

О.Ю. МЄШКОВ

Херсонський національний технічний університет

Д.М. БАРАНОВСЬКИЙ

Вінницький національний технічний університет

РОЗРОБКА ТА ДОСЛІДЖЕННЯ КРИТЕРІЮ АУТЕНТИФІКАЦІЇ ОСОБИСТОСТІ НА ОСНОВІ ЛОКАЛІЗОВАНИХ СТРУКТУР ГОЛОСОВОГО СИГНАЛУ

У статті розглянуто розроблений автором критерій аутентифікації особистості. Даний критерій базується на локалізованих структурах голосового сигналу людини, які формуються у просторі характеристик «частота-структура». Охарактеризовано спосіб побудови даних структур та визначення їх геометричних параметрів. Визначено критерій аутентифікації особистості як відношення площі перекриття локалізованих структур аналізованого сигналу до площі базової локалізованої структури диктора. Визначено основні показники аутентифікації для даного методу та проаналізовано їх динаміку.

Ключові слова: голосовий сигнал, аутентифікація, критерій аутентифікації, локалізовані структури, персоніфікований голосовий еталон.

O.YU. MIESHKOV

Kherson National Technical University

D.M. BARANOVSKIY

Vinnytsia National Technical University

DEVELOPMENT AND INVESTIGATION OF PERSONAL AUTHENTICATION CRITERIA BASED ON THE LOCALIZED VOICE SIGNAL STRUCTURES

An author's developed personal authentication criteria is given in the article. This criterion is based on the localized human voice signal structures that are formed in the characteristics space "frequency-structure". The method for these structures formation and their geometric parameters definition is characterized. The personal authentication criterion is defined as a cross-area ratio of the analyzed signal localized structures and the basic localized structure of the person. Main authentication parameters for this method are defined and their dynamics are analyzed. Selection of previously unsettled parts of the general problem. In our opinion, the unsolved problem is the development of the criterion of personality authentication by voice signal, which would allow to carry out this procedure with minimal computational complexity. It is also important to conduct a study of the features of this criterion, taking into account the dynamics of the human voice signal, which is used as the basis of this criterion. Formulating the purpose of the study. The purpose of the article is the development and experimental study of the criterion of identity authentication based on localized structures of the human voice signal. During the study, the criteria for personal authentication by voice were developed. This criterion involves the construction of localized structures in the space of the characteristics of the voice signal. Preliminary processing and allocation of these characteristics is performed on the basis of algorithms previously developed by the author. The geometric peculiarities of constructing localized structures in the characteristics space are determined for reducing the computational complexity of the authentication procedure. The authentication criterion is defined as the average value of the particle area of the localized structures of all phonemes that have fallen into the base areas of the corresponding phonemes. The dynamics of the developed authentication criterion based on FAR and FRR authentication indicators was studied. It is determined that over time the value of the developed criterion is stabilized at the level of 0.7, however for the authentication task it is expedient to set it at the level of 0.8-0.85. However, given the significant dynamics of the voice signal as an authentication base, it is advisable to update personalized voice standards with a frequency of 28-30 days.

Keywords: voice signal, authentication, authentication criteria, localized structures, personified voice standard.

Постановка проблеми. Для розв'язання задачі голосової аутентифікації на сьогоднішній день використовується велика кількість різних методів, на основі яких будується сучасне програмне забезпечення. У більшості методів з голосового сигналу виділяється ряд характеристик і на основі порівняння їх з певними еталонами приймається аутентифікаційне рішення. При цьому важливим моментом є саме критерій аутентифікації, на основі якого проводиться порівняння вхідного та базового сигналу та винесення рішення про вдалу аутентифікацію чи відмову у ній.

У роботах [1, 2, 3] розглянуто та описано методи первинної обробки та виділення характеристик голосового сигналу, необхідних для побудови системи аутентифікації. Також у даних роботах наведено опис критерію, який може використовуватись для задачі аутентифікації. Однак використання даного критерію потребує експериментального дослідження та визначення його адекватних значень. Тому дослідження критерію аутентифікації особистості за голосом є актуальною проблемою.

