

## РОЗРОБЛЕННЯ І ДОСЛІДЖЕННЯ ЯКОСТІ VoIP-КОДЕРІВ МОВНИХ СИГНАЛІВ

© Тимченко О., Колодій Р., Сайдекс Х.А., 2008

**Заропоновано перспективний метод кодування мовного сигналу зі змінною швидкістю для передачі через мережі з комутацією пакетів на основі вейвлет-перетворення, досліджено його якість та стійкість до впливу мережних факторів. Наведено блок-схему його реалізації, що додатково враховує голосову активність та пакетизацію сигналу.**

**The perspective method of code of speech signal is offered variable-speed for a transmission through networks with commutation of packages on the basis of wavelet transformation, his quality and firmness is probed to influence of network factors. Blok-schema of his realization which additionally takes into account speech activity and to the package formation signal is resulted.**

### Вступ

Мовні повідомлення як джерело інформації є важливим елементом мультимедійної системи в різноманітних застосуваннях та під час передачі через телекомунікаційні канали. Тому особливу увагу привертають алгоритми цифрового оброблення мультимедійних даних, а також методи підвищення ефективності компресії та якісного відтворення мови за умов втрат даних в мережі. Це вимагає провести порівняльну оцінку алгоритмів роботи кодерів в середовищі IP (VoIP) та їх оцінки суб'єктивними та об'єктивними методами.

**Мета роботи** – дослідити та розробити нові алгоритми роботи кодерів мовних сигналів для VoIP з врахуванням об'єктивних якісних показників психоакустичного сприйняття.

### 1. Аналіз існуючих алгоритмів компресії і кодування мовних сигналів та методів оцінки якості психоакустичного сприйняття

В середовищі VoIP шлюзи конфігуруються на оцифровку мови за допомогою кодування форми сигналу та гібридних методів кодування [1]. Кодери форми сигналу (PCM, ADPCM та CELP) характеризуються різними часовими та обчислювальними затратами. Проте їх застосування (крім CELP) обмежене через необхідність виділення широкої смуги пропускання і неможливість роботи зі змінною швидкістю кодування.

Гібридні кодери (MPEG) поєднують кодування форми сигналу з кодуванням параметрів. Параметри використовуються для синтезу фрагмента мови. Такі методи кодування, як MPEG, не можуть бути застосовані до вузькосмугових каналів через велику кількість додаткової і службової інформації, а також за наявності втрат пакетів в мережі. На противагу алгоритми кодерів на основі CELP вирізняються високою якістю мовного сигналу, оскільки працюють з його формою, проте їх реалізація має велику обчислювальну складність.

Проведений аналіз та дослідження алгоритмів компресії мовних сигналів [2–4], результати яких наведено у табл. 1, дали змогу встановити, що основними шляхами покращання якості роботи кодерів в мережах з комутацією пакетів є використання:

- методів лінійного передбачення;
- психоакустичних особливостей сприйняття мовного сигналу людиною;
- маскування фрагментів, що мають незначне навантаження;
- динамічного розподілу біт під час кодування відліків;
- завадостійкого кодування найважливішої частини інформації.

Кодери для IP-мереж, що реалізують ці алгоритми, повинні бути стійкими до втрат пакетів в мережі, забезпечувати згладжування джитера і мінімальну затримку кодування, а також відповідати

особливостям трафіку IP-мереж – роботі зі змінною швидкістю кодування. Їх головне завдання – досягнення максимальної компресії мовного сигналу за мінімального суб’єктивного (на слух) спотворення. Повною мірою ні один з відомих кодерів не відповідає цим вимогам.

Тест MSE/P % (відношення квадрата середньоквадратичного відхилення MSE до потужності сигналу P у цьому пакеті) запропонований як об’єктивний алгоритм оцінки якості мовного сигналу, що є простим та доступним методом і може бути використаний для оцінки роботи існуючих та розроблених алгоритмів кодерів мовних сигналів. Відповідність тесту MSE суб’єктивним оцінкам MOS для найпоширеніших кодерів мовних сигналів показано на рис. 1.

