

АКУСТИЧНА ІДЕНТИФІКАЦІЯ ПО ГОЛОСУ

С. С. Єгоров^{1, а}, Є. А. Мачуський¹

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
Фізико-технічний інститут

Анотація

В статті описаний короткий аналітичний огляд акустичної ідентифікації по голосу, наведені принципи роботи основних методів ідентифікації. А також проведено аналіз цих методів та порівняння в залежності від різних параметрів сигналу.

Ключові слова: ідентифікація по голосу, основний тон, SVM, EER.

Вступ

Необхідність акустичної ідентифікації

Голос – така ж невід’ємна риса кожної людини, як і його обличчя або відбитки пальців. Широке поширення засобів зв’язку (стаціонарні і мобільні телефонні мережі, інтернет-телефонія і т.д.) відкривають великі можливості для застосування даного ідентифікатора.

Необхідно враховувати, що голос (наряду з почерком, ходом і т.п.) відноситься до т.зв. «Поведінкових» ідентифікаторів, і було б марним очікувати від технологій, що використовують дані ідентифікатори, високої точності і надійності.

Системи ідентифікації по голосу економічно вигідні тому, що їх можна встановлювати з обладнанням (наприклад, мікрофонами), що поставляються в стандартній комплектації з багатьма персональними комп’ютерами. Системи аутентифікації по голосу під час запису зразка і в процесі подальшої ідентифікації спираються на такі унікальні для кожної людини особливості голосу, як висота, модуляція і частота звуку. Ці показники визначаються фізичними характеристиками голосового тракту і унікальні для кожної людини.

Слід зазначити, що розпізнавання голосу відрізняється від розпізнавання мови. У той час як технологія розпізнавання мови інтерпретує те, що говорить абонент, технологія розпізнавання голосу використовується для ідентифікації особистості мовця. Оскільки голос можна просто записати на плівку або інші носії, деякі виробники вбудовують в свої продукти операцію запиту відгуку. Біометричний підхід, пов’язаний з ідентифікацією голосу, характеризується зручністю в застосуванні.

Технології та засоби ідентифікації по голосу застосовуються в ряді областей, безпосередньо пов’язаних з обробкою звернень користувачів по телефону (колл-центри і т.п.), що дозволяє прискорити обслуговуван-

ня абонентів і розвантажити операторів. У найбільш значних проектах (особливо пов’язаних з необхідністю захисту конфіденційної інформації) ідентифікація по голосу відіграє допоміжну роль по відношенню до інших біометричних технологій (перш за все ідентифікації за відбитками пальців).

З [1] можна дізнатися аналітичний огляд розпізнавання особистості по голосу.

Принцип роботи

Розпізнавання диктора підрозділяється на два напрямки [2, 3]:

- ідентифікацію;
- верифікацію.

При верифікації користувач пред’являє в тому чи іншому вигляді свій ідентифікатор і система розпізнавання повинна підтвердити або відкинути цей ідентифікатор.

При ідентифікації диктор не вказує свого ідентифікатора і система розпізнавання повинна встановити, чи належить мовної сигнал голосу одного з дикторів, які пройшли навчання. Важливий випадок ідентифікації полягає у визначенні того, чи належать два мовних сигналу голосу одного і того ж диктора.

1. Методи ідентифікації

Надійність роботи модуля поділу дикторів в значній мірі залежить від ступеня якості мови вхідних фонограм. Максимальна якість поділу мови дикторів (93-95% правильного маркування по дикторам) досягається при відношенні сигнал/шум не менше 20 дБ і тривалості мовлення кожного з дикторів на фонограмі не менше 16 секунд.

Ідентифікація і порівняння дикторів здійснюється на основі таких методів:

- метод статистик основного тону (ОТ);
- спектрально-формантний метод;
- метод повної мінливості.

