Міністерство освіти і науки України
Львівський національний університет ім. Івана Франка
Факультет електроніки та комп’ютерних технологій

Індивідуальне домашнє завдання
на тему:
"Надійність та стабільність ідентифікації за голосом"

Виконав
Студент групи ФеІ-44
Сапанюк М. І.
Перевірив
ас. Футей О. В.

Львів 2022

Вступ

Ідентифікація за голосом - це ідентифікація людини залежно від характеристик її голосу. Існує різниця між розпізнаванням диктора (розпізнавання того, хто говорить) і розпізнавання мови (розпізнавання того, що було сказано). Ці два терміни часто плутають, і «розпізнавання мови» може бути використане для обох. Крім того, існує різниця між актом аутентифікації (зазвичай він називається верифікацією диктора або аутентифікацією диктора) та ідентифікації. І, нарешті, існує різниця між розпізнаванням диктора (розпізнавання того, хто говорить) і діаризацією (розпізнавання, коли ж оратор говорить). Розпізнавання мовця може спростити завдання перекладу мови в системах, які були навчені на голоси конкретної людини або воно може бути використане для перевірки автентичності чи перевірки особистості мовця як частина процесу забезпечення безпеки. Процес розпізнавання мовця має історію, що йде назад приблизно на чотири десятиліття і використовує акустичні особливості мови з метою розрізняти окремих людей. Ці акустичні моделі відображають анатомію (наприклад, розмір і форма горла чи рота) і вивчають поведінкові моделі (наприклад, голос основного тону, стиль говоріння). Перевірка мовця заробила класифікацію «поведінкової біометрії».

Прогрес у технології штучних нейронних мереж призвів до розробки голосової біометрії, яка вважається швидшою та точнішою, ніж інші методи біометричної автентифікації. Очікується, що обсяг ринку біометричного розпізнавання голосу зросте з $1,1 млрд у 2020 році до $3,9 млрд до 2026 року при середньорічному темпі зростання 22,8%.

За даними нещодавнього дослідження, кількість шахрайських атак проти фінансових установ зросла на 269% за попередні чотири роки – це більше, ніж у більшості інших досліджуваних галузей. Тож саме фінансові установи активно впроваджують технології розпізнавання голосу і, ймовірно, в майбутньому використовуватимуть розпізнавання голосу в поєднанні зі своїми програмами для розпізнавання мови.

Варіанти розпізнавання

Перш за все, необхідно створити голосовий зліпок особи та зберегти його в базі даних для подальшої автентифікації. Розпізнавання голосу схоже на розпізнавання райдужної оболонки ока або відбитків пальців. Кожен голосовий зліпок унікальний для окремої людини, повторити його неможливо. Надалі система зіставлятиме голос особи з попередньо збереженою інформацією.

Системи голосової ідентифікації можна поділити на наступні класи: текстозалежні, текстонезалежні, дикторозалежні, дикторонезалежні.

- Дикторозалежні системи — це системи, які орієнтовані на ознаки мовлення певної людини або групи осіб, тому вони можуть використовуватися для ідентифікації тільки цієї особи (групи осіб). При зміні диктора (особи, яка ідентифікується системою) необхідно налаштовувати систему знову з використанням голосових ознак нового диктора.
- Дикторонезалежні системи — це системи, які не прив’язані до голосових ознак певної особи, та можуть використовуватися для ідентифікації будь-якої особи.

Такі системи самі виділяють необхідні ознаки голосу та порівнюють їх з еталоном з бази.

- Текстозалежні системи — це системи голосової ідентифікації, які здійснюють ідентифікацію особи за допомогою певного ключового слова або ключової фрази, яку повинна вимовити особа, що проходить ідентифікацію, наприклад, проголошення парольної фрази, яка кожного разу генерується випадковим чином. Використання індивідуальних ознак і збіг згенерованої та розпізнаної фраз підвищує надійність.
- Текстонезалежні системи — це системи, які здійснюють ідентифікацію особи за допомогою голосу без прив'язки до будь-яких ключових слів. У даному випадку важливе значення мають артикуляційні ознаки голосу людини, саме вони використовуються як головні ознаки, а фізіологічні ознаки виступають як вторинні. Текстонезалежна ідентифікація має на увазі використання тільки індивідуальних ознак.