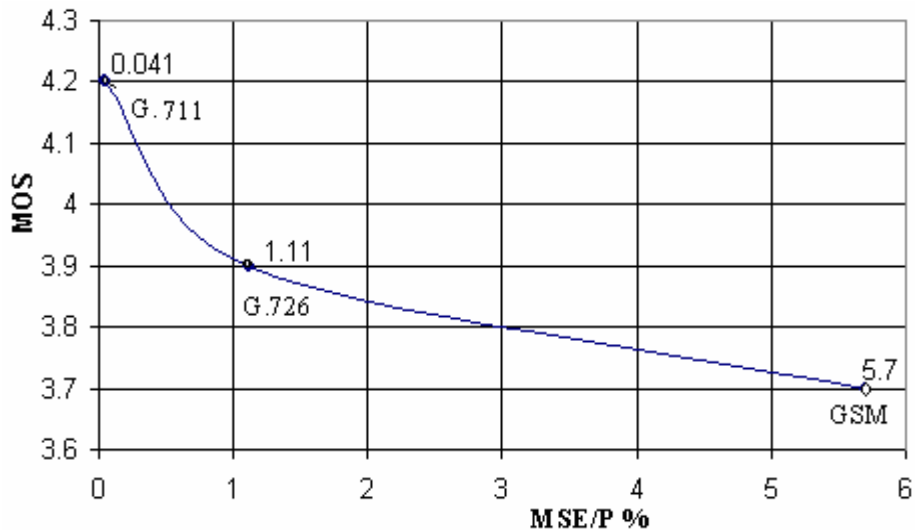


Рис. 1. Перерахунок MSE–MOS для кодерів мовних сигналів за відсутності втрат пакетів

## 2. Розроблення адаптивного алгоритму кодування та компресії мовного сигналу зі змінною швидкістю

Перспективний метод кодування на основі вейвлет-перетворення (рис. 2) реалізує кодування форми мовного сигналу і тому має практично максимальну якість MOS [5]. Він включає таку послідовність операцій. Мовний сигнал  $x(t)$  піддається аналого-цифровому перетворенню з параметрами PCM [1], що створює цифровий потік (наприклад, стандартний 64 кбіт/с). Цей потік розділяється на пакети, довжина яких відповідає мінімальному інтервалу стаціонарності (переважно по 20 мс), які обробляються детектором голосової активності (VAD), після чого сегменти мови, що залишились, піддають дискретному вейвлет-перетворенню (CWT). CWT використовується як метод оброблення сигналу, внаслідок якого отримуються коефіцієнти, що перетворюються у коефіцієнти апроксимації та деталізації (коефіцієнти A та D відповідно). Отримані коефіцієнти квантуються у квантувачах ( $Q_A$ ,  $Q_D$ ) з різною розрядністю кодування, внаслідок чого генерується вихідний потік із змінною швидкістю [6]. Оцифровані коефіцієнти CWT мультиплекуються у загальний потік та піддаються кодуванню для передачі каналом зв’язку.

Для підвищення ефективності роботи кодера в структурній схемі (рис. 2) блок пакетизації враховує стан детектора VAD для того, щоб в періоди мовчання або пауз, що виникають під час розмови абонентів, зменшувати обсяг передаваних даних і замінювати їх включенням генератора комфортного шуму на приймальній стороні. У функції останнього входить генерація фонового сигналу для слухаючого абонента, що імітує присутність зв’язку з тим, щоб абоненти не роз’єдналися у періоди пауз, вирішивши, що зв’язок втрачено. Цей блок, як правило, входить в обладнання шлюзу IP-телефонії і описаний в стандарті H-323.

У розробленому алгоритмі пропонується використовувати VAD за рівнем. Алгоритм роботи VAD дає можливість обнулити пакети мовних відліків за наявності в них відліків, менших за встановлений поріг.

