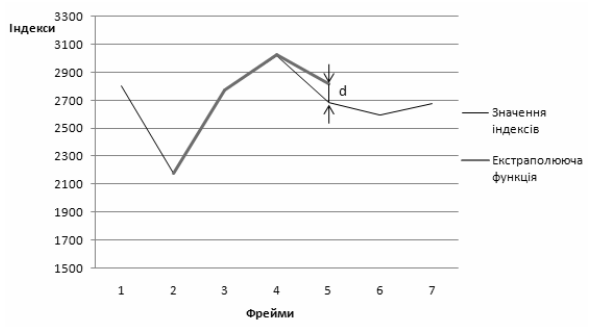

Рисунок 1: Приклад роботи методу ущільнення з екстраполяцією.

Екстраполююча функція будується за методом найменших квадратів. Суть даного методу полягає в тому, що функція для опису експериментальних даних g будується як лінійна комбінація M базисних функцій F_j [6]:

$$g = \sum_{i=0}^{M-1} c_j F_j(x). \qquad (1)$$

При цьому коефіцієнти c_j вибираються таким чином, щоб мінімізувати критерій $C_{нмк}$ – суму квадратів відхилень екстраполюючої функції від експериментальних значень:

$$C_{нмк} = \sum_{i=1}^{n} (y_i - g_i)^2 \to \min C. \qquad (2)$$

де y_i – експериментальна функція.

Досліджувалися також альтернативні методи – метод найменших модулів (МНМ) та метод Чебишева. Для МНМ критерієм мінімізації є сума відхилення модулів (3):

$$C_{мнм} = \sum_{i=1}^{n} \left| y_i - g_i \right|. \qquad (3)$$

Для методу Чебишева критерієм мінімізації є максимальне відхилення [6]:

$$C_{мч} = \max \left| y_i - g_i \right| \cdot \qquad (4)$$

Проте тільки метод найменших квадратів дозволяє знайти найкращі c_j за кінцеве число операцій, зводячи задачу до вирішення системи лінійних рівнянь. Це означає, що він є найбільш простим для обчислення, до того ж він точніше описує експериментальну функцію.

Введення затримки дозволяє корегувати значення індексів, що передаються; це необхідно в ситуаціях, коли прогноз виявляється невірним (рисунок 2). В даному випадку ми маємо спрогнозувати індекс 5-го фрейму на основі попередніх – 2-го, 3-го та 4-го. Проте прогнозування тільки погіршило результат – значення дельти зросло.

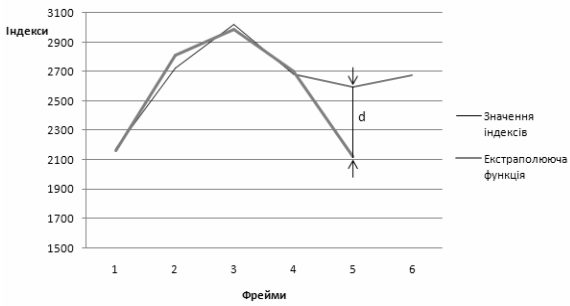

Рисунок 2: Приклад невірного прогнозу.

У випадках, коли при передачі параметрів i-го фрейму виконується нерівність (5):

$$\Delta_{i+1} > S_w, \qquad (5)$$

де Δ_{i+1} - значення дельти для i+1 фрейму, S_w - розмір вікна застосовується такий алгоритм:

1. Здійснюється апроксимація значень експериментальної функції, враховуючи наступні значення індексів i+1, i+2, …, i+k, де k – величина затримки в фреймах. Для апроксимації також використовується метод найменших квадратів.
2. Здійснюється корекція значення i-го індексу відповідно до апроксимуючої функції, щоб отримати кращий прогноз для наступного фрейму.
3. Здійснюється прогноз i+1 фрейму на основі попередніх фреймів i, i-1, i-2, ..., i-p, де p – кількість фреймів для створення прогнозу, яка визначається порядком моделі прогнозування.
4. Якщо різниця між прогнозованим та реальним значенням виходить за межі вікна, то корекція проводиться знову - відбувається перехід до п.2. У граничному випадку, коли при максимально припустимому відхиленні i-го фрейму прогноз є неуспішним, то відбувається перехід до п.5.
5. Різниця між корегованим значенням та прогнозованим, враховуючи значення індексів для i-1, i-2, ..., i-(p+1) фреймів, передається у канал зв'язку.

