

РОЗРОБКА МОДЕЛІ VAD В КАНАЛІ ЗВ'ЯЗКУ VOIP ІЗ ЗАСТОСУВАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

Будішевський Д.С., студент; Дегтяренко І.В., к.т.н., доц.

(Донецький національний технічний університет, м. Донецьк, Україна)

Основними специфічними проблемами при розгортанні мереж IP-телефонії стали проблеми шумозаглушення та придушення відлуння (ехопоглинання). Є також труднощі забезпечення параметрів QoS, але це проблема всієї IP-мережі та для її вирішення використовуються загальновідомі методи [1].

Для шумозаглушення використовується пристрій виявлення голосової активності (voice activity detection, VAD). VAD представляє собою апаратний або програмний модуль, який виявляє голос на вхідному акустичному каналі та відділяє його від фонового шуму або тиші. VAD також сприяє економії ресурсів каналу: якщо мовний сигнал в даний момент часу не поступає, то дані в мережу не передаються. В іншому випадку, за відсутністю VAD, в мережу передаються дані о шумі, що збільшує обсяг переданої інформації в каналі.

Область застосування детекторів мовної активності також включає пристрої для аудіо конференцій, слухові апарати, бортові системи безпеки польотів. Типова структурна схема VAD зображена на рисунку 1.

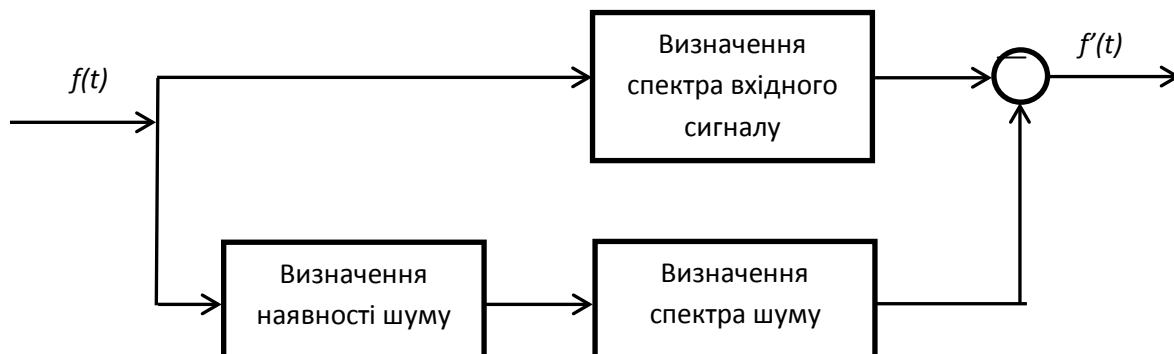


Рисунок 1 – Структурна схема VAD

На вхід системи поступає вхідний мовний сигнал з мікрофону $f(t)$. Елемент «Визначення наявності шуму» виявляє, чи є у вхідному сигналі шум. Ця процедура виконується за допомогою різноманітних алгоритмів виявлення порога шуму та мови. Прикладом є алгоритми, основані на інформаційному підході (застосування розрахунку ентропії сигналу) або статистичної моделі (постановка гіпотез о наявності шуму або мови). Після того, як шум виявлено, розраховується його спектр та відбувається віднімання спектра шуму від спектра вхідного сигналу. На виході системи отримуємо сигнал $f'(t)$, який містить тільки мову.

Загальною властивістю VAD-алгоритмів є те, що вони включають в себе навчання (обчислення характеристик шуму) і спектральне віднімання. Найчастіше в якості ознак, що визначають початок і кінець слова, обираються енергетичні та спектральні характеристики сигналу, а також число переходів через нуль [2-3].

Модель VAD з використанням вейвлет-перетворення була розроблена в середовищі matlab (Додаток А). Також використовувався wavelet toolbox даного

середовища. На вхід моделі поступає зашумлений звуковий wav сигнал з частотою дискретизації 44100 Гц – це слова «you'll see very soon», вимовлені жіночим голосом.