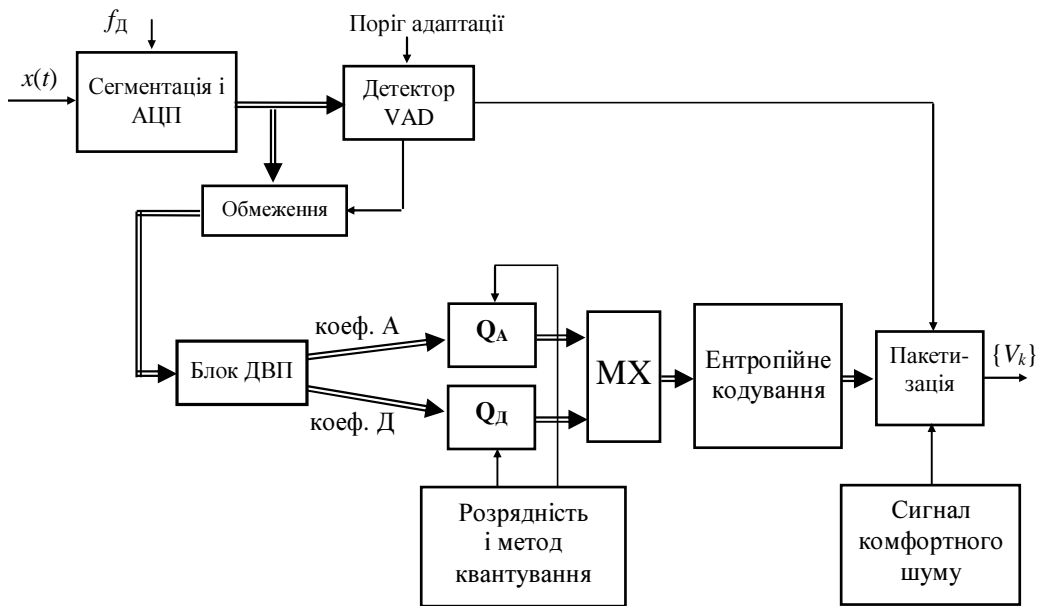


Рис. 2. Структурна схема кодера за методом вейвлет-перетворення

Робота алгоритму зводиться до розрахунку  $X_{max}$  – максимального відліку сигналу в усіх сегментах фрагмента мови, що піддається обробленню (по замовчуванню встановлюється  $X_{max}=1$ ). Потім:

- 1) встановлюється  $X_{nor}$ . (виражається в кількості рівнів  $k_{рівн.}=1...32$ ) по відношенню до  $X_{max}$ :

$$X_{nor} = \frac{k_{розр}}{k_{max}} X_{max}, \text{ де } k_{max}=127;$$

- 2) проводиться порівняння усіх відліків у сегментах з  $X_{nor}$ . та визначається їх кількість  $m_{відліків} = \sum_{i \in k} (x_i < X_{nor})$ , після чого встановлюється поріг для кількості відліків, які можуть бути відкинуті в сегменті (наприклад, 80 %) і менші  $X_{nor}$ . Якщо  $m_{відліків} < K \cdot 0,8 \cdot k_{рівн}$  (в подальшому моделюванні вибрано  $K=160$  відліків), то приймається рішення, що усі відліки сегменту дорівнюють нулю;

- 3) після оброблення одного сегменту переходимо до наступного (до п. 2), поки не перевіримо весь файл.

Зауважимо, що цей алгоритм VAD є поточним і обробляє мовний сигнал в міру його поступлення. Після VAD сегменти мови подаються на кодер, який виконує вейвлет-перетворення, після чого проводиться зворотне декодування і збирання сегментів.

Після оброблення мовного фрагмента детектором VAD активні фрагменти піддаються дискретному вейвлет-перетворенню (CWT), який пропонується реалізовувати за допомогою ліфтинг-схеми. Така реалізація вимагає мінімальних обчислювальних і часових затрат і здійснюється за алгоритмом, після якого отримують дві послідовності відліків – апроксимації  $\{CA_{j,i}\}$  та деталізації  $\{CD_{j,i}\}$ .

Базис вейвлет-перетворення представляється дискретними значеннями вейвлет-фільтрів  $\psi_t$ ,  $\varphi_t$ , де  $t=0...t_m$ . Нехтуючи знаком вибірки, що враховується за зворотного перетворення, пряме вейвлет-перетворення зводиться до обчислення дискретної згортки:

$$CA_{j,i} = \sum_{k=0}^{(R_p+P_p)+t_m} |XP_{j,k}| \cdot \varphi_{i-k}; \quad CD_{j,i} = \sum_{k=0}^{(R_p+P_p)+t_m} |XP_{j,k}| \cdot \psi_{i-k},$$

де  $R_p$  – довжина пакета;  $P_p$  – перекриття між пакетами;  $XP_{j,i} = x_k$  – значення відліку ( $i$ ) в пакеті ( $j$ ),  $i=0..(R_p+P_p)$ ,  $k=i+jR_p$ , при цьому:

$$\begin{cases} \psi_x = \psi_t, \varphi_x = \varphi_t; j = 0..t_m; \\ \psi_x = \varphi_x = 0; j \neq 0..t_m, \end{cases}$$

де  $\{CA_{j,i}\}, \{CD_{j,i}\}$  – масиви коефіцієнтів апроксимації і деталізації.

