

ДОСЛІДЖЕННЯ ХАРАКТЕРИСТИК МОВНОГО СИГНАЛУ В ЗАДАЧАХ РОЗПІЗНАВАННЯ

© Фаді Аль-Келані, 2004

Розглянуто імовірнісні і статистичні параметри мовних сигналів, розраховані усереднені спектральна густина і кореляційна функція, наведені і обчислені критерії оцінок якості компресії.

In this paper probabilistic and statistical parameters of linguistic signals, average spectral density and correlation function are calculated, criteria of estimations of compression quality are carried out.

Вступ. Джерелом акустичної мовної хвилі є артикуляційний апарат людини, який при проходженні через нього повітря утворює звукові коливання. Якщо сказати інакше, то утворюються звуки людської мови. Для полегшення розуміння подальшого матеріалу приведемо декілька визначень частотних параметрів, що властиві лише для мовного сигналу:

Голосовий тракт – гортань і ротова порожнина.

В акустиці голосовий тракт розглядається як система резонаторів, характеристики яких повільно змінюються в часі. Частоти і області резонансів називаються відповідно формантними частотами і областями, якщо ж скорочено, то формантами.

Піки в частотній області, що виникають через квазіперіодичні коливання голосових зв'язок, називають періодом основного тону.

Звуки, при формуванні яких голосові зв'язки роблять коливальні рухи, називають вокалізованими. Вони мають в частотній області чітко виражений період основного тону ('а', 'о', 'е'...). Усі решта звуків можна віднести до невокалізованих. Точніше до останніх належать фрикативні звуки, що виникають у разі утворення турбулентного широкосмугового шуму ('с', 'ш'...), і вибухові, які формуються утворенням в тракті перепони, на шляху повітря, з подальшим вивільненням стиснутого в області за перепорою повітря ('ч', 'п', 'к'...).

Такими параметрами, як період основного тону, форманти, вокалізованість і невокалізованість можна лише поверхнево охарактеризувати тип сигналу, оскільки артикуляційний апарат є нелінійною параметричною системою [1], яка не піддається точному математичному опису.

Тому в роботі будуть розглянуті імовірнісні і статистичні параметри мовних сигналів, розраховані усереднені спектральна густина і нормована кореляційна функція. Крім того, будуть наведені і розраховані критерії оцінки якості компресії (об'єктивні і суб'єктивні) мовних сигналів, що описуються частотними параметрами.

Як правило, мовний сигнал є нестационарним випадковим процесом [1]. Тому оцінка його таких імовірнісних характеристик, як функція імовірності, спектральна густина, кореляційна функція і інші, не можлива лише на основі усереднення в часі однієї реалізації. Таке усереднення можна виконати лише на коротких інтервалах часу, де характеристики можна вважати постійними. Однак при проектуванні систем передачі, в яких властивість нестационарності не використовується, можна вважати сигнал стаціонарним та ергодичним і тоді операція усереднення за великий інтервал часу набуває змісту оцінки характеристик близького до мовного сигналу стаціонарного випадкового процесу. У результаті одержуються усереднені за часом статистичні характеристики мовних коливань.

Імовірнісні і статистичні параметри мовних сигналів. Проведені статистичні дослідження показали, що найпростішою апроксимацією функції густини імовірності мовного сигналу є закон Лапласа [2]

$$p(x) = \frac{1}{\sqrt{2} \cdot G_x} \cdot \exp\left(-\frac{\sqrt{2} \cdot |x|}{G_x}\right). \quad (1)$$

Більш точним є Гамма розподіл [2]

$$p(x) = \left(\frac{\sqrt{3}}{8\pi G_x \cdot |x|}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{\sqrt{3}}{2G_x} \cdot |x|\right), \quad (2)$$

де G_x – середньоквадратичне відхилення мовного сигналу.

Як бачимо з рис. 1, функція Гамма розподілу, якщо аргумент дорівнює нулю, прямує до нескінченності. Ця властивість вказує на наявність у мовному сигналі великої кількості пауз, в яких значення сигналу прямує до нуля.

