НАПРЯМКИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ГОЛОСОВИХ СИСТЕМ АВТЕНТИФІКАЦІЇ

Кіщенко М. І., Пастушенко М.С.

Кафедра «Інфокомунікаційної інженерії ім. В.В. Поповського», ХНУРЕ, Україна

E-mail: marharyta.kishchenko@nure.ua
mykola.pastushenko@nure.ua

Abstract

The scientific problem of determining the directions of increasing the efficiency of voice authentication systems is being solved. As the main direction, it is proposed to use the phase data of the voice signal, which is currently ignored. In this case, the main attention is paid to the procedures for preprocessing the registration materials and the linear prediction coefficients, which are calculated based on the phase data.

В останні десятиліття досягнення науки та новітні інфокомунікаційні технології, як ніколи раніше, визначають динаміку економічного зростання, рівень добробуту населення, конкурентоспроможність держави у світовому співтоваристві, ступінь забезпечення її національної безпеки та рівноправної інтеграції у світову економіку. Стрімкий розвиток та широке використання сучасних інформаційно-телекомунікаційних систем ознаменували перехід людства від індустріального суспільства до суспільства інформаційного, в основі якого лежать нові системи комунікації, надійність яких не завжди відповідає вимогам, що зростають.

Разом з тим, процес інформатизації світової спільноти породжує комплекс негативних явищ, в першу чергу, розкраданням фінансових, інформаційних та обчислювальних ресурсів, а також персональних даних. Дійсно, висока складність і одночасно вразливість усіх систем, на яких базуються регіональні, національні та світові інформаційні простори, а також фундаментальна залежність від їхньої стабільності державних інфраструктур призводять до виникнення принципово нових загроз, які в деяких випадках можна вирішити за рахунок удосконалення систем доступу.

У зв'язку з широким поширенням розподілених систем у всіх сферах людської діяльності гостро постає завдання забезпечення інформаційної безпеки у таких системах. Одним із основних заходів щодо захисту фінансових засобів, інформаційних даних та обчислювального ресурсу є забезпечення надійної автентифікації користувача.

На даний момент існує безліч підходів до автентифікації та ще більше реалізацій цих підходів. Однак, в повному обсязі не всі класичні рішення завдання автентифікації підходять для використання у розподілених системах. А різні типи систем висувають унікальні вимоги до підсистем автентифікації. Крім того, активний розвиток комп'ютерної техніки дозволяє легко зламувати алгоритми автентифікації, які ще 10-15 років тому вважалися надійними. Наприклад, за 2019 рік сукупний розрахунковий дохід шахраїв, отриманих за допомогою банківських карток в Україні, зріс з 245,8 млн. грн. до 361,99 млн. грн. (зріс на 47,3%), про що повідомила заступник директора Української міжбанківської Асоціації членів платіжних систем ЕМА Олеся Дальниченка. Значною мірою це зумовлено ненадійністю парольного захисту банківських карток.

У зв'язку з цим ведеться безперервна робота у галузі дослідження та розробки методів автентифікації. Постійно з'являються нові та удосконалюються існуючі алгоритми, спрямовані на

забезпечення захищеної автентифікації користувачів. Все більш актуальною стає проблема автентифікації користувачів, які мають доступ до громадських та особистих інформаційних ресурсів. Особливо важлива ця проблема для відкритих, масових телекомунікаційних та інформаційних систем. Один із найперспективніших напрямів захисту подібних систем від несанкціонованих впливів – біометричні методи ідентифікації користувачів. Однак, незважаючи на всю привабливість, даний підхід пов'язаний з низкою серйозних проблем.

Спочатку розвиток та впровадження біометричних систем пов'язували зі статичними біометричними ознаками користувача (зображення обличчя, папілярний візерунок пальця та райдужна оболонка ока), які добре зарекомендували себе у криміналістиці. Однак, на цей час ці надії зруйновані, в першу чергу, через простоту підробки.

Тому останнім часом багато досліджень проводиться у сфері застосування динамічних (поведінкових) біометричних систем автентифікації. Серед цих біометричних систем особливе місце займає голосова автентифікація, яка відрізняється простотою, дешевизною та зручністю. Але, як і всі біометричні системи, голосова автентифікація має найнижчі якісні характеристики. У зв'язку з цим у сфері голосової автентифікації проводяться інтенсивні дослідження.

