

УДК 621.372. 542: 376.56

Тимченко О.В., Мохаммед Абул Кашем*

Українська академія друкарства, ДУ “Львівська політехніка”*

АНАЛІЗ ЯКОСТІ ПЕРЕДАЧІ СИГНАЛУ МОВИ ВОКОДЕРНОЮ СИСТЕМОЮ

© Тимченко О.В., Мохаммед Абул Кашем*, 2000

Аналізується відношення сигнал-шум та кількість частотних піддіапазонів розбиття спектра сигналу мови вокодерної системи з метою отримання мінімальної швидкості передачі сигналу.

Для дослідження якості систем обробки та зміни часового масштабу мовного сигналу застосовують різні його моделі [1]. Найбільш точна модель мовного сигналу представляє собою нестационарний гауссовий процес, в якому повільно змінюються дисперсія та спектральна густина. Використання такої моделі дозволяє отримати найбільш точний результат при оцінці параметрів будь-яких систем обробки мовного сигналу. Але параметри такої моделі значною мірою залежать від індивідуальних особливостей абонента, таких як швидкість мовлення, емоційний стан, жіночий чи чоловічий голос, гучність та інші. В зв'язку з цим через неможливість визначення більшості параметрів модель мовного сигналу у вигляді нестационарного гауссового процесу практично не застосовується.

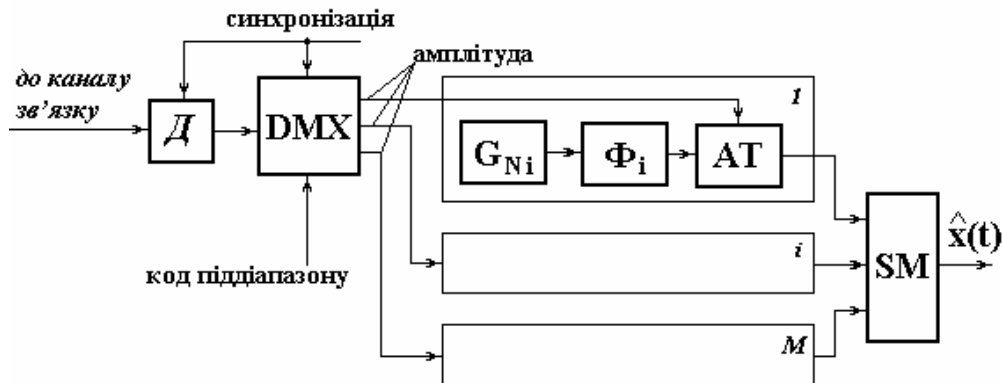
Застосовується модель опису мовного сигналу як нестационарного гауссового процесу з повільно змінною дисперсією та постійною густиною спектра потужності, визначених на основі часового усереднення експериментальних даних. При достатньо великих інтервалах реалізації процес стає близьким до стаціонарного, що власне і дає підставу вважати мовний сигнал квазістаціонарним [2].

Найбільш часто для оцінки параметрів систем обробки сигналу мови застосовується більш проста для дослідження та моделювання стаціонарна модель з постійною дисперсією σ_x^2 і усередненим значенням густини спектра потужності $S_x(\omega)$. Застосування такої моделі дозволяє отримати результати розрахунків, які збігаються з експериментальними даними з достатньою для інженерних розрахунків точністю. Зауважимо, що застосування попередніх нестационарних моделей для аналізу систем обробки сигналів мови практично неможливо здійснити аналітичними методами через відсутність загальних методів аналізу нестационарних сигналів.

Розглянемо застосування останньої моделі для оцінки якості відновленого на приймальній стороні мовного сигналу. Оцінку якості проведемо визначенням співвідношення сигнал-шум.

У вокодерних системах передачі для зменшення обсягу даних, що передаються, застосовують поділ спектру $S_x(\omega)$ центрованого випадкового процесу $x(t)$ на смуги і формування в кожній такій смузі шумоподібного (близького до білого шуму) сигналу відповідної потужності [3]. На приймальній стороні із потужностей окремих смуг формують апроксимацію спектра мовного сигналу. На рис.1 показана спрощена структурна схема приймальної частини аналізованої вокодерної системи, в якій позначені: Д –

амплітудний детектор, DMX – аналоговий демультимплексор, G_{Ni} – генератор шуму i -го піддіапазону, Φ – фільтр, АТ – атенюатор, SM – суматор, M – вибране число піддіапазонів.



Спрощена структурна схема приймальної частини вокодера

Модель густини спектра потужності мовного сигналу подамо у вигляді

$$S_x(\omega) = \alpha \sigma_x^2 \left[\frac{1}{\alpha^2 + (\omega - \omega_0)^2} + \frac{1}{\alpha^2 + (\omega + \omega_0)^2} \right]; \quad (1)$$

$$\alpha = \frac{2\Delta\omega}{1+k^2}; \quad \omega_0 = \frac{2k\Delta\omega}{1+k^2}.$$

Тут $\Delta\omega$ – ефективна ширина спектра сигналу; ω_0 – частота основного тону; k – коефіцієнт пропорційності.