^аyehorovss2000@gmail.com

1.1. Метод статистик основного тону [4]

Виділення основного тону (ОТ) здійснюється за допомогою спектрального аналізу сигналу фонограми. Основу методу складає алгоритм, який базується на аналізі значень гармонік ОТ в спектрі сигналу. Алгоритм спрямований на подолання проблем, пов'язаних з придушенням частини спектра сигналу в каналі.

Наприклад, в телефонному каналі в смузі від 0 до 300 Гц сигнал відсутній, однак саме в цьому діапазоні найчастіше знаходиться значення частоти ОТ. У цьому випадку алгоритм виділення ОТ дозволяє вирішити поставлену задачу за рахунок обліку гармонік частоти ОТ в смузі від 0,3 до 3,4 кГц.

Даний метод ідентифікації досить стійкий і до відношення сигнал/шум. Якщо рівень гармонік частоти основного тону хоча б на кілька децибел перевищує рівень шуму, алгоритм за допомогою спеціальних методів «підкреслення» гармонік ОТ дозволяє визначити шукану частоту.

На етапі обчислення статистичних характеристик ОТ обчислюється набір значень статистик ОТ за отриманою кривою ОТ. За одержаними статистичними оцінками характеристик ОТ будується модель диктора з використанням апроксимації даних з використанням Суміші Гауссового Розподілу (СГР).

На етапі прийняття рішення про ідентифікацію проводиться порівняння статистичних характеристик ОТ, обчислюються значення помилок першого і другого роду (FA і FR) по кожному з отриманих значень статистик ОТ, а також значення FA і FR по результатуючій метриці статистик ОТ як зважена сума всіх відносних відхилень значень статистик ОТ для двох порівнюваних фонограм.

1.2. Спектрально-формантний метод [5]

Даний метод заснований на тезі про унікальність геометрії вокального тракту у кожної людини і відображенні даного факту в різних спектральних характеристиках мови. Найбільш явно відмінність спектральних характеристик проявляється у взаємному розташуванні формант в частотній області. Даний метод заснований і на виділенні і порівнянні становища і динаміки поведінки трьох і більше формант.

Побудова моделі голосу диктора проводиться таким чином.

Будується середній спектр мовного сигналу, що використовується для побудови ідентифікаційної моделі диктора. Далі на його основі для даної фонограми будується нормалізуюча функція, яка використовується для обчислення миттєвих спектрів (кадрів) мовного сигналу. На кожному нормалізованому кадрі визначається положення всіх максимумів, з яких виділяються формантні частоти. Враховується відносна безперервність траєкторії формант і їх типові частотні області розташування кожної з формант для чоловічих і жіночих голосів.

В якості ідентифікаційних ознак використовуються положення 3-х спектральних максимумів, найбільш відповідних формантам мовного сигналу на

кожному спектральному зрізі, де їх вдалося досить надійно виділити. Ці ознаки найбільшою мірою визначають індивідуальні особливості геометрії голосового тракту.

Щільність розподілу ідентифікаційних ознак моделюється з використанням СГР.

Безпосереднє прийняття рішення про ідентифікацію проводиться на основі використання SVM-класифікатора (Support Vector Machine, машина опорних векторів). За результатуючою дистанцією SVM обчислюються помилки першого/другого роду (FR/FA).

1.3. Метод повної мінливості

Метод, названий в даній системі СГР або TotV (Total Variability, «Повна мінливість»), є найбільш сучасним методом ідентифікації по голосу.

Побудова моделі голосу диктора проводиться таким чином.

Обчислюються мовні ознаки MFCC (Mel Fourier Cepstrum coefficients, кепстральних коефіцієнти за шкалою мел) – показники, що характеризують покadroве спектральне подання мовного сигналу на тональних ділянках мови. Щільність розподілу ідентифікаційних ознак моделюється з використанням суміші гауссових розподілів (СГР).

Параметри СГР-моделі далі за допомогою спеціально пристосованого факторного аналізу представляються у вигляді вектора малої розмірності в так званому просторі «повної мінливості», в якому будуються підпростори так званих «власних каналів» і «власних голосів», які використовуються в алгоритмі JFA (Joint Factor Analysis, об'єднаний факторний аналіз). Відмінною особливістю такого подання СГР-моделі є її висока інформативність і малий розмір даних.