Аналіз останніх досліджень та публікацій. Дослідженню процедур голосової аутентифікації та аналізу голосового сигналу приділяли значну увагу вітчизняні та зарубіжні вчені, зокрема: Я.П. Драган, А.П. Кравченко, Е.Г. Жиляков, Г.Є. Ляшенко, Є.А. Первушин, Д.С. Голубинський, Е.В. Булгакова, Г. Фант, Дж. Фланган, Л.Р. Рабінер, Т. Матсуї та інші [4–6]. В основному їх дослідження спрямовані на розробку методів дослідження голосових сигналів та побудову систем аутентифікації у частотному просторі, що передбачає значну попередню обробку сигналу з метою виділення необхідних характеристик. Тому розробка сучасних методів, які передбачають проведення процедури аутентифікації у часовому просторі у режимі реального часу на основі критерію аутентифікації з мінімальним обсягом обчислень є доволі актуальною.

Виділення невирішених раніше частин загальної проблеми. На наш погляд, невирішеною

залишається проблема розробки критерію аутентифікації особистості за голосовим сигналом, який би дозволяв проводити дану процедуру з мінімальною обчислювальною складністю. Також важливим є проведення дослідження особливостей даного критерію з урахуванням динамічності голосового сигналу людини, який використовується як основа даного критерію.

Формування мети дослідження. Метою статті є розробка та експериментальне дослідження критерію аутентифікації особистості на основі локалізованих структур голосового сигналу людини.

Виклад основного матеріалу дослідження. У дослідженні [3] запропоновано критерій аутентифікації особистості за структурами, утвореними у просторі характеристик голосового сигналу. У вказаній роботі ці структури названі хмарними з огляду на їх візуальну подібність до хмар, а не до хмарних обчислювальних структур. Тому, з урахуванням того, що кожна вокалізована ділянка голосового сигналу локалізується у певній частині простору характеристик, більш логічно називати їх локалізованими структурами голосового сигналу. Таке визначення буде використовуватись надалі.

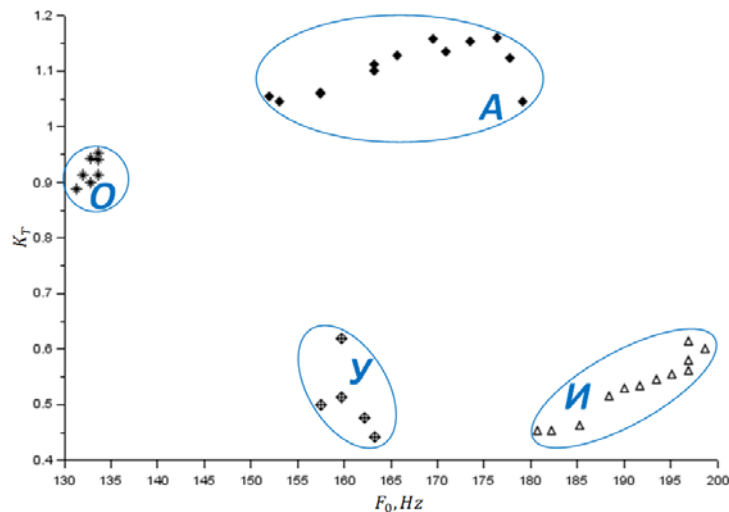


Рис. 1. Приклад локалізованих структур голосового сигналу у просторі характеристик

Виділення даних локалізованих структур базується на алгоритмах обробки голосового сигналу, описаних у роботах [1, 2].

Запропонований критерій аутентифікації передбачає попереднє формування бази персоніфікованих голосових еталонів. Для цього 60 дикторами було виконано акустичний запис паролльної фрази «Справи ідуть добре». Диктор вимовляє паролльну фразу у довільному стані, сидячи на стільці, тримаючи голову рівно. Запис паролльної фрази проводився у 5 сесій по дві вимови паролльної фрази послідовно. Інтервал між сесіями складав 15–20 хвилин. У результаті цього формувалися доволі широкі локалізовані структури для кожної з аналізованих фонем – «А», «И», «У» та «О» (рис. 2).