Для дослідження якості систем звичайно вибирають $\omega_0 = 300$ Гц, $\alpha = 1000$ Гц.

Після поділу спектра сигналу на смуги, які мають границі (ω_i, ω_{i+1}) , вважається, що потужність в межах кожної смуги σ_i^2 залишається незмінною за час стаціонарності сигналу (квазістаціонарна модель). Значення σ_i^2 може визначатись на середній частоті вибраної смуги, тобто значенням $S_x(\frac{\omega_i + \omega_{i+1}}{2})$. Тоді з (1) можна визначити потужність шуму на приймальній стороні у вигляді

$$\sigma_y^2 = \sigma_x^2 \sum_{i=1}^{M-1} \left[\frac{\alpha}{\alpha^2 + \left(\frac{\omega_i + \omega_{i+1}}{2} - \omega_0\right)^2} + \frac{\alpha}{\alpha^2 + \left(\frac{\omega_i + \omega_{i+1}}{2} + \omega_0\right)^2} \right] (\omega_{i+1} - \omega_i) - \arctg \frac{2\alpha\omega_{i+1}}{1 + (\omega_{i+1} - \omega_0)(\omega_{i+1} + \omega_0)} - \arctg \frac{2\alpha\omega_i}{1 + (\omega_i - \omega_0)(\omega_i + \omega_0)}, \quad (2)$$

де M – кількість смуг, а σ_x^2 в даному випадку є коефіцієнтом пропорційності.

Знайдемо відношення сигнал-шум для двох випадків розбивання діапазону мовного сигналу на 13 та на 20 смуг (табл.1) [3]. У першому випадку вибрано інтервал частот 125...3375 Гц, причому відомо, що обмеження спектра мовного сигналу до цього діапазону не призводить до помітного погіршення розбірливості, але дозволяє мінімізувати швидкість

передачі сигналу. Ширина кожної смуги вибрана 250Гц. У другому випадку вибраний інтервал частот 100-10000Гц і розбиття проведено за рівноартикуляційною ознакою з однаковою часткою формант в кожній смузі $\Delta D = 0.05$. У цьому випадку смуги мають різну ширину (табл.1).

Таблиця 1

| № смуги | Частотний діапазон 125...3375Гц | | Частотний діапазон 100...10000Гц | |
|---------|------------------------------------|---------------------|-------------------------------------|------------------|
| | Граничні частоти, Гц | Смуга частот, Гц | Граничні частоти, Гц | Смуга частот, Гц |
| 1 | 125...375 | 250 | 100...420 | 320 |
| 2 | 375..625 | 250 | 420...570 | 150 |
| 3 | 625...875 | 250 | 570...710 | 140 |
| 4 | 875...1125 | 250 | 710...865 | 155 |
| 5 | 1125...1375 | 250 | 865...1030 | 165 |
| 6 | 1375...1625 | 250 | 1030...1220 | 190 |
| 7 | 1625...1875 | 250 | 1220...1410 | 190 |
| 8 | 1875...2125 | 250 | 1410...1600 | 190 |
| 9 | 2125...2375 | 250 | 1600...1780 | 180 |
| 10 | 2375...2625 | 250 | 1780...1960 | 180 |
| 11 | 2625...2875 | 250 | 1960...2140 | 180 |
| 12 | 2875...3125 | 250 | 2140...2320 | 180 |
| 13 | 3125...3375 | 250 | 2320...2550 | 230 |
| 14 | - | - | 2550...2900 | 350 |
| 15 | - | - | 2900...3300 | 400 |
| 16 | - | - | 3300...3660 | 360 |
| 17 | - | - | 3660...4050 | 390 |
| 18 | - | - | 4050...5010 | 960 |
| 19 | - | - | 5010...7250 | 2240 |
| 20 | - | - | 7250...10000 | 2750 |

Для знаходження співвідношення сигнал-шум в кожній смузі використаємо формулу (2). Результати розрахунків зведемо в таблицю (табл.2).

Таблиця 2

| Число смуг | Частотний діапазон мовного сигналу | Співвідношення сигнал-шум | Відношення σ_i^2 / σ_x^2 |
|------------|------------------------------------|---------------------------|--------------------------------------|
| 13 | 125...3375 | 10.30 | 0.093 |
| 20 | 100...10000 | 10.27 | 0.094 |

Аналіз даних табл.2 дозволяє зробити висновок, що обидві вокодерні системи з вибраними методами розбиття піддіапазонів є еквівалентними, забезпечують приблизно однакову якість відтвореного сигналу і легко реалізуються (див. рис.) у вигляді ВІС.

1. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів, 1997.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: Пер. с англ. — М., 1989.
3. ССІТТ Volum V – Supplement №13.