Результат роботи даного алгоритму можна бачити на рисунку 3. Після корекції значення індексу для 4-го фрейму прогноз відбувається успішно – значення дельти залишається незначним.

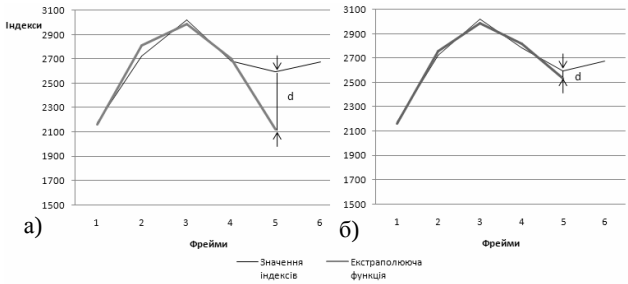

Рисунок 3: Графіки експериментального та екстрапольованого значення індексів:
а) до виконання корекції;
б) після виконання корекції.

2. Покращення алгоритму визначення голосової активності

Рекомендація ITU-T G.729 Annex B розроблялась як доповнення до вокодера G.729. Рішення про вокалізованість приймаються для фреймів тривалістю 10 мс на основі таких характеристик:

1. Диференційна енергія сигналу для всього діапазону спектру – ΔE_f;
2. Диференційна енергія сигналу для низькочастостного спектру (0-1 кГц) – ΔE_l;
3. Диференційна частота переходів через нуль – ΔZCR;
4. Спектральне спотворення – ΔLSF.

Схему даного VAD наведено на рисунку 4.

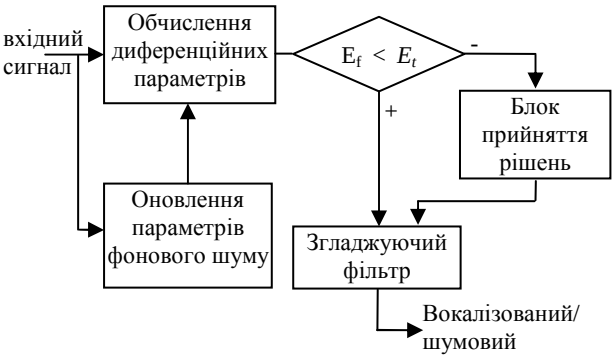

Рисунок 4: Блок схема алгоритму ITU-T G.729 Annex B.

Вхідні параметри отримуються з мовленнєвого сигналу використовуючи авторегресійну модель першого порядку, якщо виконується умова $\Delta E_f < E_t$, де E_t - задане порогове значення. Проте дані параметри потребують значної складності обчислень, також використання частоти переходів через нуль та порогу різниці енергії привносять похибки при високій зашумленості сигналу.

Для покращення якості запропонованого методу пропонується прискорити оцінку спектральної обвідної за рахунок згладжування за допомогою МНК. За рахунок цього значно легше обчислити пологість спектру, яка є ключовою при визначенні вокалізованості фрейму сигналу (для невокалізованого фрагменту - спектральна обвідна буде більш полога). Приклад такого згладжування спектральної обвідної зображено на рисунку 5.

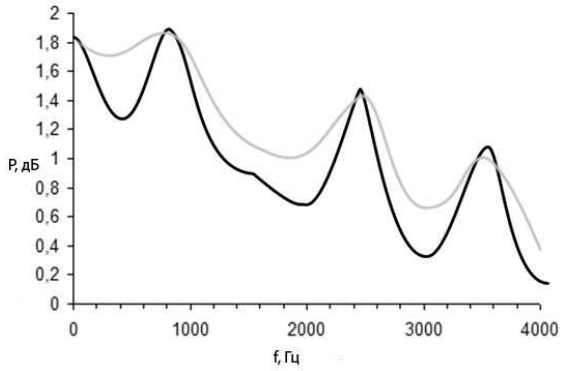

Рисунок 5: Приклад згладжування спектральної обвідної сигналу.