Додатково, згідно зі структурою кодера (рис. 2), потік коефіцієнтів апроксимації та деталізації  $\{CA_{j,i}\}, \{CD_{j,i}\}$  піддається ентропійному кодуванню, що реалізує усунення довгих послідовностей однакових значень коефіцієнтів. Це додатково приводить до збільшення коефіцієнта компресії.

Після кодування вихідний потік коефіцієнтів перетворення пакетизують. Пакетизація відбувається шляхом заповнення поля корисного навантаження протоколу RTP сформованими мовними кадрами та додавання до них заголовків транспортного, каналного та фізичного рівнів згідно з моделлю OSI. Формується бітовий потік для пакетизації за стандартом H-323, встановленим для мереж з підтримкою послуги VoIP.

Результати моделювання і дослідження методу кодування мовного сигналу показано на рис. 3–5.

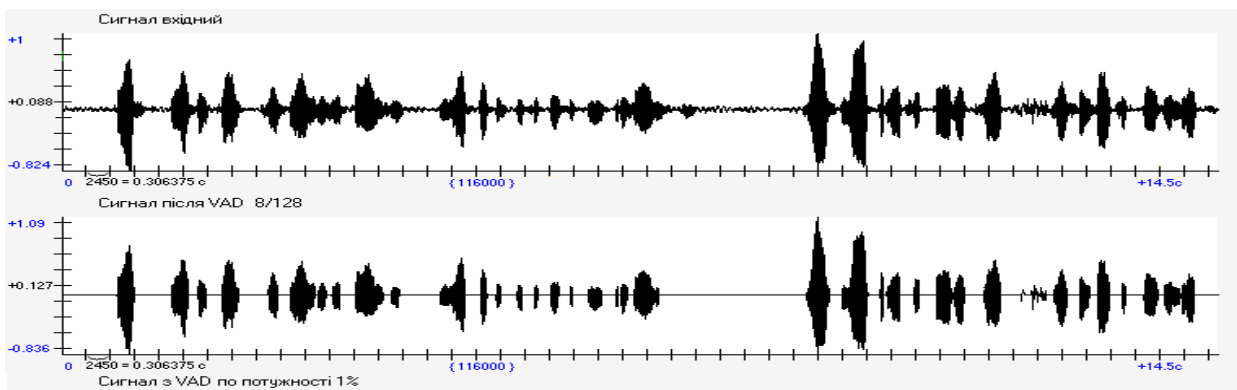


Рис. 3. Дослідження роботи VAD: а – вхідний сигнал; б – вихідний сигнал з VAD по рівню 1/16

Для оцінки якості відтворення мовного сигналу використано запроповану об'єктивну оцінку MSE (рис. 4). Також реалізована можливість прослуховувати вихідний файл, тобто суб'єктивно оцінити отриманий мовний сигнал на слух.

