

SCI-CONF.COM.UA

EUROPEAN CONGRESS OF SCIENTIFIC DISCOVERY



**PROCEEDINGS OF VII INTERNATIONAL
SCIENTIFIC AND PRACTICAL CONFERENCE
JUNE 23-25, 2025**

**MADRID
2025**

РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ ЛЮДИНИ ЗА ЇЇ ГОЛОСОМ

Омельченко Сергій Васильович

к.т.н., доцент кафедри інформаційно-мережна інженерія
Харківський національний університет радіоелектроніки,
м. Харків, Україна

Вступ. Розпізнавання та моніторинг емоцій може відігравати важливу роль у контролі психологічного здоров'я та при взаємодії людини з комп'ютером. Це є також важливим при безконтактній психодіагностиці, що включає себе оцінку стресу, депресії або емоційної лабільності. Також може використовуватись в системах віртуальних помічників, що потребує адаптації відповіді до емоційного стану користувача. Застосовується також в розумних службах підтримки, які враховують модулювання реакцій оператора в залежності від виявленої емоції. Ці характеристики також можуть бути корисними для персоналізованої реклами та використовуватися як додаткові дані про користувача. 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. Наявні методи визначення емоційного стану не завжди гарантують високу точність в умовах впливу шумів та завад.

Мета роботи. Мета цієї роботи спрямована на розробку ефективних методів розпізнавання емоційного стану людини за голосом в умовах дії завад.

Мовний сигнал попередньо розбивається на вибірки однакової тривалості, які як показано з попередніх експериментальних досліджень раціонально вибрати тривалість 10-30 мс. Такий вибір обумовлений тим, що при збільшенні вибірки проявляється нестационарність, а зменшенні збільшується дисперсія оцінки параметрів. Вибірki слугують для оцінок сукупності ознак, що змінюються в часі.

Методи розпізнавання. Попередня обробка мовного сигналу включає нормалізацію та видалення корельованих або вузькосмугових завад.

Прийняття рішення про наявність або відсутність сигналу мовлення людини в заданій вибірці дозволяє суттєво зменшити затрати необхідних обчислень на етапі оцінювання ознак та прийняття рішення.

Експериментальні дослідження показали, що оцінки розподілу частоти основного тону, що характерні для людини, раціонально використовувати для розпізнавання емоцій диктора. Фізичні особливості голосового тракту впливають на спосіб вираження емоцій під час мовлення.

Одним із ключових аспектів є вибір надійних ознак для класифікації, таких як різнорідні оцінки параметрів мовного сигналу, що базуються на моделях мовотворення та слуху. Розпізнавання емоцій за частотою основного тону – один із ключових напрямів в аналізі мовлення. Частота основного тону є акустичним маркером, що добре корелює з емоційним станом мовця. Підвищення основного тону часто свідчить про наявність вираженої емоції, як наприклад, страху, збудження або радості. Зниження частоти основного тону є типовим в випадку печалі або втоми людини. Флуктуації (нестабільність) основного тону може сигналізувати про нервозність або збудженість. Монотонність частоти основного тону асоціюється з емоційними станами, що проявляються як апатія або байдужість.

В якості вектора ознак доцільно для розпізнавання емоційного стану сумісно використовувати нормовані моментні та кумулянтні функції оцінок частоти основного тону, формантні характеристики, а також Мел-кепстральні частотні коефіцієнти (MFCC) [1-3]. MFCC добре моделюють сприйняття людського слуху: вони зменшують спектральну роздільну здатність у високих частотах, подібно до того, як це роблять наші вуха. Хоч MFCC не містять інформацію про абсолютну висоту тону (на відміну від F_0), вони несуть цінні спектральні ознаки, пов'язані з формантами, тембром і артикуляцією – усі вони змінюються під впливом емоційного стану.

Формантні частоти можна розглядати як резонансні частоти голосового

тракту людини. Найважливішими серед них є перша форманта з частотою F1, формування якої пов'язано з відкритістю рота (висота язика), друга формантна частота F2, що характеризує передньо-задню позицію язика та третя форманта F3 яка пов'язана з більш складними акустичними особливостями. При цьому емоції людини змінюють напруженість її м'язів, артикуляцію і навіть довжину вокального тракту (через зміну положення гортані), що впливає на розміщення формантів. Для різних емоційних станів відбуваються характерні зміни формантів. Для емоції радість (збудження) як правило спостерігається підняття F1 і F2 (вища артикуляційна енергія). При емоції гніву формантна частота F1 висока, але частоти формант F2 та F3 можуть змінюватись. Для емоцій сум (туга) відбувається зниження F1, та зменшення енергії в спектрі. Для емоції страх характерним є зміщення формантів через тремтіння голосу та нестабільну артикуляцію.

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

Результати досліджень. В експериментальних дослідженнях точність розпізнавання оцінювалась при використанні в якості ознаки 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. З проведених експериментальних досліджень отримано, що середній показник точності розпізнавання емоційних станів 0,89.

Таблиця 1

Точність розпізнавання окремих емоцій при використанні в якості ознак MFCC

№	Емоція	Точність
1	Відраза	0,93
2	Гнів	0,91

3	Здивування	0,88
4	Нейтральна	0,89
5	Радість	0,89
6	Смуток	0,85
7	Спокій	0,90
8	Страх	0,89

Комплексний підхід, що поєднує ознаки MFCC, формантні частоти, оцінки частоти основного тону та її нормовані моменти, дозволив підвищити точність розпізнавання емоційного стану до 93 %.

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