Таким чином, це дозволяє зменшити кількість обчислень, оскільки пологість спектру рахується тільки по кільком точкам, та підвищити точність алгоритму, особливо у низькочастотній смузі сигналу.

Для обчислення оптимального порогового значення вокалізованості сигналу застосовується новий підхід – геометрично адаптивний енергетичний поріг, запропонований Езером та Таньєром [7]. Даний підхід базується на використанні ймовірнісного розподілу амплітуд та дозволяє встановлювати порогове значення адаптивно не тільки використовуючи шумові сигнали, але й вокалізовні. Даний метод є стійким до нестаціонарного шуму, але іноді може давати неправильний результат при різких й коротких стрибках шуму. Завдяки простоті обчислення адаптивного енергетичного порогу також зменшується складність алгоритму VAD.

3. Дослідження методу дельта ущільнення

Тестування методу прискореного пошуку векторів відбувалося на основі фонетичного матеріалу, який складався з двох фонетично повних текстів [4]. При використанні методу дельта-ущільнення з прогнозуванням значення індексу в кодовій книзі для наступного фрейму, прогнозування відбувається шляхом створення екстраполюючої функції для індексів кількох фреймів. Досліджувалися декілька методів побудови екстраполюючої функції, а саме метод найменших квадратів (МНК), метод Чебишева та метод найменших модулів (МНМ).

За критерій оцінювання методів використовувались дисперсія розподілу ймовірності дельти D між послідовними фреймами та спектральне спотворення SD. Результати оцінювання наведено в табл. 1.

Таблиця 1 Результати дослідження методів екстраполяції для кодової книги розміром 4096 векторів

Назва методу	D	Середнє SD, дБ	Відсоток фреймів, для яких SD>2 дБ, %	Відсоток фреймів, для яких SD > 4 дБ, %
МНК	21531	0,882	1,78	0
Метод Чебишева	31095	0,912	1,96	0
МНМ	141950	0,956	2,15	0

Як можна бачити з таблиці 1, найкращим за показниками дисперсії кінцевого розподілу та спектральним спотворенням є МНК. Для МНМ кількість фреймів зі спектральним спотворенням в межах 2-4 дБ перевищує припустиме значення в 2%, прийняте для систем низькошвидкісного ущільнення мовлення.

4. Дослідження методу визначення голосової активності

Загальноприйнятою практикою оцінювання якості методів VAD є використання одного з двох видів тестів – об'єктивних або суб'єктивних. До суб'єктивних відносять методи, в яких робота алгоритму визначається за допомогою оцінок слухачів. Проте така схема не дозволяє визначити кількість невірних спрацювань при високому рівні шуму, оскільки це виражатиметься в збільшенні кількості даних в каналі, а не якості результуючого

мовлення, тому було обрано об'єктивний метод оцінювання VAD. Для тестів було обрано мовленнєву базу TIMIT, в записи якої додано випадковий шум та поділено на фрейми, які потім розмічено вручну на вокалізовані та шумові, тобто фактично порівняння проводилось з ідеальним варіантом. Дані файли оброблялися за допомогою різних алгоритмів визначення активності – запропонованого, оригінального, ETSI AMR 1 та ETSI AMR 2, після чого підраховувалися кількості фрагментів, на яких було обрізання мовлення та на яких шум було розпізнано як мовлення. Результати тестування наведено у таблиці 2.

