



УКРАЇНА

(19) **UA** (11) **38505** (13) **U**
(51) МПК (2006)
G10L 21/00
G10L 19/00

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

видається під
відповідальність
власника
патенту

(54) СПОСІБ ДЕЛЬТА-УЩІЛЬНЕННЯ МОВНИХ СИГНАЛІВ

1

2

(21) u200810041

(22) 04.08.2008

(24) 12.01.2009

(46) 12.01.2009, Бюл.№ 1, 2009 р.

(72) ТКАЧЕНКО ОЛЕКСАНДР МИКОЛАЙОВИЧ,
УА, ФЕФЕРМАН ОЛЕГ ДМИТРОВИЧ, УА, ХРУЩАК
СЕРГІЙ ВІКТОРОВИЧ, УА

(73) ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ
УНІВЕРСИТЕТ, УА

(57) Спосіб дельта-ущільнення мовних сигналів,
що включає кодування вибірок мовного сигналу на
основі моделі лінійного прогнозування, який **відрі-**
зняється тим, що спочатку виконують попередню
структуризацію векторів у кодовій книзі згідно з

відношенням мажорювання їх відстаней до зада-
них точок відліку, далі за допомогою аналого-
цифрового перетворювача перетворюють мовний
сигнал в набір оцифрованих вибірок та кодують
фрейми отриманих вибірок у вектор коефіцієнтів
лінійного прогнозування на основі моделі лінійного
прогнозування, перетворюють коефіцієнти лінійно-
го прогнозування в вектор лінійних спектральних
пар, проводять буферизацію групи фреймів, за
алгоритмом Вітербі виконують перерахунок оціню-
вальної функції для квантованих значень фреймів,
в канал зв'язку для кожного фрейму передають
різницю між індексами в кодовій книзі від поточно-
го до попереднього квантованого значення.

Корисна модель відноситься до галузі обчис-
лювальної техніки, а саме до ущільнення мовних
сигналів, і може бути використана в засобах запи-
су, відтворення та передачі мовних сигналів, іден-
тифікації диктора та при розпізнаванні мови.

Аналогом даного способу є «Спосіб та при-
стрій для кодування аудіосигналів (варіанти)» [па-
тент України №48191 М. кл. H03M3/00, опубліко-
ваний 15.08.02, бюлетень №98041711], в якому
запропоновано спосіб та пристрій кодування звуку,
який полягає в тому, що здійснюється кодування
аудіосигналу з малою швидкістю передачі бітів та
невеликою затримкою, порівняно із затримкою, яка
виникає під час кодування сигналу високої якості
та передача першого кодованого сигналу на деко-
дер до початку передачі на цей декодер принаймні
одного кодованого сигналу, який окремо або разом
з першим кодованим сигналом створює під час
декодування сигнал вказаної високої якості.

Недоліком даного способу є великі матеріаль-
ні витрати, пов'язані з потребою використання ча-
стотних каналів з широкою смугою пропускання,
що зумовлено високою бітовою швидкістю, необ-
хідною для передачі кодованого мовного сигналу.

У [статті авторів K. Paliwal, S. Atal "Efficient Vec-
tor Quantization of LPC Parameters at 24
Bits/Frame", IEEE Transactions on Speech and Audio
Processing, Vol.1, No. 2, 1993 - pp.3-14] описаний

спосіб кодування мовного сигналу на основі моде-
лі лінійного прогнозування. Для опису сигналу ви-
користовуються лінійні спектральні частоти, які
додатково квантуються за допомогою векторних
кодових книг. Для порівняння векторів під час по-
шуку найближчого квантованого значення вектора
у кодовій книзі запропоновано використовувати
зважену евклідову міру, яка враховує різні ваги
коефіцієнтів лінійних спектральних частот.

Недоліком даного способу є вузькі функціона-
льні можливості за рахунок того, що пошук вектора
в кодовій книзі, враховуючи використання зваже-
ної евклідової міри, потребує значних обчислюва-
льних затрат, що унеможлиблює практичне засто-
сування даного способу в режимі реального часу.

