КІБЕРБЕЗПЕКА ТА КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ

КБКІТ-2022

науково-практична конференція
молодих вчених
аспірантів та студентів

м.Тернопіль

ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ
ГАЛИЦЬКИЙ ФАХОВИЙ КОЛЕДЖ ІМ. В. ЧОРНОВОЛА

КІБЕРБЕЗПЕКА ТА КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ (КБКІТ – 2022)

науково-практична конференція
молодих вчених, аспірантів та студентів

29–31 серпня 2022
Тернопіль

Збірник матеріалів науково-практичної конференції молодих вчених, аспірантів та студентів «Кібербезпека та комп'ютерно-інтегровані технології» (КБКІТ - 2022), Тернопіль, 2022. - 118 с.

Редактор коректор: Гуменний П.В.
Технічний редактор: Давлетова А.Я.

Адреса редакції:
Західноукраїнський національний університет, кафедра кібербезпеки,
вул. Чехова 8, м. Тернопіль 46003
Контактний телефон: (0352) 50-17-87
e-mail: kb.tneu@gmail.com

УДК 004.56.5(043.2)

Кузик В.М.[1] Продан Т.І.[2], Івасьєв С.В.[2], Слєпцова О.Я.[1]

[1]Галицький фаховий коледж імені В'ячеслава Чорновола
[2]Західноукраїнський національний університет

БІОМЕТРИЧНА СИСТЕМА АУТЕНТИФІКАЦІЇ З ВИКОРИСТАННЯМ ГОЛОСОВИХ ДАНИХ

Вступ. Ідентифікація людини за голосом є одним із біометричних систем аутентифікації. Біометрична система аутентифікації – система перевірки особистості (аутентифікації) людини за її біометричними показниками Біометричний параметр – параметр, що є частиною самої людини.

Біометричні системи аутентифікації досить зручні для користувачів. На відміну від парольних або ключових систем автентифікації, біометричні використовують параметри, які неможливо забути або втратити. Проблеми збереження автентифікаційних даних не виникає.

Мета роботи полягає у тому щоб розглянути один із способів побудови біометричної системи ідентифікації користувача за голосом, яка є незалежною від ключової парольної фрази. Основними перевагами використовуваного підходу є простота реалізації та висока надійність.

1. Біометричний метод аутентифікації за голосом

Біометричний метод аутентифікації за голосом є простим у застосуванні. Він не вимагає спеціальної апаратури, достатньо лише мікрофона та звукової плати. В даний час технологія розвивається, так як цей метод широко використовується у сучасних бізнес-центрах. Основним недоліком методу голосової автентифікації є його низька точність. Голос людини може змінюватись залежно від стану здоров'я, настрою, віку тощо. Крім того, не варто забувати про сторонні шуми. Через високу ймовірність помилок другого роду його застосовують переважно у приміщеннях середнього рівня безпеки [1]. Один із способів аутентифікації людини за голосом – це завдання системі безлічі зразків голоси однієї і тієї ж людини для порівняння з автентифікаційним ключем, що отримується в майбутньому. Далі існує безліч способів порівняння [1, 2].

Система, яку передбачається розробити, має бути незалежною від конкретної фрази. Через це може виникнути проблема з підміною голосу за допомогою звукозаписних пристроїв, наприклад, диктофонів. Однак цю проблему можна обійти за допомогою генерації простих фраз, які людина має вимовити. Не можна заздалегідь передбачити, яка фраза буде затребувана системою, що розпізнає, отже зловмисник не зможе записати ключову фразу. Такий підхід додає завдання перекладу мови до тексту, проте вона не є основною і може бути вирішена за допомогою сторонніх бібліотек [3].

На вхід системи ідентифікації надходить запис ключового повідомлення користувача. Одним із простих форматів для обробки є WAV.WaveformAudioFileFormat – формат файлу-контейнера для зберігання записи оцифрованого аудіопотоку, підвид RIFF. Цей контейнер, як правило, використовується для зберігання несжатого звуку в імпульсно-кодовий модуляції.

Отримані значення амплітуд можуть не збігатися для двох однакових записів через зовнішній шум, різні гучності вхідного сигналу і так далі. Одним із найбільш ефективних способів попередньої підготовки звуку є нормалізація [4].