Важливою характеристикою системи голосової ідентифікації є швидкість (швидкодія) визначення особистості. Підвищення швидкодії може бути досягнуто за рахунок використання нових швидких алгоритмів обробки даних.

Порівняльний аналіз методів реалізації голосової біометрії

Наразі існує два основних етапи для розпізнавання мовних шаблонів: навчання та порівняння. Для цього необхідно добре сформувати математичні основи та створити послідовність представлення шаблонів для його надійного порівняння. Існують такі підходи до реалізації голосової біометрії:

- на основі шаблону – при авторизації особи її мовлення порівнюється з набором попередніх записів для знаходження найкращої відповідності. Перевагою є те, що даний підхід досить продуктивний для пошуку точних моделей, а недоліком – фіксування раніше записаних шаблонів;
- стохастичний підхід – відхилення в мовленні моделюється статистично. Перевага його в тому, що використовується автоматична процедура статистичного навчання. Недоліком є те, що алгоритм повинен приймати попередні припущення в моделюванні, що може призвести до помилкового рішення, негативно впливаючи на продуктивність системи;
- підхід на основі знань (штучний інтелект) – поєднання акустичного фонетичного підходу та розпізнавання образів. Даний підхід варто застосувати для моделювання варіацій у мовленні, але такі знання важко здобути та уміло використати для успішної реалізації, тому цей підхід не є практичним.

Під час дослідження математичних алгоритмів було вирішено виокремити та розглянути шість методів, що найчастіше використовуються для реалізації голосової біометрії.

1. Метод опорних векторів (Support Vector Machine (SVM)) – це набір методів навчання з наглядом, що використовуються для класифікації, регресії та виявлення викидів. SVM є одним із найпопулярніших алгоритмів контрольованого навчання, який використовується для задач класифікації та регресії. Однак, насамперед, він використовується для задач класифікації в машинному навчанні. Метою алгоритму SVM є створення найкращої лінії або межі рішення, яка може розділити n-вимірний

простір на класи, що сприятиме легкому розміщенню нової точки даних у правильну категорію. Ця межа найкращого рішення називається гіперплощиною. Основні переваги SVM: ефективність у великих просторах та у випадках, коли кількість вимірів переважає над кількістю зразків; використання підмножини навчальних точок у функції прийняття рішень; різні функції ядра можна вказати для функції прийняття рішень; надаються звичайні ядра і можна вказати власні. До недоліків можна віднести те, що коли кількість функцій набагато більше, ніж кількість зразків, необхідно уникати підбору функцій ядра, і термін її регуляризації є вирішальним; SVM не надають безпосередньої оцінки ймовірності, вони виконуються за допомогою дорогої п'ятикратної перехресної перевірки.

2. Модель суміші Гаусса (Gaussian Mixture Model (GMM)) – це імовірнісна модель, яка передбачає, що всі точки даних генеруються на основі суміші гауссових розподілів з невідомими параметрами. Модель гауссової суміші може бути використана для: кластеризації, яка є завданням групування набору точок даних у кластери; пошуку кластерів у наборах даних, де кластери не можуть бути чітко визначені; оцінювання ймовірності того, що нова точка даних належить кожному кластеру. Гауссові моделі суміші також відносно стійкі до викидів, це означає, що вони все ще можуть давати точні результати, навіть якщо є деякі точки даних, які однозначно не вписуються в жоден із кластерів. Це робить GMM гнучким і потужним інструментом для кластеризації даних. Переваги моделі: найшвидший алгоритм для вивчення моделей сумішей; алгоритм не зміщує середнє значення до нуля або не зміщуватиме розміри кластерів, тому що він максимізує лише ймовірність. До недоліків можна віднести те, що при недостачі точок, оцінювання матриць стає важким процесом, алгоритм розходиться і знаходить рішення з нескінченною ймовірністю; цей алгоритм буде завжди використовувати всі компоненти, до яких має доступ, потребуючи закритих даних або теоретичних критеріїв інформації.

