

АЛГОРИТМ ВИЯВЛЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ НА ОСНОВІ КОРОТКОЧАСОВИХ ХАРАКТЕРИСТИК В УМОВАХ ВИСОКОЇ ЗАШУМЛЕНОСТІ

Дюжаєв Л. П., к.т.н., доцент; Коваль В. Ю., магістрант

Національний технічний університет України

«Київський політехнічний інститут», Київ, Україна

Алгоритм виявлення голосової активності (*Voice Activity Detection, VAD*) — є одним з важливих компонентів попередньої обробки звукового сигналу в системах, що працюють з голосовими даними.

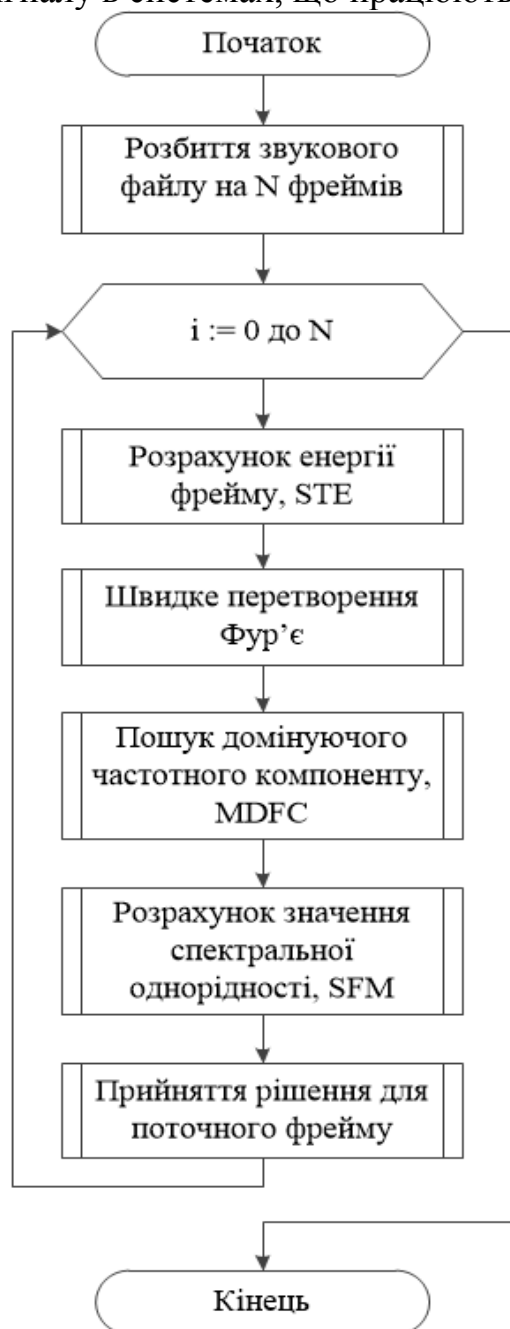


Рис. 1 Алгоритм роботи детектора

Такі алгоритми найчастіше використовуються в областях кодування та розпізнавання мови, локалізації джерела мовного сигналу, підвищення якості голосового сигналу тощо.

Згідно [1], основними характеристиками ідеального детектору голосового сигналу є надійність роботи, стійкість до шумових завад, адаптивність, простота, можливість використання в реальному часі. Досягнути стійкості до шумових завад виявилось найскладнішим завданням, оскільки в умовах низьких значень відношення сигнал/шум (*Signal-to-noise ratio, SNR*), якість роботи найпростіших VAD алгоритмів значно зменшується.

Питання якісного виділення вокалізованих ділянок голосового сигналу досліджується вченими різними напрямків. Основну увагу приділяють виділенню завадостійких ознак і вибору правил для класифікації вокалізованих ділянок звукового сигналу.

Сьогодні існує багато методів виявлення голосової активності, які відрізняються характеристиками, що класифікують вокалізовані ділянки. Найчастіше використову-

ються алгоритми на основі спектрального та кепстрального аналізу, пошуку основного тону, статистичного моделювання, використання порядкових фільтрів [2, 3]. В деяких дослідженнях [4, 5] пропонують використання комбінацій різноманітних алгоритмів, що дозволяє збільшити стійкість детектору до шумових завад, але обчислювальна складність таких методів наближається до складності самого алгоритму виявлення голосової активності.

