

АДАПТИВНИЙ АЛГОРИТМ ВИЗНАЧЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ

Корнієнко О. О.; Куц С. М., к.т.н, доц.

Національний технічний університет України
«Київський політехнічний інститут», м. Київ, Україна

Важливим етапом попередньої обробки мовного сигналу (МС) у системі автентифікації мовця є визначення сегментів сигналу, що містять голосові дані.

Ефективність алгоритму визначення голосової активності (АВГА) оцінюється його надійністю, стійкістю, точністю, адаптивністю та складністю обчислень [1]. Сучасні АВГА, що використовують, наприклад, ZCR (*Zero Crossing Rate*) [2] та енергетичний метод [3], характеризуються зниженням точності виявлення сегментів голосу та тиші (СГТ) при зменшенні співвідношення сигнал/шум [4]. Для підвищення точності розпізнавання СГТ АВГА запропоновано використання метрики Махаланобіса [5] та алгоритму, оснований на розрахунках різниць спектральних компонент (*Spectral subtraction (SS)*) [6]. Недоліком модифікованих алгоритмів є необхідність проведення переналагодження класифікаторів голосу та тиші (КГТ) при зміні співвідношення сигнал/шум у каналі зв'язку [8].

Метою роботи є розробка адаптивного алгоритму формування КГТ.



Рисунок 1. Структурна схема адаптивного АВГА.

Розроблений алгоритм можна розділити на дві частини: навчання КГТ та виділення СГТ. Навчання проводиться для МС, якщо не менше ніж 10 % від загальної кількості фреймів містять голосові дані. Структурна схема адаптивного АВГА зображена на рисунку 1.

На 1-му етапі, для формування вектору ознак сегментів для кожного фрейму, проводиться розрахунок мел-частотних кепстральних (*MFCC*) та вейвлет пакетних коефіцієнтів (*WPC*). На 2-му етапі, для збільшення співвідношення сигнал/шум, використовується *SS* [7]. На третьому етапі роботи алгоритму виділяється 10 % фреймів з найбільшою та найменшою енергією, що характеризують сегменти голосу та тиші. Процес навчання класифіка-

торів голосу та тиші на вибірці векторів ознак вибраних фреймів проводиться на 4-му етапі. Критерієм закінчення процедури налаштування є не стільки мінімізація похибки навчання, скільки мінімізація помилки узагальнення для ансамблю можливих сигналів. На п'ятому етапі проводяться розрахунки щільностей умовних ймовірностей та порогового значення, на основі яких на шостому етапі, у вирішувальному пристрої, визначається належність сегменту до одного з класів.

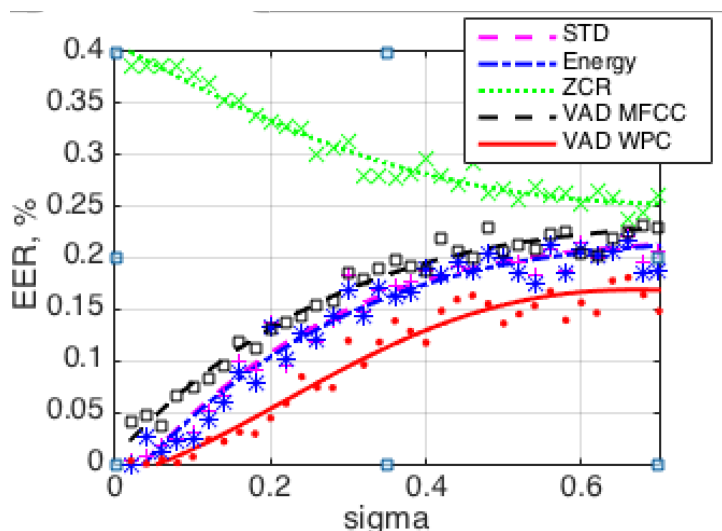


Рисунок 2. Залежність помилки EER від дисперсії АБГШ для АВГА: STD — з метрикою Махаланобіса; $Energy$ — енергетичний; ZCR — з визначенням частоти перетину нуля, $VADMFCF/WPC$ — розроблений адаптивний алгоритм з $MFCC$ та WPC відповідно.