3. Прихована марківська модель (Hidden Markov Model (HMM)) – це модель, в якій спостерігається послідовність викидів, але невідомою є послідовність станів, через які пройшла модель, щоб створити викиди. Аналіз прихованих марківських моделей спрямований на відновлення послідовності станів зі спостережуваних даних. При правильному застосуванні цієї моделі для вирішення ряду важливих питань та прикладних задач може привести до позитивних результатів. Для HMM є такі проблеми: враховуючи параметри моделі та спостережувані дані, необхідно оцінити оптимальну послідовність прихованих станів і розрахувати ймовірність даних; враховуючи лише спостережувані дані, потрібно оцінити параметри моделі. Першу та другу задачу можна вирішити за допомогою алгоритмів динамічного програмування, відомих як алгоритм Вітербі та алгоритм «Вперед-Назад», відповідно. Останній може бути вирішений за допомогою ітераційного алгоритму максимізації очікування (EM), відомого як алгоритм Баума-Велча.

4. Метод динамічного викривлення часу (Dynamic Time Warping (DTW)) – є одним з алгоритмів для вимірювання подібності двох тимчасових рядів, які можуть відрізнятися за швидкістю. Метою методів порівняння часових рядів є отримання метрики відстані між двома вхідними часовими рядами. Подібність або несхожість подвійних рядів, зазвичай, обчислюється шляхом перетворення даних у вектори та обчислення евклідової відстані між цими точками у векторному просторі. DTW дає нелінійне

(пружне) вирівнювання між подвійними рядами. Даний метод шукає найкраще узгодження між дворазовими рядами. Це створює більш інтуїтивну міру подібності, що дозволяє схожим фігурам збігатися, навіть якщо вони не зберігаються за фазою на осі часу. Добре використовується для поєднання зразка голосової команди з командою інших, навіть якщо людина говорить швидше або повільніше, ніж попередньо записаний зразок голосу.

5. Векторне квантування (Vector Quantization (VQ)) – це блочний метод просторової області, який став дуже популярним з початку 1980-х років. У VQ вхідні дані зображення спочатку розкладено на k-вимірні вхідні вектори. Дуже важливою проблемою у VQ є дизайн кодової книги. Для розв'язання цієї проблеми найчастіше використовується алгоритм Лінде-Бузо-Грея (LBG), який є узагальненням алгоритму ЛлойдаМакса для скалярного квантування. Векторне квантування має продуктивність, яка конкурує з продуктивністю перетворення кодування. Хоча складність декодера є незначною (таблиця пошуку), висока складність кодера та високі вимоги до пам'яті методу все ще обмежують його використання на практиці. Як і кодування трансформації, VQ має проблему блокування артефактів на дуже низьких швидкостях.

6. Штучні нейронні мережі (Artificial Neural Networks (ANN)) – це обчислювальна модель, що складається з кількох елементів обробки, які отримують вхідні дані та видають вихідні дані на основі їх попередньо визначених функцій активації. Ці мережі імітують біологічну нейронну мережу, але використовують скорочений набір концепцій біологічних нейронних систем. Зокрема, моделі ANN моделюють електричну активність мозку та нервової системи. Елементи обробки (також відомі як нейрод або персептрон) з'єднані з іншими елементам. Зазвичай, нейроди розташовуються в шарі або векторі, при цьому вихід одного шару служить вхідним сигналом для наступного шару і, можливо, інших шарів. Нейрод може бути з'єднаний з усіма або підгрупою нейродів у наступному шарі, при цьому ці з'єднання імітують синаптичні зв'язки мозку. Сигнали зважених даних, що надходять у нейрод, імітують електричне збудження нервової клітини та передачу інформації всередині мережі або мозку. Теоретично, для моделювання асинхронної діяльності нервової системи людини елементи обробки штучної нейронної мережі також повинні бути активовані зваженим вхідним сигналом асинхронно. Однак більшість програмних і апаратних реалізацій штучних нейронних мереж реалізують більш дискретизований підхід, який гарантує, що кожен елемент обробки активується один раз для кожного представлення вектора вхідних значень.