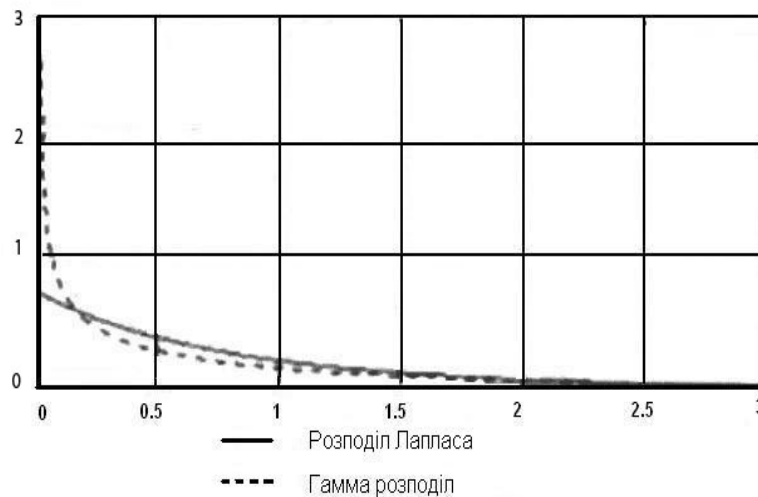


Рис. 1. Апроксимовані функції густини імовірності для мовного сигналу

Для усередненої спектральної густини потужності $W(f)$ і нормованої кореляційної функції $\mathfrak{R}(\tau)$ [2], запишемо відповідно апроксимуючі вирази

$$W(f) = 2a \left[\frac{1}{a^2 + 4\pi^2 (f + f_0)^2} + \frac{1}{a^2 + 4\pi^2 (f - f_0)^2} \right]; \quad (3)$$

$$\mathfrak{R}(\tau) = e^{-\alpha|\tau|} \cdot \cos(2\pi f_0 \tau). \quad (4)$$

На рис. 2 і 3 наведено розраховані для мовного сигналу при $f_0 = 400$ Гц і $\alpha = 10^3$ 1/с відповідно усереднену спектральну густину потужності і усереднену нормовану кореляційну функцію.

Якщо мовний сигнал, який описується як частотними статистичними параметрами, будемо передавати по каналу зв'язку, то необхідно провести його стиснення. Нижче розглянемо критерії оцінок якості компресії мовних сигналів.

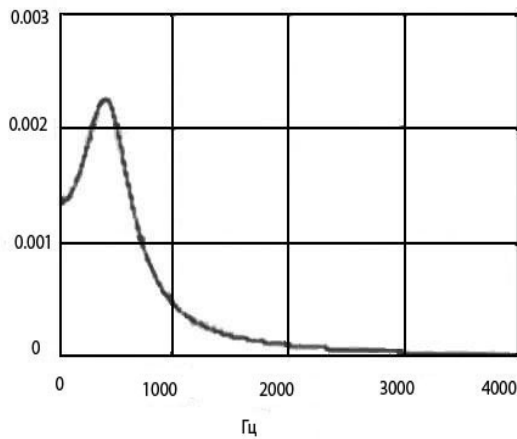


Рис. 2. Усереднена спектральна густина потужності

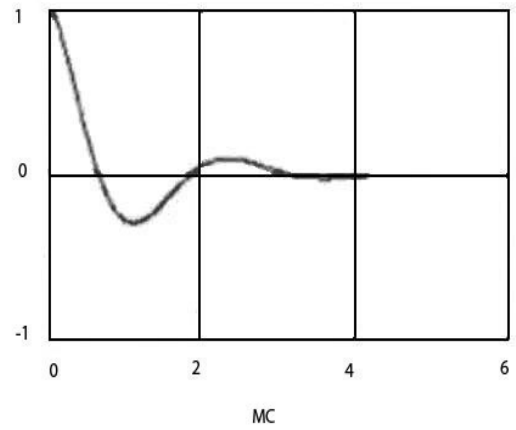


Рис. 3. Усереднена нормована кореляційна функція

Критерії оцінки якості омпресії

1. Об'єктивні критерії. Останніми роками, коли стало можливим для проведення досліджень використання ЕОМ, на перше місце, серед оцінок якості передачі, висунуті об'єктивні оцінки, які базуються на статистичних математичних параметрах сигналів для яких не потрібно значних матеріальних і часових затрат.