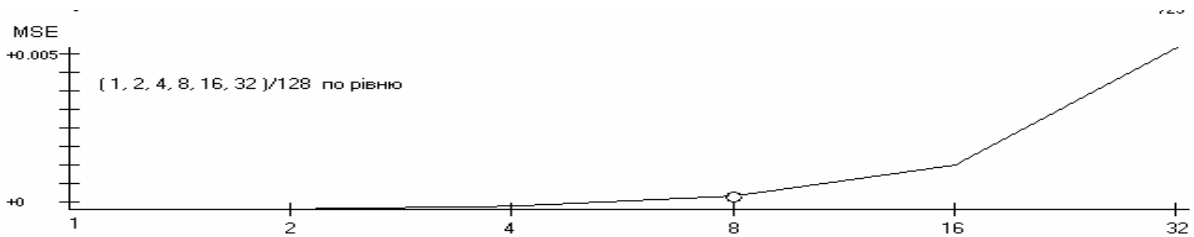


Рис. 4. Дослідження роботи VAD MSE вихідного сигналу з VAD за рівнем

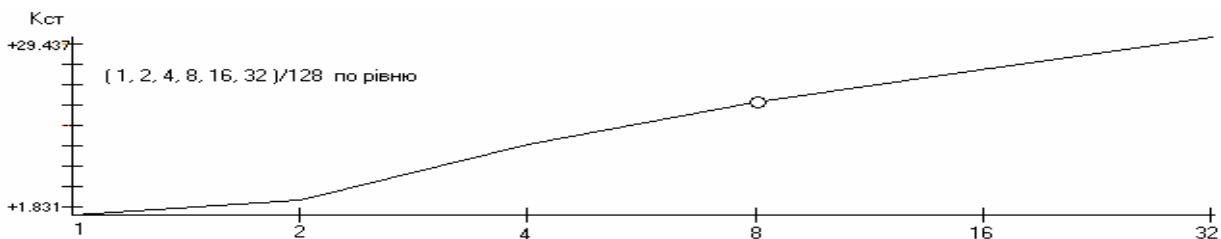


Рис. 5. Коефіцієнт стиску методу з VAD за рівнем

Врахування стану детектора VAD під час кодування мовного потоку даних приводить до створення змінної швидкості генерування пакетів на виході кодера, що відповідає особливостям навантаження мережі, та підвищує ефективність використання пропускну здатності мережі з використанням IP-технологій. Також застосування алгоритму VAD істотно підвищує ефективність компресії мовного потоку. Така залежність запропонованого алгоритму продемонстрована на рис. 5.

### Висновки

Отримання високої якості зв'язку в мережах з комутацією пакетів вимагає використання складних алгоритмів компресії мовних сигналів, що мають незначне навантаження та динамічний розподіл біт під час кодування відліків і завадостійке кодування найважливішої частини інформації. Основними характеристиками цих алгоритмів є можливість компресії джерела повідомлення, що дає змогу підвищити ефективність використання телекомунікаційних каналів зв'язку. Кодери повинні орієнтуватися на реалізацію алгоритмів компресії мовних сигналів зі змінною швидкістю кодування, що відповідає трафіку IP-мереж.

Розроблений алгоритм кодування мовного сигналу та реалізована структура кодера враховує особливості роботи IP-мережі – формує змінний в часі потік даних, працює з формою сигналу, що дає можливість кодувати мовний сигнал безпосередньо з його часового представлення, при цьому метод є стійким до впливу негативних мережевих факторів і забезпечує високу якість відновлення сигналу навіть за 20 % втрат пакетів в мережі.

1. Тимченко О.В. *Методи різницевого кодування форми сигналів в системах передачі мовної інформації*. – Львів: Вид-во УАД, 2006. – 320 с. 2. Тимченко О.В., Колодій Р.С., Смолинський М.В. *Моделі і методи опису мовних сигналів в телекомунікаційному каналі // Моделювання та інформаційні технології: Зб. наук. пр. ІПМЕ НАН України*. – К., 2003. – Вип.21. – С.178–187. 3. Tymchenko O., Kolodiy R., Smolinsky M. *Wavelet Transformation Application for Multimedia Data Processing in MPEG-4 Standard // Modern problem of radio engineering, telecommunications and computer science. Proceeding of the International Conference TCSET 2004. Lviv Polytechnic National University. February 24–28, 2004. Lviv-Slavsko, Ukraine*. – P. 346–349. 4. Тимченко О., Колодій Р. *Якість кодерів мовного сигналу на основі вейвлет-перетворення для VoIP // Науково-технічна конференція молодих вчених і спеціалістів “Моделювання” / ІПМЕ НАН України: Тези конф. 13 січня 2006 р.* – К., 2006. – 37 с. – С.29–30. 5. Тимченко О., Колодій Р. *Моделі для оцінки якості IP-телефонії // XXIV Науково-технічна конференція “Моделювання” / ІПМЕ НАН України: Тези конф. 11–12 січня 2005 р.* – К.: 2005. – 46 с. – С.35–36. 6. Тимченко О.В., Хазем Аль Сайдех. *Дослідження якості методів адаптивного багатошвидкісного широкосмугового кодування мови AMR-WB в 3G // Моделювання та інформаційні технології: Зб. наук. пр. ІПМЕ НАН України*. – К., 2007. – Вип.44. – С.161–170.