Ідентифікація мовника та її надійність

Голосова ідентифікація — одна із найпривабливіших систем для ідентифікації, однак існуючі на даний момент проблеми у даному виді біометричних систем повинні бути, як мінімум, враховані у працюючих системах. Наприклад, розпізнавання за голосом може ефективно використовуватись як додатковий метод, наприклад, до розпізнавання за обличчям, оскільки ймовірність помилки самостійного розпізнавання за голосом складає 2–5 %. Сьогодні напрямок ідентифікації особи за голосом активно розвивається. Перевагою голосової біометрії є простота реалізації системи, яка, зазвичай, складається із голосового приймача, диктофона, голосового модулятора, біометричного програмного забезпечення та бази даних голосів. На відміну від інших біометричних технологій, голосова біометрія дозволяє здійснювати верифікацію на великій відстані. Одним з перспективних шляхів підвищення надійності голосової ідентифікація є

залучення характеристик динаміки підсвідомих рухів, що активно використовується при ідентифікації по підпису. З іншого боку, існують галузі застосування, в яких голосова ідентифікація є найбільш зручною, наприклад, віддалений доступ до телекомунікаційних каналів зв'язку з аналізу голосових даних.

Метод розпізнавання за голосом ідентифікує особу за сукупністю унікальних характеристик голосу. Алгоритми аналізують основні ознаки, за якими приймається рішення про особу диктора: голосового джерела, резонансних частот мовленнєвого тракту, їх затухань, а також динаміку управління артикуляцією.

Спираючись на багатий досвід, сучасні науковці при ідентифікації за голосом використовують дві групи ознак, які характеризують голос.

Перша група — це фізіологічні (анатомічні) ознаки, які пов'язані з особливостями механізму мовотворення людини. Друга група — це так звані артикуляційні ознаки, які засновані на особливостях роботи нервової системи людини, яка визначає характер використання фізіологічних ознак.

Фізіологічні ознаки. Фізіологічні ознаки засновані на моделі мовленнєвого тракту. В даному випадку як основні ознаки виступають декілька параметрів, які характеризують голос:

— енергія мовного сигналу;

— частотний діапазон мовного сигналу;

— основна частота — визначає довжину мовного тракту;

— форманти — визначають концентрацію мовного сигналу за частотою та характеризують голосні звуки.

Артикуляційні ознаки. Якщо фізіологічні ознаки відображають статистичні властивості мовного апарату, то артикуляційні ознаки дозволяють здійснити опис поведінки мовного апарату у часі, тобто відобразити артикуляційну динаміку мови. Головним фактором, який впливає на цю групу ознак, є соціально обумовлені мовленнєві навики людини, її індивідуальний опит, темперамент та особливості характеру. Артикуляційні ознаки враховують інтонацію мовлення, ритм, наголоси, гучність. Для того, щоб отримати ці характеристики, використовується поняття синтагма.

Синтагма — це ритмічно-мелодична одиниця мови, граматично оформлена та визначена у межах більш складної цілої структури (наприклад, речення) із закінченою думкою. У межах синтагми відокремлюють сегменти характеристики мови та інтонаційні характеристики мови, а саме: інтенсивність голосу; мелодійність голосу; система наголошень; часові характеристики — довжина сегментів та пауз; темп мовлення; тон мовлення.

Слід зазначити, що сучасні системи ідентифікації за голосом можуть одночасно використовувати фізіологічні й артикуляційні ознаки.

Підвищення надійності голосової ідентифікації є важливим не тільки для такого напрямку, як розмежування доступу до фізичних та інформаційних об'єктів, наприклад, доступу до операційної системи персонального комп'ютера або віддаленого доступу до телекомунікаційних каналів зв'язку з аналізу голосових даних. Певний інтерес є і для суміжних напрямів мовних технологій: розпізнавання усного мовлення, управління

голосовими командами тощо. Сьогодні широкого поширення набув електронно-цифровий підпис для захисту конфіденційних документів у вигляді захищеного електронного пристрою (token), у зв’язку з цим перспективним напрямком є розробка захисту конфіденційних документів на основі мовного підпису.