Нормалізація звуку – процес вирівнювання частотних характеристик студійного звукозапису на магнітний носій. Корекція необхідна, оскільки процес намагнічування покриття плівки відбувається нерівномірно стосовно спектру аудіочастот. Якщо не проводити корекцію, навіть перше відтворення запису звучатиме несхоже на оригінал.

Існують два способи нормалізації [4]:

- пікова нормалізація – це спосіб нормалізації, за якого рівень звукового сигналу піднімається до максимально можливого значення цифрового звуку без появи спотворень. Орієнтиром служить найвищий пік амплітуд. Цей спосіб повністю виключає обмеження амплітуди сигналу (кліпінг), проте, за наявності у файлі сильно виділяється піку, то нормалізація за його рівнем може призвести до тому, що звуковий сигнал залишиться досить тихим, незважаючи на досить високу гучність оригіналу. Розмір звуку при пікової нормалізації вимірюється у відсотках;
- RMS-нормалізація - нормалізація за середньоквадратичним значенням рівня звуку у файлі. Є повною протилежністю до пікової нормалізації. При цьому способі величина звуку вимірюється у децибелах. Цей спосіб найбільш підходить для людського вуха, однак за високої гучності можливий кліпінг.

Оскільки передбачається, що людина вимовлятиме ключове вираз спокійний, то очевидна перевага у пікової нормалізації внаслідок того, що в необробленому звуку не повинно бути надто великих перепадів амплітуд. Так як розмір унікальних характеристик навіть для секундного зразка звуку величезний, то робити складні операції над такими обсягами даних неможливо. Крім того, не зовсім зрозуміло, як порівнювати об'єкти з різною кількістю унікальних характеристик.

Обчислювальну складність завдання можна зменшити, розбивши її на менш складні підзавдання. Це дозволить за допомогою встановлення фіксованого розміру підзадачі та усереднення результатів обчислень за всім завданням одержати наперед задану кількість ознак для класифікації. Як розбиття мається на увазі використання поділу звукового сигналу на звані кадри певної довжини. Кадри повинні перекривати один одного, тому що у випадку, якщо вони будуть поруч один з одним, то звук спотворюватиметься.

Для усунення небажаних ефектів при обробці кадрів кожен елемент кадру множиться на вікно. Вікно – вагова функція, яка використовується для керування ефектами, зумовленими наявністю бічних пелюсток у спектральних оцінках (розтікання спектра).

У більшості завдань цифрової обробки немає можливості досліджувати сигнал на безкінечному інтервалі. Немає можливості дізнатися, який був сигнал до увімкнення пристрою і який він буде в майбутньому. Також обмеження інтервалу дослідження може бути обумовлено нестаціонарністю досліджуваного сигналу.

Обмеження інтервалу аналізу рівносильне добутку вихідного сигналу на віконну функцію. Таким чином, результатом віконного перетворення Фур'є не спектр вихідного сигналу, а спектр твори сигналу та віконної функції. Спектр, отриманий за допомогою віконного перетворення Фур'є [5] є оцінкою спектра вихідного сигналу і принципово допускає спотворення.

Спотворення, що вносяться застосуванням вікон, визначаються розміром вікна та його формою. Виділяють дві основні властивості частотних характеристик вікон: ширина головної пелюстки та максимальний рівень бічних пелюсток. Застосування вікон, відмінних від прямокутних, обумовлене бажанням зменшити вплив бічних пелюсток за рахунок збільшення ширини головного.

Типи віконних функцій:

Прямокутне вікно

$$w(n) = \begin{cases} 1, n \in [0, N-1] \\ 0, n \notin [0, N-1] \end{cases}. \quad (1)$$

Вікно Ханна

$$w(n) = 0.5(1 - \cos(\frac{2\pi n}{N-1})). \quad (2)$$

Вікно Хемінга

$$w(n) = 0.53836 - 0.46164(\cos(\frac{2\pi n}{N-1})). \quad (3)$$

Вікно Блекмена

$$w(n) = a_0 - a_1 \cos(\frac{2\pi n}{N-1}) + a_2 \cos(\frac{4\pi n}{N-1}). \quad (4)$$

Вікно Кайзера

$$w(n) = \frac{\left| l_0(\beta \sqrt{1 - \left(\frac{2n - N + 1}{N-1} \right)^2}) \right|}{|l_0(\beta)|}. \quad (5)$$

Найбільш простий і підходящою для вирішення задачі є функція вікна Хеммінгу.