Таблиця 2 Результати тестування якості

Назва методу	Фреймів з обрізанням мовлення, %	Шумових фреймів розпізнаних як вокалізовані, %
Покращений ITU-T G.729B	83%	2%
ITU-T G.729B	68%	5%
AMR 2	87%	0,3%
AMR 1	76%	5,5%

Результати тестування показали що покращений метод ITU-T G.729B дозволяє отримати кращу якість розпізнавання мовленння на зашумлених фреймах, ніж звичайний метод, але дещо гіршу, ніж досягається в методі з використанням статистики високого порядку, який, проте, є більш обчислювально складним. Якщо враховувати, що в середньому при двосторонній бесіді кожен з співбесідників розмовляє близько 40% [2], то даний метод може зменшити навантаження на канал зв’язку до 60%, хоча в реальних умовах цифра не перевищує 45%.

Для дослідження швидкодії запропонованого покращеного методу визначення голосової активності використовувались матеріали з набору TIMIT, оскільки це дозволяє порівнювати отримані результати з зарубіжними аналогами. Результати порівняння покращеного методу ITU-T G.729 Annex B та існуючих методів для різного розміру аналізованих блоків наведено у таблиці 3.

Таблиця 3 Середня кількість операцій для запропонованого та відомих методів при різному розмірі аналізованого блоку

Назва методу	Кількість математичних операцій для заданого розміру аналізованого блоку, мс			
	10	20	50	100
Покращений ITU-T G.729B	$1{,}7\cdot10^4$	$2{,}4\cdot10^4$	$5{,}2\cdot10^4$	$14{,}3\cdot10^4$
ITU-T G.729B	$2{,}2\cdot10^4$	$2{,}7\cdot10^4$	$6{,}1\cdot10^4$	$18{,}6\cdot10^4$
AMR 2	$2{,}9\cdot10^4$	$2{,}6\cdot10^4$	$4{,}8\cdot10^4$	$9{,}8\cdot10^4$
AMR 1	$1{,}6\cdot10^4$	$2{,}0\cdot10^4$	$3{,}9\cdot10^4$	$7{,}2\cdot10^4$

Як видно з таблиці, вдосконалений метод працює швидше, ніж оригінальний, та швидше, ніж метод, що базується на статистиці високого порядку. Метод з використанням адаптивних енергетичних рівнів є найшвидшим, але й найменш точним.

5. Висновки

Дослідження методу дельта-ущільнення з прогнозуванням показало, що його використання дозволяє зменшити кількість бітів для передачі мовленнєвих параметрів з 24 до 20 при затримці в 1 фрейм та незначному (SD = 0,882 дБ) спотворенні сигналу. Застосування дискретної передачі пакетів дозволяє додатково зменшити навантаження на канал на 45%. Запропонований вдосконалений метод ITU-T G.729B дозволяє отримати кращу якість розпізнавання зашумлених фремів, ніж звичайний та меншу обчислювальну складність, ніж метод ETSI AMR 2.

6. Список літератури

[1] W. C. Chu Speech Coding Algoritms – Foundation and Evolution of Standardized Coders. New Jersey:Wiley, - 2003. P. 553.

[2] J. Ramirez, J. M. Gorriz and J. C. Segura (2007). Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, Robust Speech Recognition and Understanding, Michael Grimm and Kristian Kroschel (Ed.), ISBN: 978-3-902613-08-0, I-Tech Education and Publishing, Vienna, Austria.

[3] A. Benyassine, E. Shlomot, and H.-Y. Su, “ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data application,”

[4] Біліченко Н. О., Ткаченко О. М., Феферман О. Д., Хрущак С. В. LSF-вокодер на основі векторного квантування //Реєстрація, зберігання і обробка даних. – 2007. – №1. – С. 35 – 41.

[5] Ткаченко О.М., Феферман О.Д., Хрущак С. В. Метод дельта-ущільнення мовних сигналів //Інформаційні технології та комп’ютерна інженерія. Вінниця: УНІВЕРСУМ-Вінниця. - 2008.- № 1(11). – С. 8-13.

[6] Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика: Учебное пособие. - М.: Высшая школа, 1982. - 256 с.

[7] H. Özer and S. G. Tanyer, “A geometric algorithm for voice activity detection in nonstationary Gaussian noise,” in Proc. EUSIPCO’98, Rhodes, Greece, Sept. 1998.