Крім того, практичні застосування таких досліджень корисні для правоохоронних органів, наприклад, ототожнення особи за фізичними параметрами голосу.

Однак розвиток технологій та технічний прогрес несе в собі не лише позитивні моменти, а й розширює можливості злочинців. Одним із основних правопорушень у телекомунікаційному середовищі є телефонне шахрайство, яке стрімко набирає популярності та перетворюється у справжню епідемію. Жертвами злочинців стають усі без винятку — це і бізнесмени, і чиновники, і зірки шоу-бізнесу, і звичайні громадяни.

Нижче наведені основні види телефонного шахрайства та засоби боротьби із ними.

Найрозповсюджений вид телефонного шахрайства — так званий «Родич у біді». Як це організовано? Людині дзвонять із невідомого номера. Злочинець представляється родичем або знайомим та схвильованим голосом повідомляє, що він затриманий співробітниками поліції та звинувачується у скоєнні того чи іншого злочину. Це може бути як дорожньо-транспортна пригода, так і зберігання зброї, наркотичних засобів, нанесення тілесних пошкоджень та навіть вбивство. Далі в розмову вступає так званий співробітник поліції, який впевненим тоном повідомляє, що неодноразово допомагав таким людям. Для вирішення питання необхідна певна сума грошей, яку слід привезти в умовлене місце та передати якомусь чоловіку. Ціна питання зазвичай складає від однієї до кількох тисяч доларів.

Схожі голоси, голоси однієї групи, які не мають відмінностей, можуть кодуватися в системах сотового зв’язку приблизно однаково, тому виявлятимуться подібними до ступеня змішування при слуховому сприйнятті та порівнянні інтегральних акустичних параметрів.

Саме в такій особливості передачі мовлення по сотовому зв'язку лежать передумови здійснення телефонного шахрайства, коли при зверненні по сотовому зв'язку досить висока ймовірність помилкового впізнання чужого голосу як знайомого. Як відомо, емоційний стан людини суттєво впливає на характеристики голосу, манеру розмови тощо. Траплялися випадки, коли, отримавши таке повідомлення, людина піддається на обман, навіть якщо особа, про яку йдеться у повідомленні, знаходиться поруч. Так, схвильовані батьки: вони завжди переймаються своїми дітьми, і реакція на можливу загрозу для них — дуже сильна. Проста, ефективна та нахабна схема, яка використовує сильні почуття та базові інстинкти. На такому ж принципі оснований вид телефонного шахрайства — «Мамо, у мене проблеми, не дзвони, перекажи гроші на цей рахунок».

Подібно відбиткам пальців у судовій криміналістичній експертизі слідів рук (дактилоскопічна експертиза), у експертизі відео-, звукозапису використовують свої об’єкти судової експертизи, а саме відео-, звукозаписи, зафіксовані на носіях інформації. У зв’язку з цим для судової криміналістичної експертизи використовують свої специфічні методи і технічні засоби.

Фізичною основою верифікації за голосом служить анатомія мовленнєвого тракту, властивості системи управління артикуляцією і особливості голосового джерела.

Анатомія тракту визначає спектральні характеристики звуків мовлення, система управління артикуляцією впливає на темп мовлення, швидкість перехідних процесів і

тривалість мовленнєвих сегментів, а голосове джерело визначає частоту основного тону і тембральні характеристики мовленнєвого сигналу.

Досліджуються тільки такі ознаки, які можуть бути безпосередньо виміряні в мовленнєвому сигналі. Разом з тим, як показують результати досліджень, верифікація дикторів у просторі акустичних параметрів забезпечує характеристики, що задовольняють та найбільше підходять до оточення реального застосування.

