

**Андрусенко В.В.**

*студент,*

*Харківський національний університет радіоелектроніки*

## **РОЗПІЗНАВАННЯ МОВИ. ЗАСТОСУВАННЯ ПЕРЦЕПЦІЙНОЇ ТРАНСФОРМАЦІЇ ДО МЕТОДУ КОДУВАННЯ З ЛІНІЙНИМ ПЕРЕДБАЧЕННЯМ**

Виділення інформаційних ознак [1] вважається ключовим етапом процесу автоматичного розпізнавання мови, який безпосередньо відповідає за відокремлення одного звукового сигналу від іншого. Головною метою даної складової є обчислення акустичних векторів, які по суті є значно більш компактною репрезентацією вхідних даних системи. Таким чином, процес виділення інформаційних ознак мови відповідає за видалення непотрібної інформації та збереження найбільш важливої з точки зору дикторонезалежних систем розпізнавання. Він включає в себе перетворення звукових сигналів у цифрову форму, а також вимірювання таких важливих складових сигналу, як енергія та частота. Одними з найбільш поширених, серед сучасних систем автоматичного розпізнавання мови, методів виділення інформаційних ознак є мел-частотні кепстральні коефіцієнти, кодування з лінійним передбаченням, а також їх гібрид – перцепційне лінійне передбачення. У межах даної роботи буде більш детально розглянуто методи на базі лінійного передбачення.

Кодування з лінійним передбаченням [2] вважається найбільш ефективним методом для виділення базових параметрів мови. У своїй оригінальній формі цей метод частіше за все використовується для декодування мови високої якості із низькою бітовою швидкістю. Ідея, що лежить в основі даного методу, полягає у тому, що будь-яке наступне значення звукового сигналу мови може бути розраховано на базі певної кількості попередніх значень сигналу. Кількість цих значень називається порядком прогностичної моделі і вважається ключовим параметром процесу виділення інформаційних ознак даним методом. Таким чином, даний метод базується на припущенні відносно того, що окреме значення сигналу залежить від попередніх значень, у наслідок чого даний алгоритм досить чуттєвий до такої невід'ємної складової мови, як вигуки. Описана прогностична модель може бути записана наступним чином:

$$\sum_{i=0}^{p-1} a_i x(p-i) = x(p+1), \quad (1)$$

де  $p$  – порядок прогностичної моделі;

$x(j)$  – окреме значення звукового сигналу мови;

$a_i$  – коефіцієнт лінійного передбачення.

Саме з коефіцієнтів лінійного передбачення будується вихідний акустичний вектор. Дані коефіцієнти описують форманти звуків – акустичні характеристики звуків мови, що пов'язані із рівнем частоти голосового тону.

Послідовність отримання коефіцієнтів лінійного передбачення зображено на рисунку 1.



**Рис. 1. Обчислення коефіцієнтів лінійного передбачення**

Таким чином, спочатку вхідний сигнал б'ється на фрейми, до кожного з яких застосовується віконна функція, зазвичай – вікно Хемінга. Далі виконується автокореляція отриманих даних наступним чином:

$$r(m) = \sum_{n=0}^{N-1-m} x(n) * x(n+m), \quad (2)$$

$$m = 0, 1, \dots, p,$$

де  $N$  – довжина фрейму;

$x(j)$  – значення фрейму з індексом  $j$ ;

$p$  – порядок прогностичної моделі.

Останній крок алгоритму – безпосередньо отримання коефіцієнтів лінійного передбачення, яке зводиться до застосування методу Левінсона-Дурбіна до наступного рівняння:

$$Xa = p, \quad (3)$$

де  $X$  – матриця усіх можливих послідовностей значень сигналу довжиною, що відповідає порядку моделі;

$a$  – вектор коефіцієнтів лінійного передбачення;

$p$  – вектор значень, що необхідно передбачити.

Отже, до переваг методу кодування з лінійним передбаченням можна віднести його точність та надійність у виділенні параметрів мови. Окрім цього, для даного методу характерним є достатньо високий рівень швидкості обчислення. Метод лінійного передбачення є надзвичайно корисним під час кодування мови, записаної із низькою бітовою швидкістю. В той самий час до недоліків даного методу можна віднести складність відокремлення схожих звуків, складність представлення мови через припущення щодо стаціонарності вхідних сигналів, а також втрату певних даних сигналу через існуючу під час обчислення коефіцієнтів передбачення похибку. Зважаючи на існуючі переваги та недоліки методу, зовсім не дивним є той факт, що у продовж свого існування він неодноразово вдосконалювався. В результаті цього процесу з'явився такий метод виділення інформаційних ознак, як перцепційне лінійне передбачення [3]. Головною відмінністю даного методу від оригінального є обробка отриманих спектральних даних сигналу таким чином, щоб вони відповідали ключовим характеристикам слухової системи людини. Подібно до мел-частотних кепстральних коефіцієнтів, даний ефект досягається завдяки застосуванню цифрового фільтру, однак, на відміну від мел-коефіцієнтів, для даного методу його основою виступає барк-шкала. Так само, як і мел, барк –