За прототип обрано «Спосіб стиснення мовно-
го сигналу шляхом кодування зі змінною швидкіс-
тю, схема та пристрій для стиснення акустичного
сигналу» [патент України №43311, М. кл.
G10L21/04, G10L19/00, опублікований 17.12.01,
бюлетень №93004697], який включає операції ви-
значення рівня сигналу мовної активності для кад-
ру оцифрованих вибірок мовного сигналу, вибір
для вказаного кадру швидкості кодування із групи
швидкостей залежно від вказаного визначеного
рівня сигналу мовної активності, кодування вказа-
ного кадру відповідно до попередньо визначеного
формату кодування для вибраної швидкості, при-

(19) **UA** (11) **38505** (13) **U**

чому кожній швидкості відповідає відмінний від інших формат кодування і різні формати кодування відповідають різним наборам параметрів сигналів, що визначають оцифровані вибірки мовного сигналу відповідно до мовної моделі, та формування для даного кадру відповідного пакету даних з указаними параметрами сигналів. За основу взято вокодер, побудований на основі моделі лінійного прогнозування. При кодуванні параметрів мовного сигналу використовуються скалярні кодові книги.

Недоліком наведеного способу є значні спотворення декодованого сигналу, зумовлені кодуванням параметрів за допомогою скалярних кодових книг. За умов передавання однакових обсягів даних, використання скалярних кодових книг призводить до більш значних спотворень сигналу, порівняно з векторними кодовими книгами. Проте значні обчислювальні витрати при використанні векторних кодових книг не дозволяють використовувати їх у режимі реального часу.

В основу корисної моделі поставлено задачу створення способу ущільнення мовних сигналів, який враховує кореляцію між сусідніми фреймами мовного сигналу. За рахунок цього можливо зменшити швидкість, необхідну для передавання мовного сигналу по каналам зв'язку. Це приводить до зменшення вимог до необхідної пропускної спроможності каналу для передачі мовного сигналу, зменшення цін на послуги цифрового зв'язку, збільшення кількості абонентів систем цифрового зв'язку.

Поставлена задача досягається тим, що в способі дельта-ущільнення мовних сигналів спочатку виконують попередню структурування векторів у кодовій книзі згідно з відношенням мажорування їх відстаней до заданих точок відліку, далі за допомогою аналого-цифрового перетворювача перетворюють мовний сигнал в набір оцифрованих вибірок та кодують фрейми отриманих вибірок у вектор коефіцієнтів лінійного прогнозування на основі моделі лінійного прогнозування, перетворюють коефіцієнти лінійного прогнозування в вектор лінійних спектральних пар, проводять буферизацію групи фреймів, за алгоритмом Вітербі виконують перерахунок цінової функції для квантованих значень фреймів, в канал зв'язку для кожного фрейму передають різницю між індексами в кодовій книзі від поточного до попереднього квантованого значення.

Спосіб здійснюється наступним чином: сигнал представляється за допомогою набору параметрів: спектральної обвідної сигналу, що описується коефіцієнтами лінійного прогнозування, періоду основного тону та коефіцієнту підсилення. Оцінка параметрів виконується для кожного фрейму звукового сигналу, довжина яких становить 20мс.

Коефіцієнти лінійного прогнозування (LPC), обчислені з використанням моделі лінійного прогнозування, є параметрами передатної функції $A_p(z)$, що описує голосовий тракт людини:

$$A_p(z) = 1 + \sum_{k=1}^p a_k \cdot z^{-k}, \quad (1)$$

Надалі отримані LPC параметри перетворюються в лінійні спектральні пари (LSP), які, в свою чергу, є коренями поліномів:

$$P(z) = P_{p+1}(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}); \quad (2)$$

$$Q(z) = Q_{p+1}(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}). \quad (3)$$

Коренями даних поліномів $P(z)$ та $Q(z)$ будуть значення x_j та y_j , що лежать на одиничному колі:

$$x_j = \cos \omega_{2j-1} \pm i \sin \omega_{2j-1}; \quad (4)$$

$$y_j = \cos \theta_{2j} \pm i \sin \theta_{2j}; \quad 1 \leq j \leq p/2, \quad (5)$$

де $\{\omega_{2j-1}\}$ та $\{\theta_{2j}\}$ є наборами LSP. Перехід до лінійних спектральних пар дозволяє здійснити подальше кодування параметрів за допомогою векторних кодових книг. Векторна кодова книга містить набір векторів значень LSP, але, оскільки квантизація всього вектора LSP, довжина якого становить 10 елементів, потребує великого обсягу пам'яті та значно ускладнює пошук, використовується розбиття вектора LSP на компоненти і квантизація кожного з компонентів окремо. Таким чином, враховуючи розбиття вектору, фактично формуються окремі кодові книги для кожного підвектора.