При діагностиці використовують відомі ознаки класу або групи об'єктів і порівнюють їх з ознаками конкретного об'єкта, в результаті чого визначається приналежність його до даного класу або групи. Для досягнення мети експерти вирішують різні види завдань.

Важливою характеристикою системи голосової ідентифікації є стійкість. Під перешкодами розуміються спотворення, шуми, імпульсні перешкоди тощо. Сучасні методи класифікації, що використовуються в системах голосової ідентифікації, дуже чутливі до шуму, що призводить до зниження надійності при впливі шуму. Прикладом спотворення каналу може бути реверберація звуку, тобто звук багато разів відбивається від предметів у приміщенні. Навколишній фон, в деяких випадках будучи перешкодою, може мати значний вплив на голосову ідентифікацію. Він може мати «значний» рівень сигналу (наприклад, звуки транспортних засобів; звуки, вироблені пристроями та засобами побутового призначення; звуки, властиві механізмам, приборам, апаратам, пристроям та засобам, які супроводжують роботу цих джерел; звуки живої природи; звуки явищ природи; звукові сигнали механізмів; звуки приборів, апаратів та пристроїв, спеціально призначених для створення, посилення та випромінювання звуку тощо) та «перекривати» діапазон мовленнєвого сигналу.

У телефонному каналі перешкодами можуть бути клацання, перевантаження, музичні сигнали (тональні сигнали) тощо. Оскільки сучасні цифрові мобільні пристрої зазвичай мають вбудований мікрофон і продуктивні апаратні засоби, то створення системи автентифікації за голосом із залученням більш витратних за обчисленнями методів цілком вирішуване завдання для мобільних платформ.

Проте забезпечити мінімальні обчислювальні витрати при збереженні точності, завадостійкості до різних видів перешкод і достатню надійність при поширених апаратних засобах все ж необхідно.

Ідентифікація диктора — процес, за допомогою якого система може визначити, хто є диктором на основі інформації з мовленнєвого сигналу. Останнім часом в інтегрованих системах безпеки, системах відеоспостереження, системах охоронного телебачення (СОТ) широко застосовується також звукозапис. Інформація, що отримується, використовується і для виявлення порушників, і для аналізу стану аудіообстановки з метою контролю дій персоналу та охорони.

Аудіоінформація використовується також у системах передачі інформації (СПІ) телефонних переговорів, у системах оповіщення, тривожного виклику тощо. У зв'язку з цим актуальним стає вирішення завдань, що пов'язані з аналізом звукової інформації, яка отримана під час запису в системах безпеки і яка може використовуватися надалі для аналізу в спеціалізованих лабораторіях правоохоронних органів, лабораторіях і центрах судової експертизи, науково-дослідних і навчальних центрах із метою:

— ідентифікації особи за допомогою записаної фонограми;

— аналізу шумового фону, діагностики акустичної обстановки й умов проведення звукозапису;

— ідентифікації засобів звукозапису;

— підвищення якості та розбірливості у вже існуючих записів;

— захисту мовного сигналу від несанкціонованого доступу;

— стискання мовних повідомлень;

— встановлення дослівного змісту низькоякісних записів.

Під час вирішення завдань охорони фізичних об’єктів та інформаційних ресурсів від кримінальних і терористичних загроз дуже цікавим є використання аудіоінформації (звукових голосових записів) у системах контролю та управління доступом (СКУД). Особливість таких систем полягає в тому, що вони допускають віддалену (за допомогою телефону) та приховану автентифікацію, що інколи є єдиним можливим засобом встановлення особистості співрозмовника.

Можливості злому голосової біометрії

Існує дослідження, яке докладно розглядає питання "обману" комерційних систем біометричної ідентифікації за допомогою відкритих інструментів з клонування голосу.

Природно, дослідження чіткої однозначної відповіді не дає, але скоріше каже, що на шляху зловмисників насамперед постає недосконаліть систем клонування голосу, кількість та якість записів, отриманих шахраями, акценти та інші недосконалості світу. Відсотки "обману" за наявності ряду таких труднощів там не вражають.

Тепер детальніше про дослідження.