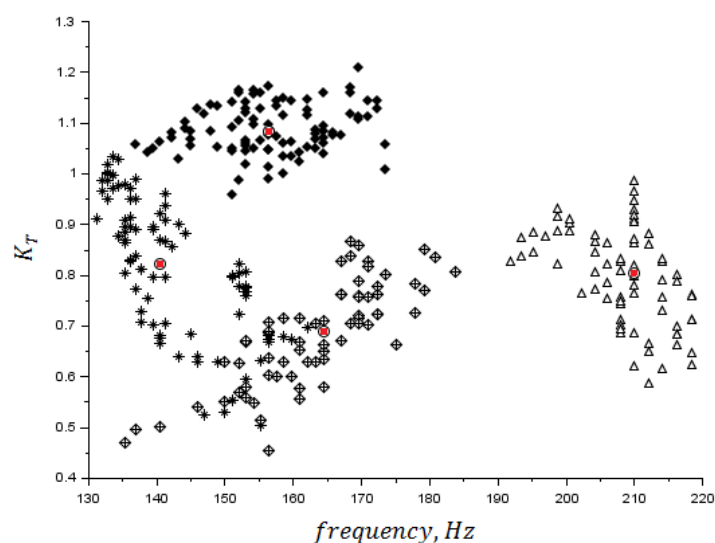


Рис. 2. Локалізовані структури голосового сигналу, які використовувались у дослідженні

Для кожної із сформованих таким чином локалізованих структур визначається середньозважений центр та деяка область простору навколо нього. Координати центру даної області визначаються як середнє арифметичне значень частоти фреймів та коефіцієнтів середньоквадратичного відхилення сигналів K_T відповідно:

$$K_{Tj} = \frac{\sum_{i=1}^{n_j} K_{Tji}}{n_j},$$

$$F_{0j} = \frac{\sum_{i=1}^{n_j} F_{0ji}}{n_j},$$
(1)

де K_{Tj}, F_{0j} – відповідні координати центру j -ї області;
 K_{Tji}, F_{0ji} – відповідні координати i -ї точки j -ї області;
 n_j – кількість точок в j -й області;
 $j = 1..m$ – кількість аналізованих звуків і відповідних областей простору характеристик.

Область простору, що виділяється навколо середньозваженого центру, має форму прямокутника. Це пов'язано з необхідністю швидкого та простого визначення меж координат точок простору характеристик, які потрапляють чи не потрапляють у дану область. Сторони кожного з цих прямокутників визначаються середньоквадратичним відхиленням точок простору, що відповідають фреймам досліджуваних сигналів:

$$\overline{S_{KT}} = \frac{\sqrt{\sum_{i=1}^{n_j} (K_{Tji} - K_{Tj})^2}}{n_j},$$

$$\overline{S_{F0}} = \frac{\sqrt{\sum_{i=1}^{n_j} (F_{0ji} - F_{0j})^2}}{n_j},$$
(2)

де $\overline{S_{KT}}, \overline{S_{F0}}$ – відповідні середньоквадратичні відхилення коефіцієнту структури та частоти.

Сторона кожного з утворених прямокутників приймається вдвічі більшою, ніж отримане значення середньоквадратичного відхилення відповідної характеристики голосового сигналу, оскільки відхилення враховується як в більшу, так і в меншу сторону (рис. 3).

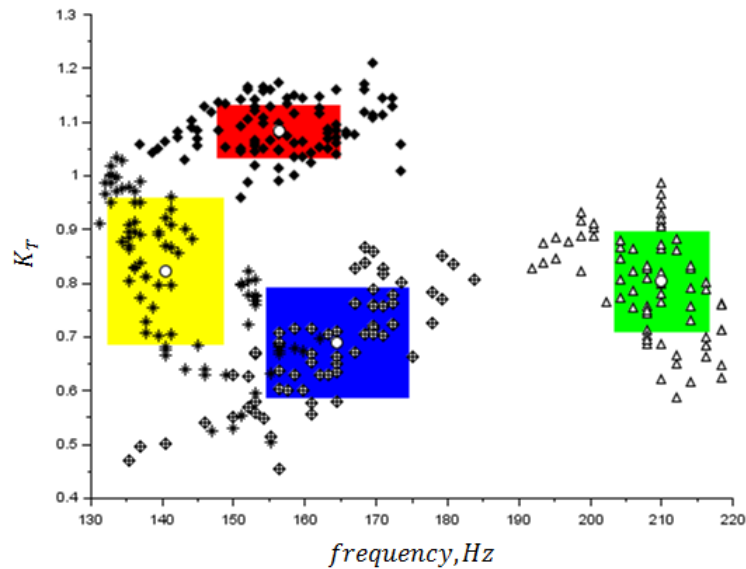


Рис. 3. Формування локалізованих структур у просторі характеристик «структура–частота»

Для кожного окремо взятого диктора система формує описані локалізовані структури на основі даних його власних акустичних записів, які несуть індивідуальну інформацію про цього диктора.

