

АНАЛІЗ ЕМОЦІЙ З ВИКОРИСТАННЯМ ГОЛОСОВИХ ОЗНАК

Дмитрієва І.С.¹, Бімалов Д.В.²

¹Дніпровський металургійний інститут, Український державний університету науки і технологій, к.т.н., доцент, Україна

²Дніпровський металургійний інститут, Український державний університету науки і технологій, аспірант, Україна

Анотація. У роботі описані методи розпізнавання емоцій людини. Ця сфера штучного інтелекту стрімко розвивається і відіграє ключову роль у підвищенні ефективності взаємодії між людиною та комп'ютером. У процесі дослідження аналізуються наявні підходи щодо розпізнавання емоцій за участю голосових ознак. У цій роботі розглядаються методи аналізу емоцій на основі голосових ознак. Описуються основні параметри, такі як частота, темп, інтонація, тембр і гучність, які дають змогу визначити емоційний стан людини за її мовленням. Розглядаються сучасні підходи, включно з використанням алгоритмів машинного навчання та нейромережевих моделей для опрацювання аудіофайлів і класифікації емоцій. Це дослідження сприяє розвитку вивчення емоцій, надаючи глибше розуміння людських емоційних станів.

Ключові слова: розпізнавання емоцій, розпізнавання емоцій через звук, машинне навчання, аналіз емоцій, обробка мовлення.

Мета і завдання досліджень. Дослідити модель розпізнавання емоцій за допомогою голосових даних для підвищення точності та стабільності системи в умовах реальної взаємодії, що дозволить точніше інтерпретувати емоційні стани людини.

Розглянемо технологію розпізнавання емоцій за голосом, відому як паралінгвістика, що досліджує, як люди передають емоційні стани через особливості мови, такі як інтонація, темп, гучність і висота голосу. Цей процес відіграє важливу роль у психолінгвістиці, когнітивних науках і штучному інтелекті. Коли ми слухаємо когось, ми інтуїтивно розпізнаємо не лише зміст слів, а й емоції, які передаються. Наприклад, підвищений темп мовлення і високий тон можуть свідчити про радість або хвилювання. Низький і повільніший голос може вказувати на смуток чи розчарування. Приклади технологій: Emotion Recognition (SER), Neural Networks (CNN)[1]

До основних аспектів паралінгвістики, що дозволяють розпізнати емоції через голос, відносяться такі параметри мовлення:

1. Тональність голосу може передавати емоції навіть точніше, ніж слова. Високий тон зазвичай асоціюється з позитивними емоціями, як-от радість або збудження. Низький тон може натякати на сум, розчарування чи втому. Зміни в тоні (монотонність), можуть свідчити про хвилювання, здивування або інтерес.

2. Гучність голосу є ефективним показником емоційного стану, оскільки люди часто підвищують її у моменти сильних емоцій і знижують у спокійних або сумних ситуаціях.

3. Темп мовлення, тобто швидкість вимови слів, відіграє важливу роль у розпізнаванні емоцій. На емоційному фоні він часто змінюється: швидка мова може свідчити про збудження, тривогу, стрес або хвилювання; повільна — про смуток, втому, депресію або роздуми.

4. Інтенація мови це зміна висоти голосу та інших його характеристик під час мовлення, вона допомагає передати емоційні відтінки. Підйом інтонації може вказувати на здивування, хвилювання або інтерес, а спад — на завершення думки, спокій або смуток. Змінна інтонація, коли вона варіюється протягом фрази, може свідчити про збудження, радість чи інтерес.

Для аналізу та розпізнавання голосових емоцій часто використовують методи машинного навчання. Моделі класифікації, такі як логістична регресія, метод опорних векторів (SVM) та нейронні мережі, можуть бути навчені на датасеті з попередньо розміченими емоціями, наприклад аудіофайли з різними емоціями (гнів, радість, смуток тощо)[2].

Акустичні ознаки та їх ваги визначають наступним чином:

$$E = w_1 \times F0 + w_2 \times V + w_3 \times T + w_4 \times P,$$

де:

E - емоційний показник,

F0 - основна частота,

V - гучність,

T - темп мовлення,

P - тривалість пауз,

w_1, w_2, w_3, w_4 - ваги, які визначаються в процесі навчання моделі.