Голос робився на старій версії синтезу мови, яка ще мала низку яскравих дитячих "болячок".

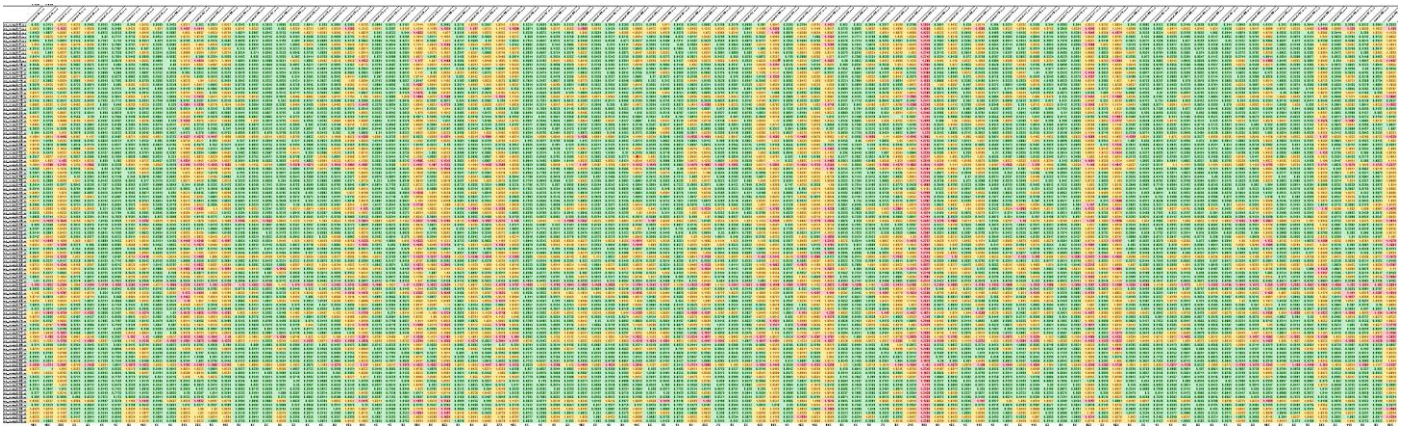
Також важливо сказати, що:

- Записано було близько кількох годин у відносно гарний мікрофон (зараз досить навіть 15 хвилин мови);
- Власник голосу – не диктор, просто умовно випадковий співробітник із яскравими артефактами дикції (системний адміністратор);
- Мікрофон був добрим, але на фоні трохи гудів кондиціонер;

Описана вище система голосової біометричної ідентифікації видає скоріше менше 1.1, якщо вона вважає, що у двох аудіо-файлах говорить одна і та сама людина, і скоріше більше 1.1, якщо вважає, що люди - різні (ці пороги були виставлені замовником або вендором системи) .

100 реальних аудіо людини порівнюються зі 100 різними згенерованими прикладами (включаючи приклад з таким самим текстом), створеними на основі текстів цих 100 реальних аудіо.

На малюнку нижче ви можете побачити порівняння 100 реальних аудіо зі 100 згенерованими аудіо та скора системи (позначено кольором). Зеленим кольором (не менше 1) позначено, що система "вважає" цю пару записів однією людиною. Червоним - це "різні" люди (скоріше більше 1.1). Жовтий – це прикордонні значення (швидкий між 1 і 1.1).



Якщо взяти середнє значення за точкою відсікання 1.1 (поріг встановлений замовником), то виходить, що в середньому НЕ ВДАЛОСЯ обдурити систему десь у 10% випадків. Тобто в 90% випадків систему обдурити ВДАЛОСЯ.

За загальною "матрицею" також видно, що "погані" (або "хороші", як дивитися) приклади згруповані. Основні помилки ставилися скоріше до недоліків старої системи синтезу:

- Нестабільна робота;
- Нестабільність на більш довгих аудіо та на дуже коротких аудіо;
- Система не брала більше 140 символів на вхід;

Ці артефакти були виправлені з переходом на новий синтез мови і в поточній версії синтезу в рамках однієї пропозиції визначити чи це вже практично неможливо.