складався з 12 $MFCC$, визначених при використанні 27 мел-частотних фільтрів. Для розрахунку WPC використовувалось дерево декомпозиції MC [8]. При формуванні КГТ було використано 32 гаусових розподіли. Точність результатів, одержаних при використанні розробленого адаптивного АВГА, порівнювалась з результатами, одержаними при тих же умовах, для АВГА з метрикою Махаланобіса (STD) [3], енергетичного методу [3] та методу ZCR [4]. Залежності коефіцієнтів, що характеризують рівність ймовірності помилок 1-ого та 2-ого роду (EER , *Error equal rate*) від дисперсії адитивного білого гаусового шуму (АБГШ), зображені на рисунку 2. Виявлено, що значення EER для алгоритму з використанням WPC в широкому діапазоні змін дисперсії гаусівського шуму менше, ніж у інших методах. Наприклад, при дисперсії АБГШ $\sigma=0,1$ для цього алгоритму $EER = 1,4$ %, а EER для адаптивного АВГА з $MFCC$ складає 7,8 %, ZCR — 36,7 %, STD — 4,6 %, а енергетичного — 4,4 %.

Розроблений алгоритм, у порівнянні з відомими методами, характеризується мінімальними значеннями EER і може бути використаний для під-

Експериментальні дослідження по визначенню точності виділення СГТ проводились на вибірці з десяти (п'яти чоловічих та п'яти жіночих) голосових записів, тривалістю 5 хв кожен, з частотою дискретизації 16 кГц. При цьому стадія попереднього навчання була завершена. Тривалість фрейму була прийнята 30 мс (480 відліків) з перекриттям 10 мс (160 відліків). Вектор ознак для фрейму MC

вищення ефективності аудіоавтентифікації мовця у системах контролю доступом.

Перелік посилань

1. Savoji M. H. A robust algorithm for accurate end pointing of speech / M. H. Savoji // *Speech Communication*. — 1989. — pp. 45–60.
2. Atal B. A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition / B. Atal, L. Rabiner // *Acoustics, Speech, and Signal Processing [IEEE Transactions on Signal Processing]*. — 1976. — Vol.24, Issue: 3. — pp. 201 - 212.
3. Kinnunen T. An overview of text-independent speaker recognition: from features to supervectors / T. Kinnunen, H. Li // *Speech Communication*. — 2010. — Vol. 52, No. 1. — pp. 12–40.
4. Greenwood M. SUVing: Automatic Silence/Unvoice/Voiced Classification of Speech / M. Greenwood, A. Kinghorn; University of Sheffield. — 1999. — UK.
5. Sahidullah M. Comparison of Speech Activity Detection Techniques for Speaker Recognition / M. Sahidullah and G. Saha // *ArXiv e-prints*. — 2012.
6. A tutorial on text-independent speaker verification / [Bimbot F., Bonastre J.-F., Fredouille C., Gravier G., Magrin-Chagnolleau I., Meignier S., Merlin T., Ortega-Garcia J., Petrovska-Delacretaz D., Reynolds D.A.]; *EURASIP Journal on Applied Signal Processing*. — 2004. — Vol. 2004, no. 4. — pp. 430–451.
7. McLaren M. A simple and effective speech activity detection algorithm for telephone and microphone speech / M. McLaren, D. van Leeuwen // *Proc. NIST SRE 2011 workshop*. — 2011. — Atlanta, US.
8. Sarikaya, R. Wavelet Packet Transform Features With Application To Speaker Identification / R Sarikaya, BL Pellom, JHL Hansen // *Third IEEE Nordic Signal Processing Symposium*. — 1998. — pp. 81-84.

Анотація

У роботі запропонований адаптивний алгоритм визначення голосової активності, який може бути використаний для підвищення ефективності системи автентифікації мовця в системах біометричного контролю доступом.

Ключові слова: Визначення голосової активності, вейвлет пакетні коефіцієнти, автентифікація мовця.

Аннотация

В работе предложен адаптивный алгоритм определения голосовой активности, который может быть использован для повышения эффективности систем аутентификации диктора в системах биометрического контроля доступа.

Ключевые слова: Определение голосовой активности, вейвлет пакетные коэффициенты, аутентификация диктора.

Abstract

Adaptive voice activity detector was proposed. It can be used for efficiency increasing of speaker recognition systems in biometric access control systems.

Keywords: Voice activity detection, wavelet packet coefficients, speaker verification.