Тривалість сигналу – 3 секунди. Кожен фрейм містить 454 відліку (20 мс при частоті дискретизації 44100 Гц). В якості шуму в програмі використовується функція matlab awgn з варіативним рівним відношення сигнал-шум (SNR) (рис. 2)

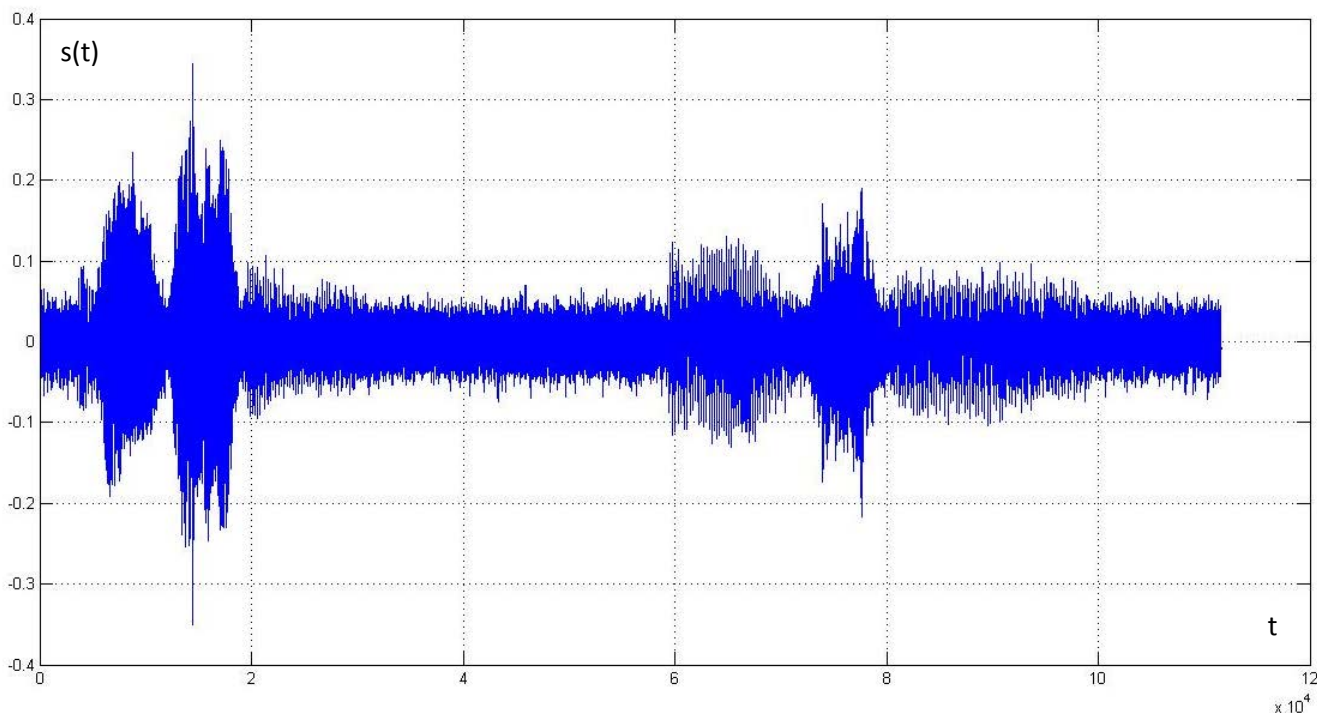


Рисунок 2 – Зашумлений вхідний сигнал (SNR=35dB)

Алгоритм роботи моделі виконується в три етапи (рис. 3)

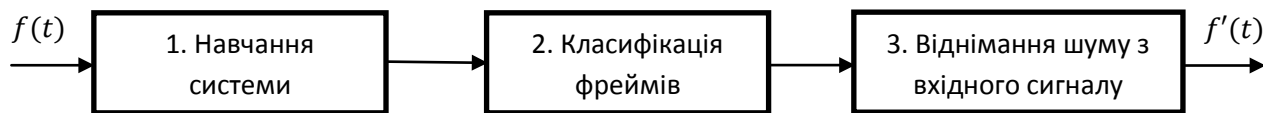


Рисунок 3 – Схема роботи моделі

На блок «Навчання системи» надходить вхідний зашумлений сигнал $f(t)$. В якості шуму використовується Гаусів білий шум. Відомо, що перші п'ять фреймів сигналу не містять мови і являють собою шум. На цьому етапі визначається середнє значення спектральної ентропії $H_{f_threshold}$ перших п'яти фреймів і позначається як граничне.

У блоці «Прийняття рішення» розраховується спектральна ентропія наступних фреймів і порівнюється за граничним значенням:

$$H_f = - \sum_{k=0}^{N-1} p(k) \log(p(k)), \quad (1)$$

де $p(k) = \frac{|X_k|}{\sum_{m=0}^{N-1} |X_m|}$ – спектральна щільність k -ої компоненти спектру, X_k – спектральні коефіцієнти БПФ, або вейвлет-коефіцієнти [4-5].

$$H_{f_threshold} \leq H_f(n) \rightarrow n \in \text{Noise}, \quad (2)$$

$$E_{f_threshold} > H_f(n) \rightarrow n \in \text{Voice}, \quad (3)$$

де n – спектральна ентропія n -го фрейму.