Здається, що треба починати панікувати, адже люди заморочилися, замовили модель, провели пен-тест якоїсь супер-комерційної системи та виявилося, що у 90% випадків її можна обдурити! Жах! Але насправді, звичайно, жодного приводу для паніки немає.

Аудіо записували на якісний мікрофон у відносній тиші. Доклали максимум зусиль для обробки та чищення аудіо та вклали кілька обчислювальних ресурсів (нетривіальне, але й не десятки тисяч GPU-днів).

Абсолютно очевидно, що для масового обману цей підхід не підходить, так само як і системи клонування голосу в поточному стані (хоча цікавий прогрес, звичайно, є).

Навіть якщо шахраї будуть так само підковані (або таргетуватимуть супер "важливу" мету за допомогою деяких неадекватно дорогих і непрактичних публічних інструментів) - на їхньому шляху стане низька якість голосових семплів, які вони швидше за все зможуть отримати "лівими" способами. А зниження якості аудіо і його кількості різко знижує якість результату. Плюс багато "халявних" інструментів заточено тільки англійською мовою з очевидних причин. А інше дослідження явно вказує на зміну акценту/діалекту як основну причину різкого зниження ймовірності успіху "атаки".

Плюс треба пам'ятати, що повноцінний синтез завжди звучить краще, ніж клонування голосу. Ну і звичайно жартівливий висновок: якщо голосова фраза – єдиний ключ (а не частина складового) у вашого вендора послуг – біжіть у будь-якому випадку. Ніхто в здоровому глузді і так не розглядає голос як основний ключ (на відміну від смішних намірів деяких банків, що успішно і активно обнуляються, збирати ваші селфі).

Переваги голосової біометрії

1. Низькі операційні витрати

Кол-центри та навіть банки можуть заощадити гроші, використовуючи голосову автентифікацію. Це економить мільйони доларів, усуваючи багато кроків, необхідних у традиційних методах перевірки.

2. Покращена взаємодія з користувачем

Ще одна перевага голосових біометричних систем полягає в тому, що вони можуть значно покращити взаємодію з клієнтами. Абонентам більше не потрібно вводити паролі, PIN-коди або відповідати на контрольні запитання, щоб підтвердити свою особу.

3. Точність та надійність

Голосова автентифікація точніша та надійніша, ніж паролі, які легко забути, змінити чи вгадати шахраям. Це схоже на відбитки пальців, які є унікальними.

4. Проста у впровадженні технологія

Багато компаній цінують технологію розпізнавання голосу з точки зору використання та впровадження. Деякі біометричні технології можуть бути складними для інтеграції в бізнес, проте голосові біометричні системи зазвичай можуть бути реалізовані без додаткового обладнання та потребують мало ресурсів.

Висновок

Отже, підсумовуючи сказане вище, можна прийти до висновку, що технологія ідентифікації за голосом вже використовується в багатьох різноманітних сферах та дуже стрімко розвивається. Проте, також вона має й свої недоліки, такі як залежність від умов запису та якості мікрофону, але в майбутньому, з розвитком та покращенням технологій, вона може досягти ще кращих результатів та отримати справді дуже широке застосування на рівні з скануванням відбитку пальця та сітківки ока

Список використаної літератури

1. Коваль Л.Г., Злепко С.М., Новіцький Г.М., Крекотень Є.Г. Методи і технології біометричної ідентифікації за результатами літературних джерел.
2. Петро Бідюк, Володимир Бондарчук. Сучасні методи біометричної ідентифікації.
3. Г. М. Новіцький, С. М. Злепко, Л. Г. Коваль, І. О. Криворучко. Аналіз помилок ідентифікації й шляхи підвищення точності систем біометричної ідентифікації.
4. Валеріян Швець, Андрій Фесенко. Основні біометричні характеристики, сучасні системи та технології біометричної аутентифікації.
5. В. П. Захаров, В. І. Рудешко. Біометричні технології в ххі столітті та їх використання правоохоронними органами.