

SCI-CONF.COM.UA

**SCIENCE IN THE MODERN
WORLD: INNOVATIONS
AND CHALLENGES**



**PROCEEDINGS OF X INTERNATIONAL
SCIENTIFIC AND PRACTICAL CONFERENCE
JUNE 12-14, 2025**

**TORONTO
2025**

РОЗПІЗНАВАННЯ ВІКУ, СТАТІ ТА ЕМОЦІЙНОГО СТАНУ ЛЮДИНИ ЗА ЇЇ ГОЛОСОМ

Омельченко Сергій Васильович

к.т.н., доцент

Харківський національний університет радіоелектроніки,
м. Харків, Україна

Анотація: У статті наведено методи розпізнавання статі, віку та емоційного стану людини за її голосом в умовах дії завад. При цьому для розпізнавання в якості ознак раціонально використовувати нормовані моментні та кумулянтні функції, оцінки частоти основного тону, формантні характеристики, а також Мел-кепстральні частотні коефіцієнти (MFCC). Виконано експериментальне дослідження розроблених методів розпізнавання.

Ключові слова: розпізнавання, емоційний стан, частота основного тону, формантні частоти, MFCC.

Деякі телефонні діалогові довідкові системи потребують інформацію про вік, стать та емоційний стан користувача. Ці характеристики також можуть бути корисними для персоналізованої реклами та використовуватися як додаткові дані про користувача. Визначення віку та статі людини є також важливим у процесах аутентифікації, верифікації та розпізнавання мовлення. Наявні методи визначення віку, статі та емоційного стану не завжди гарантують високу точність в умовах впливу шумів та завад.

Розпізнавання емоцій за голосом має широкий спектр застосувань, включаючи психологічну підтримку, системи безпеки, виявлення обману, аналіз взаємодії з клієнтами та використання в відеоіграх. Оскільки людське визначення емоцій є суб'єктивним, не завжди точним і потребує значних витрат часу, постає потреба у створенні програмного забезпечення, здатного автоматизувати цей процес.

Мета цієї роботи спрямована на розробку ефективних методів розпізнавання статі, віку та емоційного стану людини за голосом в умовах дії завад. Одним із ключових аспектів є вибір надійних ознак для класифікації, таких як різномірні оцінки параметрів мовного сигналу, засновані на моделях мовотворення та слуху. Як вектор ознак доцільно для розпізнавання сумісно використовувати нормовані моментні та кумулянтні функції, оцінки частоти основного тону, формантні характеристики, а також Мел-кепстральні частотні коефіцієнти (MFCC) [1-3].

Цифровий сигнал попередньо розбивається на вибірки однакової тривалості, які з експериментальних досліджень раціонально вибирати 10-30 мс, та за якими оцінюються сукупність ознак, що змінюються в часі.

Суттєво зменшити затрати на обчислення на етапі розпізнавання можливо якщо використовувати інформативні фрагменти, в яких є корисна для розпізнавання інформація. Тому, необхідна сегментація мовного сигналу, яка включає прийняття рішення про наявність або відсутність сигналу мовлення в заданій вибірці та завдяки цьому оцінюються часові межі початку та закінчення активної мови людини.

Експерименти показали, що оцінки розподілу частоти основного тону, що характерні для людини, раціонально використовувати для розпізнавання статі, віку та емоцій диктора.

З експериментальних досліджень отримано, що оцінки частот основного тону змінювались в залежності від емоційного стану та інших факторів, та при цьому для дикторів-чоловіків як правило мали середню частоту 128 Гц з діапазоном можливих значень від 58 Гц до 238 Гц, а для жінок мали середню частоту – 256 Гц зі зміною частоти основного тону в інтервалі від 135 Гц до 522 Гц. Експериментальні дослідження мовних сигналів дозволили установити, що розподіли частот основного тону є несиметричні – щодо своєї моди: так у жіночих голосів з боку малих періодів схил крутіший, ніж для великих періодів, а у чоловіків спостерігається зворотна картина.

Для класифікації людей за віком, статтю та їх емоційним станом існують

різні методи класифікації. Зокрема раціонально використовувати класифікатори на основі моделей гауссових сумішей, та моделей сумішей Коші, регресії опорного вектора (SVR), векторної машини підтримки рішень (SVM), метода k-найближчих сусідів (kNN) та нейронних мереж.

Вік та гендерна приналежність відіграє ключову роль у системах розпізнавання емоцій, оскільки фізичні особливості голосового тракту впливають на спосіб вираження емоцій під час мовлення. З віком змінюється не лише тембр голосу, а й характерні параметри мовного сигналу, що допомагають ідентифікувати емоції. Тому застосування однієї вікової категорії сприятиме підвищенню точності класифікації. Процес передобробки включає нормалізацію та видалення шумів.

Для дослідження сумісної класифікації віку та гендерної приналежності в експериментальних дослідженнях використано три класи, що включають дорослих чоловіків, дорослих жінок та дітей без гендерної дискримінації. Для оцінювання якості розпізнавання віку та статі використано 40 зразків мови. Експериментальні дослідження показали, що точність розпізнавання статі у запропонованій системі досягла 66%. Розпізнавання віку забезпечило точність на рівні 55% при використанні виключно ознак MFCC (від 4 до 9 коефіцієнта). Використання лише формантних частот дозволило досягти 50% правильних визначень віку, тоді як точність розпізнавання статі на основі формантних частот (без урахування дітей) склала 91%. Аналіз частот двох формантів показав, що найкращі результати для розпізнавання віку було отримано при використанні першої та другої форманти, водночас найвища точність розпізнавання статі була досягнута при врахуванні всіх чотирьох формант. Комплексний підхід, що поєднує ознаки MFCC, формантні частоти, оцінки частоти основного тону та її нормовані моменти, дозволив підвищити точність розпізнавання статі до 97%.

В експериментальних дослідженнях точність розпізнавання оцінювалась при використанні в якості ознак MFCC для наступних емоцій: відраза, гнів, здивування, нейтральна, радість, смуток, спокій, страх). У таблиці 1 наведено

отримані результати оцінок точності розпізнавання кожного з 8 емоційних станів. Середній показник точності розпізнавання емоційних станів 0,89.

Таблиця 1

Точність розпізнавання окремих емоцій

Емоція	Точність
Відраза	0,93
Гнів	0,91
Здивування	0,88
Нейтральна	0,89
Радість	0,89
Смуток	0,85
Спокій	0,90
Страх	0,89

Показано ефективність створених нові методи розпізнавання статі, віку та емоційного стану людини за їх голосом, що сумісно використовують в якості ознак MFCC, кумулянтні та моментні функції оцінок формантних частот, та частоти основного тону. Експериментально підтверджено, що придушення корельованих завад в випадку їх наявності суттєво підвищує якість розпізнавання статі, віку та емоційного стану людини.

СПИСОК ЛІТЕРАТУРИ

1. Omelchenko S. Development of the method of Automatic determination of the speaker Gender on the basis of joint evaluation of frequency moments of basic tons and formant frequencies// TECHNOLOGY AUDIT AND PRODUCTION RESERVES. 2018. 3 (2(41)). P.29-33.
2. Омельченко С.В. Алгоритмы сегментации речевого сигнала на фоне коррелированной помехи//ScienceRise 4, Scientific Journal «ScienceRise». 2018. №4(45). С. 31-35.
3. Пресняков І.М., Омельченко О.В., Омельченко С.В. Автоматическое распознавание речи в каналах передачи //Радиоэлектроника и информатика. 2002. №1.С. 26-31.