При виконанні процедури аутентифікації диктора виконується одиничний запис паролльної фрази. Отриманий акустичний матеріал обробляється аналогічними алгоритмами і для нього формується такі ж локалізовані структури із середньозваженими центрами. Критерій точності аутентифікації за такою процедурою пропонується визначати як середнє значення часток площі локалізованих структур усіх фонем, які потрапили у базові області відповідних фонем:

$$square_{f_{it}} = \frac{\sum_{i=1}^m S_i}{\sum_{i=1}^m S_{0i}} * 100\%,$$
(3)

де S_i – площа локалізованої структури аналізованої фонемі, яка потрапила до базової області даної фонемі;
 S_{0i} – загальна площа області аналізованої фонемі;
 m – кількість аналізованих фонем.

Для визначення загальної площі виділеної області аналізованої фонемі S_{0i} достатньо перемножити

сторони відповідного прямокутника у просторі характеристик голосового сигналу. Для визначення площі перекриття S_i необхідно спочатку перевірити, які вершини виділеної області аналізованої фонему потрапляють у межі базової області даної фонему. Якщо усі 4 вершини знаходяться у межах базової області, то коефіцієнт відношення площ для даної фонему автоматично приймається рівним 1. Якщо ж деякі вершини виходять за межі даної області, програмний алгоритм визначає точки перетину меж базової та аналізованої областей. Ці точки обираються як вершини області перекриття і на їх основі визначаються довжини сторін області перекриття. За відомими сторонами визначається площа даної області.

Якщо отримані центри разом із виділеними областями фонему потрапляють у межі аналогічних базових областей для даного диктора, вважається, що диктор пройшов аутентифікацію. Якщо ж центр аналізованої виділеної області хоча б однієї фонему не потрапляє до базової області, або критерій точності аутентифікації має значення менше деякого визначеного рівня помилки (порогу аутентифікації), то вважається, що диктор не проходить процедуру аутентифікації. Рівень даного порогу визначається на основі експериментального дослідження розробленого алгоритму на акустичних матеріалах реальних людей.

У ході експериментального дослідження для 40 дикторів, частина з яких була занесена до бази, було сформовано голосові еталони на основі різної кількості акустичних записів паролльної фрази. Після цього через певні проміжки часу (6 годин, 1 добу, 3 доби та 7 діб) усім 40 дикторам було запропоновано пройти процедуру аутентифікації шляхом запису паролльної фрази і порівняння її зі сформованими голосовими еталонами. Запис проводився таким самим чином, що і записи при формуванні персоніфікованого голосового еталону, описаному вище. Відмінність полягала лише у тому, що диктор вимовляв паролльну фразу послідовно три рази.

Для оцінки якості та точності технологій ідентифікації та аутентифікації використовуються два показники, які визначалися для кожного часового інтервалу. Першим з них є FAR (False Acceptance Rate) – ймовірність помилки першого роду, тобто ймовірність того, що система визнає «чужого» за «свого». Другий показник – FRR (False Rejection Rate) – ймовірність помилки другого роду, тобто ймовірність того, що система не аутентифікує особу, яка є в її базі.

Будь-яку систему аутентифікації можна налаштувати на різний рівень жорсткості відбору, тобто на різний рівень помилки першого роду. Чим нижчий показник FAR, тим більш чутливою є система, тим вище ймовірність помилкового не розпізнання FRR (рис. 4).

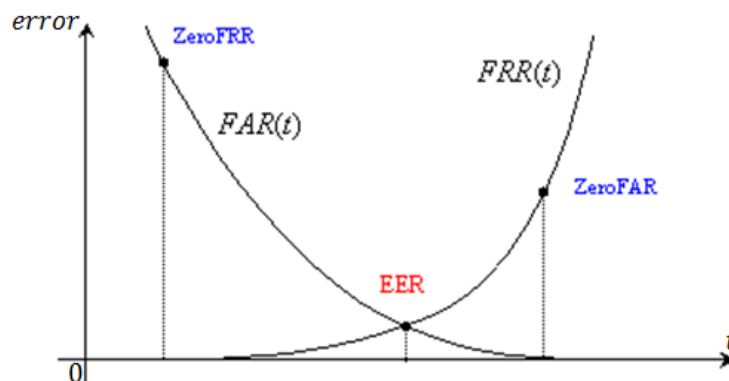


Рис. 4. Помилки першого (FAR) та другого (FRR) роду в залежності від порогу чутливості t [7]

При аналізі біометричних систем додатково також визначаються наступні показники:

- EER (Equal Error Rate) – системна помилка, за якої помилки першого та другого роду співпадають ($FAR = FRR$);
- $ZeroFAR$ – значення помилки другого роду, за якого значення помилки першого роду дорівнює 0 ($FAR = 0$);
- $ZeroFRR$ – значення помилки першого роду, за якого значення помилки другого роду дорівнює 0 ($FRR = 0$) [7].