Для подальшого скорочення обчислювальних витрат при пошуку в кодових книгах використовується їх структуризація - розбиття книги на класи згідно із заданим критерієм. Для структуризації кодових книг використовується відношенням слабого мажорування векторів. Відношення слабого мажорування виконується за таких умов: вектор Y слабо мажорує X , якщо виконується нерівність:

$$\sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i \quad k=1, 2, \dots, n, \quad (6)$$

Для структуризації кодової книги за властивістю мажорування створюється додаткова кодова книга, яка містить відстані від векторів значень LSP до заданих n точок відліку $V_0=(0, 0, \dots, 0)$, $V_1=(N_1, 0, \dots, 0)$, ..., $V_{n-1}=(0, 0, \dots, N_{n-1}, 0)$, де n - розмірність підвектора LSP-параметрів.

Оскільки LSP-параметри лежать в межах $(0, 4000)$, обрано значення $N_1=N_2= \dots = N_{n-1}=4000$.

Перехід від вектора LSP-параметрів $X=(x_1, x_2, \dots, x_n)$ до вектора відстаней $X'=(x'_1, x'_2, \dots, x'_n)$ задається формулами:

$$x'_1 = D(X, V_0), \dots, x'_n = D(X, V_{n-1}), \quad (7)$$

де D - евклідова відстань між векторами, що обчислюється за формулою:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (8)$$

Ущільнення LSP-параметрів відбувається за рахунок передачі в канал зв'язку відстані між індексами в кодовій книзі для сусідніх фреймів (дельта-кодування). Основна ідея дельта-ущільнення базується на тому, що квантоване значення наступного за даним вектору LSP-параметрів, як правило, лежить на тому ж чи сусідніх рівнях, за рахунок чого є можливим кодувати вектори як відстані між індексами в структурованій кодовій книзі. Різниця між індексами визначається розміром вікна та є значно меншою за загальну кількість векторів, що дозволяє кодувати її меншою кількістю бітів.

За рахунок введення затримки в кілька фреймів розглядається не окремий фрейм, а їх сукупність, що надає змогу додатково корегувати значення індексів для підвищення якості вихідного мовного сигналу. Кодування групи фреймів відбувається за допомогою алгоритму Вітербі. Алгоритм Вітербі забезпечує пошук шляху, цінова функція для якого є мінімальною. Цінова функція для рівня l на n -му фреймі D_{ln} обчислюється за формулою

$$D_{ln} = \sum_{j=n-k}^n D_{ij} \quad (9)$$

де $n=1,2,\dots, N$ - номер фрейма в мовному сигналі,

k - максимально припустима затримка (у фреймах),

$D_{ij}=D(X_i, X_j)$ - відстань від вектора X_i в кодівій книзі до вхідного вектора X_j на j -му фреймі, що обчислюється за формулою (10).

$$d(X_i, X_j) = \sum_{k=1}^{10} [c_k w_k (X_i[k] - X_j[k])]^2 \quad (10)$$

де c_k - коефіцієнт зваження, який враховує різну вагу елементів залежно від позиції,

w_k - коефіцієнт зваження елементу, що базується на значенні спектральної потужності сигналу.

Реалізація алгоритму Вітербі враховує розподіл векторів за рівнями мажоризації, тому цінова функція будується не для кожного вектора у кодівій книзі, а для кожного рівня мажоризації. Оскільки кількість рівнів мажоризації значно менше за кількість векторів, це дозволяє суттєво заощадити ресурси пам'яті.

У роботі алгоритму можна виділити стадії ініціалізації та виконання. На стадії ініціалізації утворюються структури даних, необхідні для програми та визначається стартовий рівень, для якого відстань від початкового вектора - мінімальна.

Виконання алгоритму зводиться до перерахунку значення відстані та знаходження мінімального значення її для кожного рівня після отримання чергового вхідного вектора. Далі аналізуються рівні, які знаходяться у вікні пошуку і за формулою (9) обчислюється сумарна відстань до даного вектора за k фреймів, де k - затримка у фреймах. При цьому для кожного рівня знаходиться та зберігається в списку траєкторій лише той шлях, для якого відстань мінімальна.

У канал зв'язку передається індекс "кращого" вектора для даного рівня. При цьому абсолютне значення цього індексу замінюється на "відносне", яке обчислюється як різниця між абсолютним значенням та індексом початку вікна, з якого даний вектор є доступним.