У сучасних системах голосової автентифікації (СГА) реєструється амплітудн-частотна інформація про полігармонічний нестаціонарний голосовий сигнал користувача. Автентифікація користувача здійснюється переважно у процесі аналізу амплітудно-частотного спектру матеріалів реєстрації [1]. Основні зусилля дослідників при цьому зосереджені на пошуку нових або вдосконаленні існуючих процедур формування (оцінки) шаблону (набору ознак – частоти основного тону, формантних даних, коефіцієнтів кепстру, мелчастотних кепстральних коефіцієнтів, коефіцієнтів лінійного передбачення та їх динамічних характеристик) користувача, і навіть на розробці вирішальних правил. Найбільш популярні серед останніх такі процедури прийняття рішень – методи гаусових сумішей (Gaussian Mixture Model, GMM) та опорних векторів (Support Vector Machine, SVM). Для цих цілей також використовуються штучні нейронні мережі та приховані Марківські моделі (Hidden Markov Models, HMM).

Метою даної роботи – дослідження впливу сучасних досягнень цифрової обробки інформації на точність оцінки окремих характеристик голосового сигналу, що аналізується в процесі формування шаблону користувача. Об'єктом дослідження є процес цифрової обробки голосових сигналів.

На нашу думку, підвищення показників якості СГА пов'язано, в першу чергу, зі зміною парадигми цифрової обробки матеріалів реєстрації, яка пов'язана з доповненням процедур аналізу амплітудно-частотного спектру сучасними досягненнями цифрової обробки інформації, в тому числі, і алгоритмами обліку фазових даних голосових сигналів. Таким чином, в даний час існує інший шлях підвищення якісних показників СГА, який базується насамперед на використанні фазової інформації голосового сигналу користувача. Давно відомо, що фаза є більш інформативним параметром сигналу, проте у сучасних СГА вона традиційно ігнорується.

Зумовлено це тим, що для отримання фазової інформації потрібні додаткові обчислювальні та алгоритмічні ресурси, які не завжди були доступними у зазначених додатках. Зауважимо, що раніше в радіолокації та радіозв'язку для отримання фазових даних використовувалися спеціальні громіздкі пристрої – фазообертачі, які неможливо було застосовувати в області обробки голосових сигналів. В даний час існують спеціалізовані мікросхеми або цифрові сигнальні процесори, які можуть бути застосовані і в цифровій обробці голосових сигналів.

Крім цього, є деякі особливості оцінки, попередньої обробки та використання фазових даних. Слід зазначити, що в даний час відсутні досвід і практика використання фази сигналу стосовно задач голосової автентифікації. Підтвердженням цього є те, що відомо лише обмежену кількість робіт, де фазові дані використовувалися для обробки мовних сигналів.

Зазначене вище наголошує на актуальність досліджень оцінки впливу фазових даних на якісні характеристики процедур голосової автентифікації. Фазові дані в голосовій автентифікації можуть використовуватися за кількома практично важливими для цифрової обробки голосових сигналів напрямками:

- підвищення відношення сигнал/шум матеріалів реєстрації (відомий напрямок використання фази в радіолокації та радіозв'язку);

- підвищення якості формування ознак для шаблонів, що традиційно використовуються, наприклад, частоти основного тону, формантної інформації та ін.;

- розробка нових процедур формування елементів шаблонів на основі фазових даних.

Очевидно, технології ідентифікації особи за голосом прийшли до систем автентифікації користувачів із криміналістики. У криміналістиці розпізнавання диктора має лише ймовірнісний характер, тобто. із зазначенням правдоподібності того, що два мовні сигнали належать одному й тому ж людині. У силу малої вибірки мовних сигналів довірчий інтервал оцінки правдоподібності приналежності двох записів мови одному й тому диктору настільки великий, що однозначне рішення неможливе.

Досить близьким є завдання сегментації дикторів. Сегментація дикторів у потоці розмови різних дикторів (audio-indexing, diarization) необхідна при розмітці звукових стенограм, телеконференцій, радіо- та телепередач, інтерв'ю і т. д. Однак, як і в криміналістиці, якість виділення диктора є низькою і неприйнятним для вирішення завдань автентифікації користувача.

Індивідуальність акустичних характеристик голосу визначається трьома факторами: механікою коливань голосових складок, анатомією мовного тракту та системою керування артикуляцією. Природно, певний вплив на акустичні характеристики може надавати канал поширення голосового сигналу (наприклад, вплив зовнішнього шуму), вплив якого в сучасних системах усувається за допомогою процедур цифрової обробки та організаційних заходів. Акустично стиль реалізується у вигляді контуру частоти основного тону, тривалості слів та його сегментів, ритміки ударних сегментів, тривалості пауз, рівня гучності.

Простір ознак, у якому приймається рішення про особистість диктора, має формуватися з урахуванням усіх факторів механізму мовлення: голосового джерела, резонансних частот мовного тракту та їх згасань, а також динаміки управління артикуляцією. Розглядаються такі параметри голосового джерела: середня частота основного тону, контур частоти основного тону, флюктуація частоти основного тону і форма імпульсу збудження. Спектральні характеристики мовного тракту описуються огинаючою спектру та його середнім нахилом, формантними частотами та їх смугами, довготривалим спектром або кепстром.