Ідеальними є такі показники помилки і відмови аутентифікації, за яких при достатньо високій надійності аутентифікації досягається досить мала відмова аутентифікації. У загальному випадку, для кожної технології ці показники різняться. Залежно від конкретної задачі система налаштовується на певний компроміс між допустимими значеннями FAR і FRR.

Для розробленого способу аутентифікації особистості показник FAR визначався як відношення кількості аутентифікованих дикторів, які відсутні у базі, до загальної кількості дикторів, відсутніх у базі. Показник FRR визначався як відношення кількості не аутентифікованих дикторів, які наявні у базі, до загальної кількості дикторів, наявних у базі. Поріг чутливості задавався критичним рівнем критерію аутентифікації $square_{crit}$. Результати дослідження з урахуванням часових особливостей подано у таблиці 1 та на рис. 5.

Помилки аутентифікації першого та другого роду розробленої системи

FAR											
Часовий інтервал	square_fit										
	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
6 годин	0,55	0,50	0,50	0,45	0,40	0,30	0,20	0,15	0,05	0,00	0,00
1 доба	0,60	0,55	0,50	0,45	0,45	0,40	0,20	0,10	0,00	0,00	0,00
3 доби	0,60	0,55	0,50	0,35	0,30	0,25	0,15	0,05	0,00	0,00	0,00
7 діб	0,65	0,55	0,40	0,30	0,25	0,20	0,15	0,05	0,00	0,00	0,00

FRR											
Часовий інтервал	square_fit										
	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
6 годин	0,00	0,02	0,05	0,08	0,15	0,18	0,22	0,37	0,58	0,85	0,93
1 доба	0,00	0,03	0,08	0,13	0,17	0,22	0,32	0,42	0,60	0,85	0,93
3 доби	0,00	0,07	0,08	0,15	0,17	0,23	0,32	0,43	0,62	0,88	0,95
7 діб	0,02	0,07	0,13	0,15	0,20	0,28	0,33	0,50	0,65	0,88	0,97