Викладемо основні принципи побудови таких оцінок. Позначимо через x_n і y_n мовні сигнали в дискретному часі на вході кодера і на виході декодера системи передачі.

Для оцінки якості передачі довільного сигналу використовується відношення сигнал/шум. У випадку компресії не завжди утворюється шумова похибка, тому прийmemo як критерій порівняння відношення сигнал/спотворення (с/сп), що загалом рівнозначне відношенню середньоквадратичного відхилення (СКВ) сигналу до СКВ похибки, виражене в децибелах

$$\rho_1 = 10 \cdot \lg \left[\frac{\sum x_n^2}{\sum (x_n - y_n)^2} \right]. \quad (4.1)$$

Для стандартних систем імпульсно кодової модуляції, при швидкості передачі 56–64 Кбіт/с, це відношення перевищує значення 30 дБ. Прийнято вважати, що при цьому відношенні сигнал/шум шум практично не відчутний на слух [3]. Основуючись на цьому, подальшу межу якості (значення с/сп) для телефонного зв'язку прийmemo 30 дБ.

Обробка мовного сигналу часто здійснюється сегментовано. В такому випадку відношення с/сп має вигляд [1]

$$\rho_2 = \frac{1}{N} \sum_{k=0}^{N-1} \left[10 \cdot \lg \left\{ \frac{\sum_{n=1}^N x_{n+kN}^2}{\sum_{n=1}^N (x_{n+kN} - y_{n+kN})^2} \right\} \right], \quad (4.2)$$

де N – кількість вибірок сигналу у сегменті обробки.

Для зменшення чутливості цього показника якості до зміни рівня сигналу існує ще одна формула обрахунку [1]

$$\rho_3 = \frac{1}{M} \sum_{k=0}^{M-1} \left[10 \cdot \lg \left\{ \frac{\mathfrak{R}_k^2}{1 - \mathfrak{R}_k^2} \right\} \right], \quad (4.3)$$

де M – кількість сегментів обробки,

$\mathfrak{R}_k = \mathfrak{R}_{xx}^{(k)}(0)$ – оцінка коефіцієнта взаємної кореляції $\mathfrak{R}_{xx}(0)$ на k -му сегменті, що розраховується за формулою

$$\mathfrak{R}_{xx}^{(k)}(\tau) = \frac{N \sum_{n=1}^{N-\tau} (x_{n+kN} \cdot y_{n+\tau+kN}) - \sum_{n=1}^N x_{n+kN} \sum_{n=1}^N y_{n+kN}}{\sqrt{\left[N \sum_{n=1}^N x_{n+kN}^2 - \left(\sum_{n=1}^N x_{n+kN} \right)^2 \right] \left[N \sum_{n=1}^N y_{n+kN}^2 - \left(\sum_{n=1}^N y_{n+kN} \right)^2 \right]}} \quad (4.4)$$

$\tau = 0, n=1, 2, 3, \dots, N$.

В системах передачі, що передбачають використання компресії, для зменшення надлишковості інформації у повідомленнях, можливе використання комплексного показника якості, що враховує як відношення с/сп, так і коефіцієнт компресії. Коефіцієнтом компресії вважатимемо відношення вихідної швидкості передачі інформації до вхідної кодера у тракці передачі

$$K = \frac{V_{x_n}}{V_{z_n}}, \quad (4.5)$$

де V_{x_n} і V_{z_n} – відповідно, швидкості передачі інформації на вході і на виході кодера.

Основоючись на формулах (4.1)–(4.3), можливе утворення трьох різних комплексних показників якості

$$\xi = K \mathcal{G}, \quad (4.6)$$

де \mathcal{G} – відношення сигнал/спотворення.

$$\xi_1 = K \frac{\sum x_n^2}{\sum (x_n - y_n)^2}; \quad (4.7)$$

$$\xi_2 = K \frac{1}{N} \sum_{k=0}^{N-1} \left[\frac{\sum_{n=1}^N x_{n+kN}^2}{\sum_{n=1}^N (x_{n+kN} - y_{n+kN})^2} \right]; \quad (4.8)$$

$$\xi_3 = K \frac{1}{M} \sum_{k=1}^M \left[\frac{\mathfrak{R}_k^2}{1 - \mathfrak{R}_k^2} \right]. \quad (4.9)$$

Використовуючи комплексний показник якості, можна лише об'єктивно порівняти системи передачі чи методи компресії. Для повної оцінки якості потрібне внесення і деяких суб'єктивних характеристик, що дадуть можливість врахувати властивості слухового апарату людини.