психофізична одиниця висоти звуку, що в свою чергу базується на критичних полосах слуху. Окрім цього, метод перцепційного лінійного передбачення супроводжується додатковою обробкою сигналу з метою зменшення чуттєвості до зовнішніх шумів, а також максимального наближення відповідності обчисленої моделі до слухової системи людини. Вихідними даними алгоритму є кепстральні коефіцієнти, обчислені на базі коефіцієнтів лінійного передбачення. Послідовність кроків, що включає в себе перцепційне лінійне передбачення, зображено на рисунку 2.



**Рис. 2. Обчислення кепстральних коефіцієнтів лінійного передбачення**

Першою відмінністю від оригінального алгоритму є фільтрація відносно барк-шкали [4]. Герци можна перевести у барки наступним чином:

$$b = 6 * \ln \frac{f}{600} + \sqrt{\left(\frac{f}{600}\right)^2 + 1}. \quad (4)$$

Форма барк-фільтру визначається наступним рівнянням:

$$\psi = \begin{cases} 0 & f_{bark} - f_{c(bark)} < -2.5 \\ 10^{(f_{bark} - f_{c(bark)}) + 0.5} & -2.5 \leq f_{bark} - f_{c(bark)} \leq -0.5 \\ 1 & -0.5 < f_{bark} - f_{c(bark)} < 0.5 \\ 10^{-2.5 * (f_{bark} - f_{c(bark)}) - 0.5} & 0.5 \leq f_{bark} - f_{c(bark)} \leq 1.3 \\ 0 & f_{bark} - f_{c(bark)} > 1.3 \end{cases} \quad (5)$$

де  $f_{c(bark)}$  – центральна частота окремого фільтру барк-шкали;

$\psi$  – ваговий коефіцієнт фільтру.

Отримані фільтри застосовуються наступним чином:

$$X_m = \sum_{k=0}^{\frac{N}{2}-1} |S[k]|^2 |\psi_m[k]|, \quad (6)$$

$$2 \leq m \leq M - 1,$$

де  $X_m$  – дані, до яких застосовано фільтр із індексом  $m$ ;

$|S[k]|^2$  – енергія спектра фрейму у певній точці;

$|\psi_m[k]|$  – ваговий коефіцієнт фільтру.

Наступним кроком виконується вирівнювання даних відносно чуттєвості слуху людини до сприйняття звуку різних частот:

$$X_{m(e)} = E_m X_m$$

$$E_m(f) = \left( \frac{f^2}{f^2 + 1.6 * 10^5} \right)^2 \frac{f^2 + 1.44 * 10^6}{f^2 + 9.61 * 10^6}, \quad (7)$$

де  $f$  – центральна частота фільтру з індексом  $m$ .

Далі виконується стиснення інтенсивності звуку за законом Стівенса:

$$\Phi_m = X_{m(e)}^{\frac{1}{3}}. \quad (8)$$

Останній крок алгоритму – отримання кепстальних коефіцієнтів:

$$C(n) = -a(n) - \frac{1}{n} \sum_{k=1}^{n-1} (n-k) * a(k) * C(n-k). \quad (9)$$

Таким чином, шляхом застосування перцепційної трансформації, було отримано більш ефективні, з точки зору автоматичного розпізнавання, параметри мови, що враховують особливості слуху людини та є більш стійкими по відношенню до зовнішніх шумів.

### Список використаних джерел:

1. Urmila Shrawankar, Dr. Vilas Thakare «Techniques for Feature Extraction in Speech Recognition System: A Comparative Study», International Journal of Computer Applications in Engineering, Technology and Sciences, pp. 412-418, 2013.
2. Namrata Dave «Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition», International journal for advance research in engineering and technology, vol. 1, no. 6, 2013.
3. Veton Z. Kepuska, Hussien A. Elharati «Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions», Journal of Computer and Communications, vol. 3, no. 6, 2015.
4. Ben J. Shannon, Kuldip K. Paliwal «A Comparative Study of Filter Bank Spacing for Speech Recognition», Microelectronic engineering research conference, 2003.