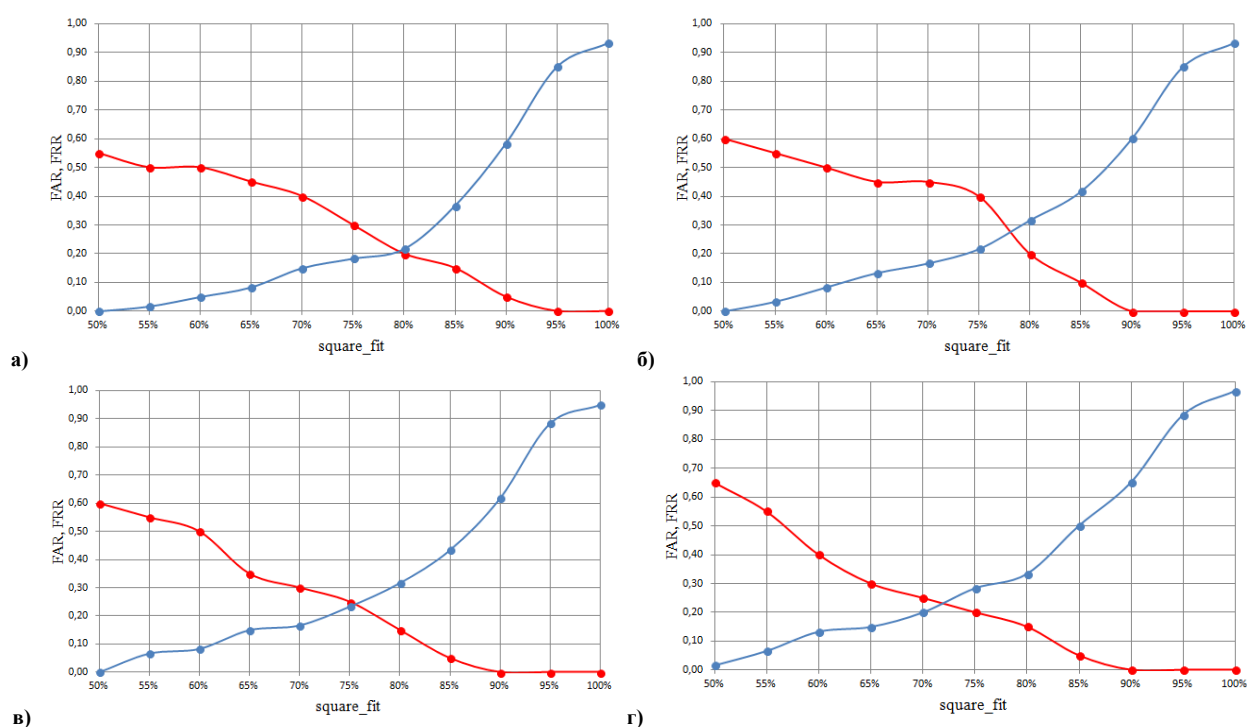


Рис. 5. Порівняння помилок аутентифікації при різних рівнях порогового значення критерію аутентифікації а – через 6 годин; б – через 1 добу; в – через 3 доби; г – через 7 діб після формування голосового еталону

Як бачимо, загальний вигляд кривих FAR та FRR є подібним до теоретичного, наведеного на рис. 4. Основна відмінність полягає у значному збільшенні помилки першого роду при високих порогових рівнях критерію аутентифікації. У той же час дана особливість вказує на те, що розроблена технологія є надто чутливою до таких рівнів критерію аутентифікації і їх використання на практиці не є бажаним.

З іншого боку, використання низьких рівнів критерію аутентифікації призводить до того, що система може аутентифікувати дикторів, які насправді відсутні у базі. З цієї причини використання критерію аутентифікації з пороговим рівнем нижче 50% також не є доцільним.

Якщо розглядати інші показники аутентифікації, то усі вони характеризуються незначною часовою динамікою. Показник **ESR** з плином часу збільшується від 20% до 27,3%. При цьому даний показник поступово зміщується в область менших порогових значень критерію аутентифікації з плином часу.

Показник **ZeroFAR** знаходиться у межах 90–95%, у той час, як **ZeroFRR** знаходиться на рівні 50%, що було вказано вище. Обидва ці показники мають тенденцію до зменшення з плином часу, однак їх динаміка також є незначною.

З огляду на отримані результати, можна стверджувати, що пороговий критерій аутентифікації **square_fit** для розробленої системи може бути встановлений на рівні 80–85% з подальшим динамічним коригуванням. Точне його значення повинно визначатись для кожної конкретної практичної реалізації у залежності від кількості дикторів в організації, а також особливостей їх голосових матеріалів. До того ж у розробленій системі передбачено динамічне регулювання даного параметру, що дозволить гнучко регулювати процедуру аутентифікації.

З іншого боку, постає питання в актуальності голосових еталонів дикторів з плином часу. З урахуванням того, що уже за тиждень рівень критерію аутентифікації для кожного диктора знижується майже до 70%, але стабілізується на цьому рівні, то такі еталони ще можна вважати актуальними. Однак через 28 днів голосові еталони для більшості дикторів уже перестають бути актуальними. Значення критерію аутентифікації для них у більшості випадків знижується до 50%, а інколи і нижче. Даний рівень критерію аутентифікації є близьким до показника *ZeroFRR* і є неприйнятним для використання у системі аутентифікації. З урахуванням цього при практичному використанні розробленої системи аутентифікації рекомендується оновлення голосових еталонів кожні 28–30 днів.

Висновки. У ході дослідження було розроблено критерії персональної аутентифікації за голосовим сигналом. Даний критерій передбачає побудову локалізованих структур у просторі характеристик голосового сигналу. Попередня обробка та виділення цих характеристик виконується на основі алгоритмів, попередньо розроблених автором.

Визначено геометричні особливості побудови локалізованих структур у просторі характеристик для зменшення обчислювальної складності процедури аутентифікації. Критерій аутентифікації визначено як середнє значення часток площі локалізованих структур усіх фонем, які потрапили у базові області відповідних фонем.

Досліджено динаміку розробленого критерію аутентифікації на основі показників аутентифікації FAR та FRR. Визначено, що з плином часу значення розробленого критерію стабілізується на рівні 0,7, однак для задачі аутентифікації доцільно встановлювати його на рівні 0,8-0,85. Однак з огляду на значну динаміку голосового сигналу як бази аутентифікації доцільно оновлювати персоналізовані голосові еталони з періодичністю 28–30 днів.