Кепстр визначає форму огинаючої спектра сигналу, в якій інтегруються характеристики джерел збудження (голосового, турбулентного та імпульсного) та форми мовного тракту. В експериментах з суб'єктивного розпізнавання диктора було встановлено, що спектр, що огинає, сильно впливає на впізнаваність голосу.

Замість обчислення спектра голосового сигналу з використанням дискретного перетворення Фур'є на короткому інтервалі часу може використовуватися також амплітудно-частотна характеристика сигналу, знайдена за коефіцієнтами лінійного передбачення мови.

Таким чином, шаблони автентифікації (розпізнавання диктора) формуються на основі цифрової обробки амплітудно-частотного спектра голосового сигналу користувача. У той же час ігнорується більш інформативний параметр голосових даних користувача, а саме фазочастотний спектр. Це може бути перспективним напрямом підвищення надійності голосової автентифікації.

У [1,2] проаналізовано можливості фазочастотного спектра голосового сигналу щодо: частоти основного тону, формантних частот, кепстральних і мел-частотних кепстральних коефіцієнтів. Аналіз одержаних результатів показує ефективність використання фазочастотного спектра голосового сигналу.

При цьому за рамками досліджень опинилися процедури попередньої обробки голосового

сигналу та не досліджені коефіцієнти лінійного передбачення.

Перспективність використання процедур попередньої обробки, у тому числі і матеріалів реєстрації, може базуватися на апріорній інформації про форму фазових даних, а також моделі аналітичного сигналу. В основу цієї обробки можуть бути покладені обчислювальні процедури. Результати попередньої обробки дозволять якісніше розрахувати складові шаблонів на основі як амплітудно-частотного, так і фазочастотного спектрів.

Перспективним напрямом є оцінка коефіцієнтів лінійного передбачення мови, які розраховані з урахуванням фазового спектра голосового сигналу. Слід зазначити, що коефіцієнти лінійного передбачення мови широко та ефективно використовуються у сучасних системах радіозв'язку, при цьому вказані коефіцієнти розраховуються на основі амплітудно-частотного спектру.

Практична важливість лінійного передбачення полягає у оцінці спектра досліджуваного сигналу з його деякому відрізку (кадрі) довжиною L відліків, і з погляду фільтрації – у отриманні рекурсивного адаптивного фільтра порядку K=M – 1 дільниці квазистаціонарності, тобто. на тому часовому відрізку тривалістю LT (T – період дискретизації), де M коефіцієнтів фільтра залишаються постійними. При цьому K – порядок передбачення. Підсумком вирішення задачі лінійного передбачення буде отримання коефіцієнтів адаптивного фільтра, амплітудно-частотна характеристика якого з гарним ступенем наближення збігається з формою спектра сигналу в кадрі. Таким чином, лінійне передбачення – це спосіб оцінки спектра сигналу, який характеризує користувача системи автентифікації на виході лінійного тракту з невідомими параметрами.

Коефіцієнти лінійного передбачення визначаються рішення рівнянь Юла-Уолкера, наприклад, з допомогою рекурсивного алгоритму Левінсона.

Спектр, отриманий методом авторегресії, називається спектром лінійного передбачення. Спектр лінійного передбачення, як і спектр розрахований за допомогою перетворенням Фур'е, явно при обробці мови не використовується. Зазвичай діапазон частот діапазону лінійного передбачення розбивається на задану кількість каналів. До кожного каналу розраховується середня потужність. Ці значення потужності використовують як коефіцієнти вектору параметрів. Коефіцієнти лінійного передбачення використовуються також для розрахунку кепстральних коефіцієнтів лінійного передбачення. В даний час коефіцієнти лінійного передбачення розраховуються на основі амплітудно-частотної інформації голосового сигналу, а фазові дані ігноруються.

Таким чином, подальші дослідження доцільно проводити за цими двома розглянутими напрямками з урахуванням фазових даних голосових сигналів.

*Л*ітература:

1. *Pastushenko M., Krasnozheniuk Ya., Zaika M.,* "Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User," International Conference "Problems of Infocommunications. Science and Technology" (PIC S&T'2020), 2020, pp. 1-5.

2. *Pastushenko M., Krasnozheniuk Y. , Lemeshko O.* Analysis of Voice Signal Phase Data Informativity of Authentication System User. International Workshop on Computer Modeling and Intelligent Systems (CMIS), Zaporizhzhya, ZNTU, 2020. P. 1-14.