На етапі порівняння моделей дикторів відповідні їм вектори ознак порівнюються за допомогою алгоритму на основі SVM-класифікатора, який добре зарекомендував себе з точки зору швидкості і якості ідентифікаційного рішення. За результатуючою дистанцією SVM обчислюються помилки першого/другого роду (FR/FA).

2. Аналіз методів

Значення рівноймовірних помилки пропуску / відхилення цільового диктора (EER) для методу порівняння статистик основного тону голосу в залежності від тривалості порівнюваних мовних фрагментів становить 16 – 20%.

Характеристики статистик ОТ голосу залежать від емоційно-психологічного та фізичного стану диктора в момент мовлення, від стилю мовлення (читання, спонтанна мова, публічне вступ, підготовлена промова), від навколишнього акустичного обстановки. Тому при порівнянні дикторів важливо враховувати сумісність мовного матеріалу порівнюваних дикторів і, при наявності чинників, що впливають на зміну ОТ, враховувати це при прийнятті загального ідентифікаційного рішення.

Табл. 1. Порівняльні характеристики методів

Параметри сигналу	Методи		
	Метод статистик основного тону	Спектрально-формантний метод	Метод повної мінливості
Тривалість	++	+++	+++
Якість сигналу	+	++	++++
Фізичний та емоційний стан	++++	+	++

Застосування спектрально-формантного методу забезпечує значення величини EER до 6 – 7%. Значення цього показника для конкретного випадку залежить від тривалості та якості порівнюваних мовних фрагментів.

Спектрально-формантний метод досить стійкий до широкопasmового рівномірного шуму типу білого або шуму приміщення, а також до відмінностей в сигналах за рахунок різних типів каналів звукозапису.

Значення EER при використанні методу повної мінливості становить 4 – 5%. Однак цей метод вимагає високу якість сигналу, має сильну залежність від навчального матеріалу, а також потребує відносно великі часові затрати на виділення біометричних признаков.

Порівняльні характеристики наведених методів наведені в табл. 1 (кількість знаків «+» показує ступінь залежності методу від параметра).

Для покращення показників ідентифікації на основі методу повної мінливості, слід використовувати:

- канал-компенсацію;
- класифікатор на основі машини опорних векторів (SVM).

Слід відмітити, що на сьогоднішній день спільне використання моделей СРР та SVM являється домінуючим в задачах ідентифікації по голосу.

Висновки

З всього вище сказаного можна зробити висновок, що метод повної мінливості має найкращі показники

ідентифікації по голосу. Але він потребує багато часу та високу якість сигналу, що може бути недоцільним в деяких випадках.

Тому, в залежності від поставлених цілей та параметрів сигналу, можна використовувати метод статистик основного тону, або спектрально-формантний метод, які мають нижчі вимоги до емоційного стану диктора, якості та тривалості сигналу.

Перелік використаних джерел

1. Сорокин В. Н., Вьюгин В. В., Тананыкин А. А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. — 1986. — Т. 12, № 1. — С. 1–30.
2. Матвеев Ю. Н. Технологии биометрической идентификации личности по голосу и другим модальностям // Вестник МГТУ им. Н. Э. Баумана. Респ. Сер. «Приборостроение». — 2012. —
3. Матвеев Ю. Н., Симончик К. К. Система идентификации дикторов по голосу. — ООО «ЦРТ»: СПб, 2010.
4. Тассов К. Л., Дятлов Р. А. Метод идентификации человека по голосу // Наука и инновации. — 2013. — Вып. №6.
5. Поповский В. В., Бобрицкий С. М., Саваневич В. Е., Ткачев В. Н. Определение частот формант голосового сигнала на выходе гребенки полосовых фильтров. — ХНУР: Харьков, 2010.