В запропонованому методі звуковий сигнал розбивається на фрейми тривалістю 10 мс. Для кожної такої ділянки одночасно обчислюється по три різних характеристики. Перша характеристика, яка найчастіше використовується в системах виявлення голосової активності як основна — короткочасова енергія (*Short-Term Energy, STE*), оскільки на вокалізованих ділянках звукового сигналу більша частина енергії зібрана в нижніх частотах, але в той же час така характеристика стає неефективною при низьких значеннях величини *SNR*. Наступна характеристика — міра спектральної однорідності (*Spectral Flatness Measure, SFM*), яка дозволяє значно підвищити стійкість алгоритму до шумових завад різного походження, розраховується за наступною формулою:

$$SFM = 10 \log \left(\frac{G_m}{A_m} \right),$$

де G_m, A_m — відповідно середнє арифметичне та середнє геометричне спектру голосового сигналу.

Останньою характеристикою є домінуючий частотний компонент (*Most Dominant Frequency Component, MDFC*), який розраховується шляхом пошуку частоти, що відповідає максимальному значенню величини спектру. Ця характеристика в поєднанні зі згаданими вище дозволяє значно зменшити величину похибки детектування вокалізованої ділянки під час обробки звукового сигналу.

Перші десять фреймів звукового файлу використовуються для визначення початкового порогового значення характеристик звукового сигналу для подальшого прийняття рішення. Алгоритм роботи такого детектору голосової активності зображено на рис. 1. В алгоритмі, після розрахунку всіх трьох короткочасових ознак, виконується блок прийняття рішення, в якому відбувається порівняння розрахованих характеристик для поточного фрейму з пороговими значеннями і за умови якщо щонайменше дві характеристики перевищили ці межі, фрейм помічається як вокалізований.

Такий алгоритм виявлення голосової активності є одночасно простим в реалізації і може використовуватись для обробки звукових сигналів в реальному часі, забезпечуючи достатню стійкість до впливу шумових завад різного походження. До недоліків алгоритму можна віднести залежність похибки детектування від темпу мови, але його вплив можна зменшити, застосувавши більш складні алгоритми прийняття рішення.

Література

1. Savoji M. H. A robust algorithm for accurate end pointing of speech / M. H. Savoji // *Speech Communication*. — 1989 — P. 45—60.
2. Yantorno R. E. A usable speech measure employed as a co-channel detection system / R. E. Yantorno, K. L. Krishnamachari // *Proc. IEEE Int. Workshop Intell. Signal Process* — 2001.
3. K. Li, N. S. Swamy. An improved voice activity detection using higher order statistics / K. Li, Swamy N. S. // *IEEE Trans. Speech Audio Process* — №13 — 2005 — P. 965—974.
4. Wuand G. D. Word boundary detection with mel scale frequency bank in noisy environment / G. D. Wuand, C. T. Lin // *IEEE Trans. Speechand Audio Processing* — 2000.
5. Lee A. Noise robust real world spoken dialogue system using GMM based rejection of unintended inputs / A. Lee, K. Nakamura // *Interspeech* — 2004 — P. 173—176.

Анотація

Алгоритм детектування голосової активності є важливою складовою в системах обробки звукової інформації. Ефективність більшості таких алгоритмів значно знижується при наявності шумових завад. Ідеальний детектор характеризується високою надійністю роботи, стійкістю до шумових завад, простотою реалізації.

В роботі запропоновано алгоритм на основі трьох короткочасових характеристик, що дозволяє зменшити вплив шумових завад на якість детектування вокалізованих ділянок та в той же час зберегти простоту реалізації.

Ключові слова: голосова активність, короткочасова енергія, міра спектральної однорідності.

Аннотация

Алгоритм детектирования речевой активности является важной составляющей в системах обработки звуковой информации. Эффективность большинства таких алгоритмов значительно снижается при наличии шумовых помех. Идеальный детектор характеризуется высокой надежностью работы, устойчивостью к шумовым помехам, простотой реализации.

В работе предложен алгоритм на основе кратковременных характеристик, что позволяет уменьшить влияние шумовых помех на качество детектирования, в то же время сохранить простоту реализации.

Ключевые слова: речевая активность, кратковременная энергия, мера спектральной однородности.

Abstract

Voice activity detection algorithm is an important component in the audio information processing systems. Effectiveness of most of these algorithms is greatly reduced in the presence of noise. Ideal detector is characterized by high reliability work, resistant to noise interference, simplicity of implementation.

An algorithm based on short-term characteristics, reducing the impact of noise on the quality of detection; at the same time maintain the simplicity of implementation.

Keywords: voice activity detection, short-term energy, spectral flatness measure, signal-to-noise ratio.