

РОЗРОБКА СТРУКТУРИ ПІДСИСТЕМИ ІДЕНТИФІКАЦІЇ ПЕРСОНАЛУ ЗА ГОЛОСОМ В АСУ

Вінницький національний технічний університет;

Анотація

В роботі описано структуру підсистеми розпізнавання мовця за голосом, робота якої побудована на принципі резонансу гармонік тембру голоса в лінії затримки. Наведено математичне обґрунтування даного принципу розпізнавання голосу мовця.

Ключові слова: спектр сигналу мови, розпізнавання мовця, інформативні ознаки голосу, ідентифікація персоналу, автоматизовані системи управління.

Abstract

The paper describes the structure of the speaker recognition subsystem by voice, the work of which is built on the principle of resonance of harmonics of the voice timbre in the delay line. The mathematical substantiation of this principle of speaker voice recognition is given.

Keywords: speech signal spectrum, speaker recognition, informative voice features, personnel identification, automated control systems.

Вступ

Однією із задач підвищення надійності роботи автоматизованих систем управління є автоматизація розмежування доступу персоналу до різних масивів інформації. Така задача вирішується шляхом впровадження підсистем автоматичного розпізнавання диктора. Аналіз робіт з даної тематики показав, що найбільш інформативними ознаками індивідуальності голосу мовця є частота основного тону (пітч) і тембр голосу, який формується гармоніками основної частоти, які виникають внаслідок імпульсного характеру голосових низок [1,2]. Однак за умови великої кількості мовців, яких потрібно ідентифікувати за допомогою підсистеми розпізнавання голосу, виникає проблема розрізнення в частотній області сигналів основного тону і їх гармонік, оскільки діапазон частоти основного тону в силу анатомічних властивостей голосового апарату людини лежить в межах 60...350 Гц. В таких умовах виділення основної частоти і її гармонік у вузьких діапазонах затрудняється в силу неможливості побудови фільтрів з вузькою смугою пропускання з високим коефіцієнтом придушення. В даній роботі автори пропонують підсистему розпізнавання мовця, в якій проблема виділення вказаних інформативних ознак переноситься з частотної області в часову шляхом використання ліній затримки звукових хвиль гармонік основного тону. При цьому використовується кореляційний резонанс прямої і багатократно відбитої звукових хвиль в неузгодженій за навантаженням лінії затримки (ЛЗ) з потрібною кількістю виводів.

Результати дослідження

Відомо [3], що спектр вокалізованого звуку на відстані l_a від початку звукової лінії визначається рівнянням

$$F_{Sl}(j\omega, t) = \frac{1}{l_a} F_S(j\omega, t) e^{-j\omega t_3}, \quad (1)$$

де $t_3 = d\varphi(\omega) / d\omega = l_a / C_v$ - час запізнення мовного сигналу в акустичному каналі довжиною l_a .

Рівняння (1) з врахуванням теореми зміщення [3] дозволяє перейти від спектру вокалізованого звуку до його представлення у вигляді комплексної амплітуди

$$Y_{lv}(t) = \frac{1}{l_a} Y_{lv}(t - t_3),$$

звідки амплітуда резонансної хвилі, яка утворюється в результаті багатократної суперпозиції прямої і відбитих хвиль гармонік на відводах лінії затримки з часом затримки t_3 , який відповідає значенню $\frac{1}{f_{im}}$, де f_{im} частота гармоніки i -го мовця, визначиться виразом

$$|V_{kv}| = \frac{Y_{LV}(t)}{Y_{LV}(t-t_3)} = \frac{1}{l_a}$$

Таким чином, різним частотам гармонік основного тону різних мовців будуть відповідати різні комбінації резонансів. Відповідно до цього пропонується структура підсистеми розпізнавання мовців за їх голосом з використанням в якості дискримінатора лінії затримки, яка представлена на рис.1.

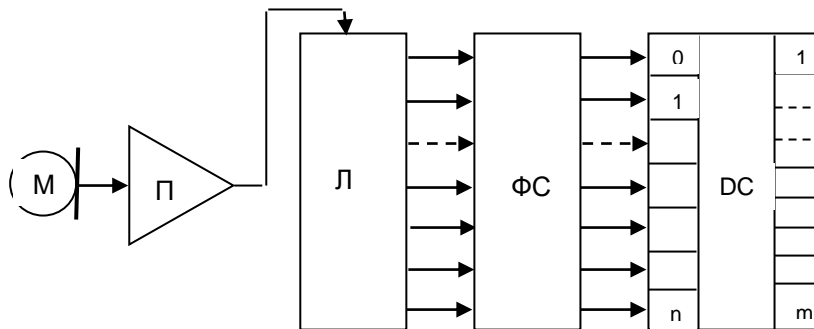


рис.1. Структурна схема підсистеми ідентифікації персоналу за голосом

Підсистема ідентифікації функціонує наступним чином: сигнал голосу мовця з мікрофона М через підсилювач П надходить на вхід лінії затримки ЛЗ, на виводах якої утворюється комбінація резонансів амплітуд, різна для різних мовців. Формувач сигналу ФС формує з них відповідні одиничні і нульові логічні сигнали $0 \dots n$, які дешифратор ДС декодує під час навчання системи в номери дикторів $1 \dots m$. Під час ідентифікації дешифратор за цими номерами звертається до таблиці пам'яті імен і ім'я розпізаного диктора видається на інформаційне табло.

Експериментальні дослідження розробленої підсистеми показали, що точність ідентифікації одного із 126 дикторів дорівнює 98,8%.

Висновки

В роботі запропоновано процедуру виділення інформативних ознак голосу мовця у вигляді гармонік тембру перенести з частотної області в часову за допомогою використання лінії затримки звукового сигналу. Наведено математичне обґрунтування реалізації дискримінації тембрів голосів шляхом використання лінії затримки, а також розроблена структурна схема для апаратної реалізації підсистеми. Експериментальні дослідження показали точність ідентифікації персоналу за голосом, достатню для сучасних вимог.

Список використаної літератури

1. Биков М.М., Ковтун В.В. Аналіз ефективності ідентифікації мовця за частотою основного тону // Вісник Хмельницького національного університету. – 2004. – №2. – Ч.1. – Т.2(60). – С. 20-23.
2. T N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P.Ouellet. Front-End Factor Analysis For Speaker Verification. IEEE Transactions on Audio, Speech, and Language Processing, 13(4):788–798, May 2011.
3. Скучик Е. Основы акустики. Т.1 / Е. Скучик. – М.: Изд-во “Мир”, 1976. – 520с.

Микола Максимович Биков — професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, e-mail: nkbykov@vntu.edu.ua.

Вадим Ігорович Дзюбенко — студент групи 2 АКІТ-176, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, м.Вінниця, e-mail: vadimdzubenko99@gmail.com.

Mykola M. Bykov — professor of Computer Control System Department, Vinnytsia National Technical University, Vinnytsia, e-mail: nkbykov@vntu.edu.ua.

Vadim I. Dziubenko - chair of computer control systems, student, Department of Computer Systems and Automation, KIV-16b group, Vinnytsia National Technical University. Vinnitsa, e-mail: vadimdzubenko99@gmail.com.