В блоці «Фінальна обробка» відбувається віднімання деталізуючих коефіцієнтів шуму з вхідного сигналу.

В якості критерію ефективності в роботі виступає ймовірність вірного визначення фрейму, що містить голос, при заданому рівні відношення сигнал-шум:

$$P_{voice} = \frac{N_{voice}}{N_{overall}}, \quad (4)$$

У ході експерименту змінювалось значення відношення сигнал-шум від 0 до 50 dBз кроком 5 dB. Результати дослідження зображені на рисунку 4.

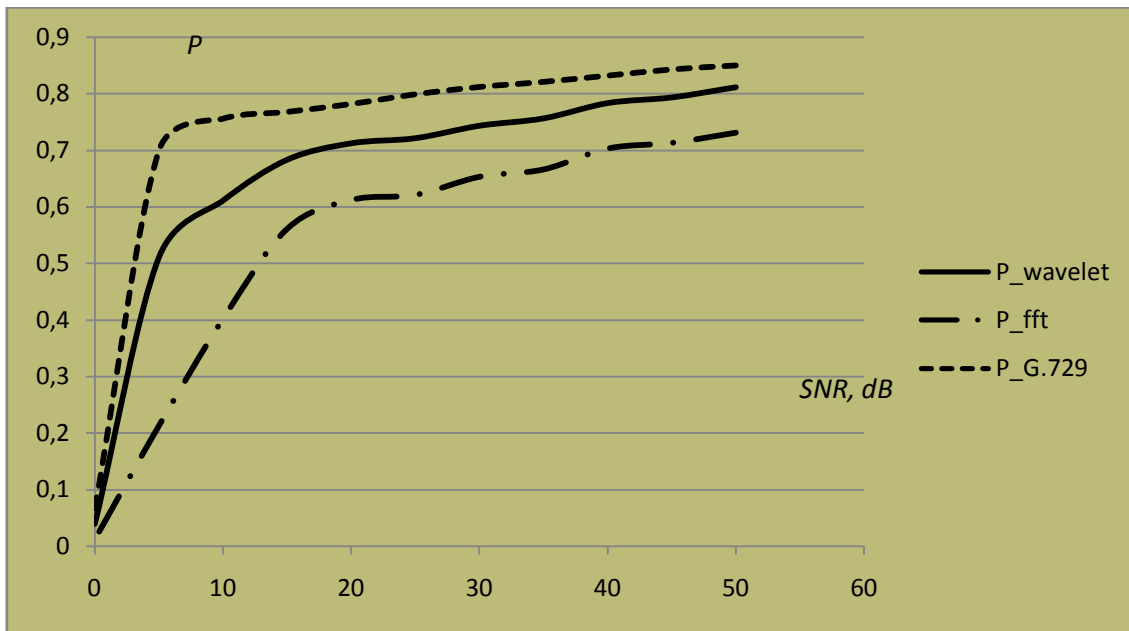


Рисунок 4 – Графік залежності ймовірності вірного визначення голосового фрейму P_{voice} від відношення сигнал-шум (SNR)

Як видно з результатів дослідження, вигрaш застосування вейвлет-перетворення замість швидкого перетворення Фур'є в ентропійному методі розпізнавання мови становить до 20%. Цей вигрaш помічається при відносно шумному сигналі – при відношенні сигнал-шум менш, ніж 15 dB. Однак, більш комплексний підхід розрахунку порога детектора G.729 дає вигрaш до 20% над запропонованою в цієї роботі системою.

Перелік посилань

1. Гольдштейн Б.С. IP-телефония./ Радио и связь. – Москва.–2001. – 335 с.
2. Kung-Ching Wang. Robust Voice Activity Detection Based on Discrete Wavelet./ Department of Information Technology & Communication. – Shin Chien University. – 2007. – 13 p.
3. R. Venkatesha Prasad, Abhjeet Sangwan, Vishal Guarav. Comparson of Voice Activity Detection Algorithms for VoIP./ CEDT. – Indian Institute of Science. – 2002. – 6 p.
4. Добеши И. Десять лекций по вейвлетам./ Регулярная и хаотическая динамика. – Ижевск. – 2001. – 464 с.
5. Смоленцев Н.К. Основы теории вейвлетов. Вейвлеты в Matlab./ - ДМК. – Москва. – 2005. – 304 с.