2. Суб'єктивні критерії. У цифрових системах передачі чіткість переданого сигналу, в основному, висока і в поняття “якість звучання” вкладаються такі суб'єктивні характеристики, як натуральність, можливість впізнання диктора за голосом і інше. Існують експериментальні способи визначення кількісних мір якості звучання, що оцінюється, наприклад, за градаціями “відмінно”, “добре”, “задовільно”, “недопустимо”. Кожній градації при формуванні кількісних оцінок приписується відповідна кількість балів [4].

Були проведені суб'єктивно-статистичні дослідження на основі яких стало можливим зв'язати суб'єктивні оцінки з обрахованими об'єктивними рівнісно [1]

$$\chi = \alpha \rho + \beta, \quad (4.10)$$

де χ – суб'єктивна оцінка якості, що дається слухачами і виражена в балах; α, β – коефіцієнти регресії, які знаходяться методом найменших квадратів за статистичними даними.

Якість передачі оцінювалась градаціями “недопустимо” – “відмінно” [1]. Кожній вказаній градації якості разом з проміжними типу “відмінно–добре” були приписані бали від 1 (“недопустимо”) до 9 (“відмінно”). Обробкою звучання 288 речень, вимовлених різними дикторами і одержаних в різних умовах передачі, були одержані оціночні значення $\alpha' = 0.156$ і $\beta' = 2.702$, для оцінки якості за формулою (4.1). Коефіцієнт кореляції між суб’єктивною оцінкою χ і розрахованою χ' за виразом (4.10) становив 0.667 при середньоквадратичному відхиленні 1.202. Таким чином зв’язок між χ і χ' є, але не настільки тісний, як можна було б цього очікувати.

При використанні відношення (4.2) коефіцієнти регресії склали $\alpha' = 0.247$, $\beta' = 1.369$ [1], а коефіцієнт кореляції 0.873 при середньоквадратичному відхиленні 0.787.

Для показника якості ρ_3 (4.3) $\alpha' = 0.336$, $\beta' = -0.486$, коефіцієнт кореляції 0.911, середньоквадратичне відхилення 0.665. Таким чином зв’язок цього показника з суб’єктивною оцінкою уже стає більш точним.

До суб’єктивної оцінки якості також можна віднести і спектральний склад похибки при компресії, а саме широкосмуговість чи вузькосмуговість спотворень, чи це періодичні імпульси, чи, може, гармонічні коливання. Від характеру помилки залежить чутливість людського слуху. За наявності стороннього впливу (корисного сигналу) відбувається зміна порогу чутливості слуху людини (до спотворень). Цей ефект називається ефектом маскування завад корисним сигналом. Основуючись на цьому, можна говорити про недопустимість оцінки якості передачі сигналу, базуючись лише на об’єктивних критеріях оцінки. Для повного представлення про якість системи потрібне проведення, в кінцевому порядку, суб’єктивно-статистичних досліджень із залученням спеціалістів в області акустики.

Висновки. На основі представлення мовного сигналу як ергодичного стаціонарного процесу побудовані його усереднені спектральна густина потужності, нормована кореляційна функція. Наведені як об’єктивні, так суб’єктивні критерії оцінки якості компресії.

1. Назаров М.В., Прхоров Ю.Н. *Методы цифровой обработки и передача речевых сигналов.* – М.: Радио и связь, 1985. – 176 с. 2. Рабинер Л.Р., Шафер Р.В. *Цифровая обработка речевых сигналов / Пер. с англ.; Под ред. М.В. Назарова, Ю.В. Прохорова.* – М.: Радио и связь, 1981. – 495 с. 3. Витерби А. Д., Омура Дж. К. *Принципы цифровой связи и кодирования / Пер. с англ.; Под ред. К.Ш. Зиганчирова.* – М.: Радио и связь, 1982. – 536 с. 4. Вемян Г.В. *Качество телефонной передачи и его оценка.* – М.: Связь, 1970. – 224 с.