Література

1. Мешков О.Ю. Метод локальних максимумів для виділення вокалізованих ділянок голосового сигналу людини / О.Ю. Мешков, О.О. Новіков, С.М. Злепко // Вісник Хмельницького національного університету. – 2018. – № 6. – С. 197–210.
2. Мешков О. Ю. Запис та обробка первинного акустичного матеріалу для задачі аналізу голосового сигналу людини та виділення його основних характеристик / О. Ю. Мешков // Наукові праці : наук. журн. / Чорном. нац. ун-т ім. Петра Могили ; ред. кол. : М. Т. Фісун (голова) та ін. – Миколаїв, 2017. – Т. 307. Вип. 295. – С. 76–81. – (Комп'ютерні технології).
3. Мешков О.Ю. Аналіз голосових сигналів людини та аутентифікація особистості за голосом : монографія / О. Ю. Мешков, О. О. Новіков, В. О. Новіков. – Херсон : ФОП Вишемирський В.С., 2018. – 168 с.
4. Ляшенко Г. Є. Дослідження ефективності методів біометричної аутентифікації / Г. Є. Ляшенко, А. А. Астраханцев // Системи обробки інформації. – 2017. – № 2(148). – С. 111–114.
5. Первушин Е.А. Обзор основных методов распознавания дикторов / Е.А. Первушин // Математические структуры и моделирование. – 2011. – Вып. 24. – С. 41–54.
6. Булгакова Е.В. Метод идентификации дикторов на основе сравнения статистик длительностей фонем / Е.В. Булгакова, А.В. Шолохов, Н.А. Томашенко // Научно-технический вестник информационных технологий, механики и оптики. – 2015. – Том 15. № 1. – С. 70–77.
7. Біометрика [Електронний ресурс]. – Режим доступу : http://www.biometrika.it/eng/wp_biointro.html

References

1. O.Yu. Mieshkov, O.O. Novikov, S.M. Zlepko Metod lokalnykh maksimumiv dlia vydilennia vokalizovanykh dilianok holosovoho syhnalu liudyny // Visnyk Khmelnytskoho natsionalnoho universytetu – 2018. – №6. – S. 197-210.
2. Mieshkov O. Yu. Zaps ta obrobka pervynnoho akustychnoho materialu dlia zadachi analizu holosovoho syhnalu liudyny ta vydilennia yoho osnovnykh kharakterystyk // Naukovi pratsi : nauk. zhurn. / Chornom. nats. un-t im. Petra Mohyly ; red. kol. : M. T. Fisun (holova) [ta in.]. – Mykolaiv, 2017. – T. 307. Vyp. 295. – 132 s. – (Kompiuterni tekhnolohii). – S. 76-81.
3. Mieshkov O.Yu., Novikov O.O., Novikov V.O. Analiz holosovykh syhnaliv liudyny ta autentyfikatsiia osobystosti za holosom: Monohrafiia [Tekst] / O. Yu. Mieshkov, O. O. Novikov, V. O. Novikov. – Kherson: vyd-vo FOP Vyshemyrskiy V.S., 2018. – 168 s. (9,17 druk. ark.)
4. Doslidzhennia efektyvnosti metodiv biometrychnoi autentyfikatsii / H. Ye. Liashenko, A. A. Astrakhantsev // Systemy obrobky informatsii. — 2017. — № 2(148). — S. 111-114.
5. Pervushyn E.A. Obzor osnovnykh metodov raspoznavaniya dyktorov // Matematycheskye struktury y modelyrovanye, vyp. 24, s. 41-54, 2011.
6. Bulhakova E.V., Sholokhov A.V., Tomashenko N.A. Metod ydentyfikatsyy dyktorov na osnove sravneniya ctatystyk dlytelnoitei fonem // Nauchno-tekhnicheskyi vestnyk ynformatsyonnykh tekhnolohiyi, mekhanyky y optyky. 2015. Tom 15. № 1. S. 70–77.
7. Biometryka [Elektronnyi resurs] – Rezhym dostupu: http://www.biometrika.it/eng/wp_biointro.html

Рецензія/Peer review : 19.10.2018 р.

Надрукована/Printed :22.11.2018 р.

Рецензент: д.т.н., проф. Кичак В.М.