Далі потрібно отримати короткочасну спектрограму кожного кадру окремо. Для цього використовується перетворення Фур'є.

Перетворення Фур'є (Fouriertransform) - це розкладання функцій на синусоїди (далі косинусні функції теж називаємо синусоїдами, оскільки вони відрізняються від «справжніх» синусоїд тільки фазою). Існує кілька видів перетворення Фур'є [5].

1. Неперіодичний безперервний сигнал можна розкласти в інтеграл Фур'є.

2. Періодичний безперервний сигнал можна розкласти у нескінченний ряд Фур'є.

3. Неперіодичний дискретний сигнал можна розкласти на інтеграл Фур'є.

4. Періодичний дискретний сигнал можна розкласти в кінцевий ряд Фур'є.

Комп'ютер здатний працювати лише з обмеженим обсягом даних, отже, реально він здатний обчислювати лише останній вид перетворення Фур'є. Отже, використовуватиметься дискретне перетворення.

На сьогоднішній день найбільш успішними є системи розпізнавання голосу,

які використовують знання про слуховий апарат. Велике поширення при розпізнаванні людського мовлення набула mel-шкала, лінійна при частотах нижче 1кГц і логарифмічна при частотах вище 1кГц. Mel-шкала була отримана в результаті експериментів із зразковими тонами (синусоїдами) в яких з випробовуваних вимагалося розділити дані діапазони частот на 4 рівні інтервалу або налаштувати частоту необхідного тону так, щоб він був в половину частоти вихідного. 1 mel визначається як 1 тисячна рівня тону на 1 кГц. як і в будь-яких інших спробах створити подібні шкали, розраховується, що шкала mel більш точно моделює чутливість людського вуха.

Перехід до нової шкали описується нескладною залежністю:

$$m = 1127 \ln (1 + \frac{f}{700}), \tag{6}$$

де m - Частота в крейдах; f – частота у герцах. Вектор ознак складатиметься з крейдяних коефіцієнтів, що розраховуються за формулою:

$$c_n = \sum_{k=1}^{K} (\log S_k) \left[n(k - \frac{1}{2}) \frac{\pi}{K} \right], \tag{7}$$

де c_n - Крейда-кепстральний коефіцієнт під номером n; S_k - Амплітуда kго значення в кадрі в крейдах; K – наперед задану кількість дрібнепстральних коефіцієнтів.

Висновок. Останньою стадією є класифікація того, хто говорить. Класифікація проводиться обчисленням міри схожості пробних даних та вже відомих. Міра схожості виражається відстанню від вектора ознак пробного сигналу до ознак уже класифікованого вектора.

Вектор ознак представляється як середнє арифметичне векторів, що характеризують окремі кадри мови. Для підвищення точності розпізнавання просто необхідно усереднювати результати не тільки між кадрами, а й враховувати показники кількох мовних зразків. Маючи кілька записів голосу, розумно не усереднювати показники одного вектора, а провести кластеризацію з допомогою нейронних мереж.

Перелік використаних джерел.

1. Болл Р. Руководство по биометрии / Р.М. Болл, Дж.Х. Коннел; [пер. с англ. под ред. Н. Агапова]. – М.: Техносфера, 2007. – 368 с.

2. Лакин Г. Биометрия / Лакин Г.Ф. – М.: Высшая школа, 1990. – 352 с.

3. Кауненко С.І., Колесніков К.В. Методи ідентифікації людини в інформаційних системах ISDMCI'2012: Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: Материалы международной научной конференции.- Херсон-Євпаторія ХНТУ,2012, 566с., С. 164-167.

4. Precise Biometrics // [Електр. Ресурс]. – Режим доступу: https://precisebiometrics.com/

5. An emerging biometric API industry standart // [Електр. Ресурс]. – Режим доступу: http://ieeexplore.ieee.org/document/820046/ 6. BioAPI Specification Version 1.1 // [Електр. Ресурс]. – Режим доступу: http://xml.coverpages.org/BIOAPIv11.pdf