Кожен з цих параметрів вносить свій внесок у загальну оцінку емоції, і їхні ваги можуть бути налаштовані відповідно до навчальних даних [3]. Деталі процесу наведено у блок-схемі (див. рис. 1).

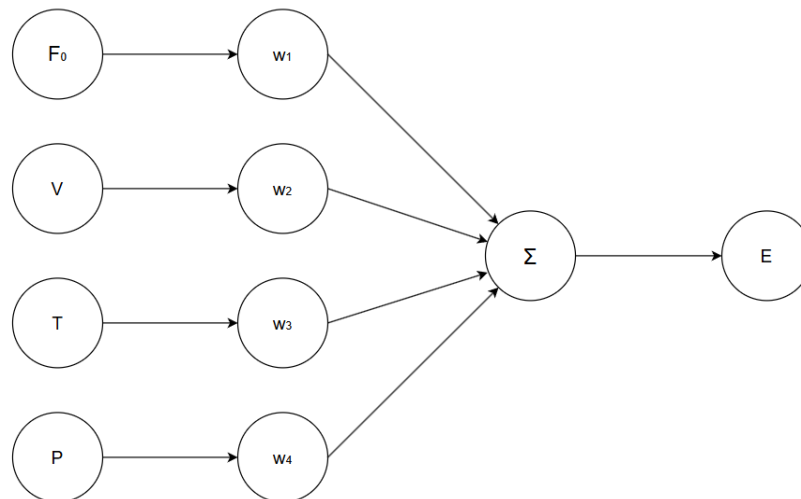


Рисунок 1 - Акустичні ознаки та їх ваги

Висновок. Визначення емоцій по голосу базується на комплексному аналізі багатьох акустичних характеристик, як-от частота, гучність, темп мовлення, інтонація та інші. Ці характеристики можуть бути проаналізовані за допомогою різноманітних математичних і статистичних моделей, включаючи машинне навчання та нейронні мережі, для точного розпізнавання емоцій.

ЛІТЕРАТУРА

1. Z. Huang, M. Dong, Q. Mao, Y. Zhan. Speech emotion recognition using CNN. Proceedings of the 22nd ACM international conference on multimedia, Association for Computing Machinery, New York, NY, USA (2014), pp. 801-804, DOI: 10.1145/2647868.2654984
2. E. Lakomkin, C. Weber, S. Magg, S. Wermter. Reusing neural speech representations for auditory emotion recognition. (2018), DOI: 10.48550/ARXIV.1803.11508.
3. Дмитрієва І.С., Бімалов Д.В. Аналіз емоцій з використанням виразів обличчя та голосових ознак. // Системні технології. Регіональний міжвузівський збірник наукових робіт. - Вип. 3(158). - Дніпро: УДУНТ, 2025. с. 21-27. DOI 10.34185/1562-9945-3-158-2025-03

ANALYSIS OF EMOTIONS USING VOICE FEATURES

Dmytriieva Iryna Serhiivna, Bimalov Dmytro Viktorovych

Abstract. *The paper describes methods for recognizing human emotions. This area of artificial intelligence is developing rapidly and plays a key role in improving the efficiency of human-computer interaction. The study analyzes existing approaches to recognizing emotions using voice features. This paper discusses methods for analyzing emotions based on voice features. The main parameters, such as frequency, tempo, intonation, timbre, and volume, are described, which make it possible to determine the emotional state of a person by his or her speech. Modern approaches are discussed, including the use of machine learning algorithms and neural network models for processing audio files and classifying emotions. This study contributes to the development of emotion research by providing a deeper understanding of human emotional states.*

Keywords: *emotion recognition, emotion recognition through sound, machine learning, emotion analysis, speech processing.*

REFERENCE

1. Z. Huang, M. Dong, Q. Mao, Y. Zhan. Speech emotion recognition using CNN. Proceedings of the 22nd ACM international conference on multimedia, Association for Computing Machinery, New York, NY, USA (2014), pp. 801-804, DOI: 10.1145/2647868.2654984
2. E. Lakomkin, C. Weber, S. Magg, S. Wermter. Reusing neural speech representations for auditory emotion recognition. (2018), DOI: 10.48550/ARXIV.1803.11508.
3. Dmitrieva I.S., Bimalov D.V. Analysis of emotions using facial expressions and voice signs. Regional inter-university collection of scientific works. - Issue 3(158). - Dnipro: USUST, 2025. p. 21-27. DOI 10.34185/1562-9945-3-158-2025-03