УДК 534.87

Мєшков О.Ю.
Херсонський національний технічний університет

РОЗРОБКА ПЕРСОНІФІКОВАНОГО ГОЛОСОВОГО ЕТАЛОНУ ДЛЯ ЗАДАЧІ АУТЕНТИФІКАЦІЇ ОСОБИСТОСТІ

У статті досліджується можливість використання характеристик голосового сигналу людини для задачі аутентифікації. Розроблено персоніфікований голосовий еталон особистості, який передбачає формування локалізованих структур у просторі характеристик голосового сигналу. Досліджено особливості використання різних парольних фраз та базових сигналів для побудови цих структур.

Ключові слова: *голосовий сигнал, аутентифікація, персоніфікований голосовий еталон.*

Постановка проблеми. Для розв'язання задачі голосової аутентифікації нині використовується велика кількість різних методів, на основі яких будується сучасне програмне забезпечення. У більшості методів із голосового сигналу виділяється ряд характеристик і на основі порівняння їх із певними еталонами приймається аутентифікаційне рішення. При цьому важливим моментом є саме форма персоніфікованого еталону, який використовується при порівнянні.

Аналіз останніх досліджень і 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, В.Н. Сорокін, Є.А. Первушин, Д.С. Голубинський, А.Ю. Трубіна, Г. Фант, Дж. Фланаган, Л.Р. Рабінер, Т. Матсуї [1-4]. В основному їх дослідження спрямовані на розробку методів дослідження голосових сигналів у частотному просторі, що передбачає значну попередню обробку сигналу з метою виділення необхідних характеристик. Тому розробка сучасних методів, які передбачають виділення характеристик у часовому просторі у режимі реального часу, є доволі актуальною.

Постановка завдання. Метою статті є розробка та дослідження персоніфікованого еталону особистості на основі характеристик голосового сигналу у часовому просторі.

Для досягнення мети необхідно виконати такі завдання:

1. охарактеризувати локалізовані структури, які формуються на основі характеристик голосового сигналу;

2. дослідити можливість використання різних парольних фраз для побудови цих локалізованих структур;

3. визначити базові сигнали, які найбільш чітко дадуть змогу розмежовувати ці структури у просторі характеристик голосового сигналу;

4. розробити персоніфікований еталон особистості за голосовим сигналом.

Виклад основного матеріалу дослідження. У попередніх роботах автора розроблено метод локальних максимумів із подальшим амплітудним уточненням. Цей метод використовується для вирішення часткової задачі сегментації сигналу, а саме виділення вокалізованих ділянок сигналу з потоку мови [5]. Кожна виділена вокалізована ділянка являє собою набір послідовних квазіперіодичних коливань і характеризується певною динамікою основної частоти сигналу та структури розподілу амплітуди у часовому просторі (рис. 1).

На основі двох виділених характеристик голосового сигналу – основної частоти сигналу та структури розподілу амплітуди – будується простір характеристик. Кожне одиничне коливання у такому просторі відповідатиме точці, координатами якої будуть ці характеристики. Для зведення структури розподілу амплітуди голосового сигналу до єдиного числового параметра використовується коефіцієнт середньоквадратичного відхилення структури сигналів, який для обробки сигналів у дискретних значеннях визначається як

$$K_T = \frac{\sqrt{\frac{1}{T}\sum_{j=1}^{N}\left(Y_{xj}-Y_{ij}\right)^2\Delta t}}{\sqrt{\frac{1}{T}\sum_{j=1}^{N}{Y_{ij}}^2\Delta t}}, \qquad (1)$$

де Y_{xj}, Y_{ij} – амплітуди j-го відліку введеного та базового сигналу;

Δt – інтервали часу між відліками;

T – період (тривалість) сигналів [6].

Загалом набір цих точок формуватиме у просторі деяку хмару, яку надалі називатимемо локалізованою структурою голосового сигналу (рис. 2).

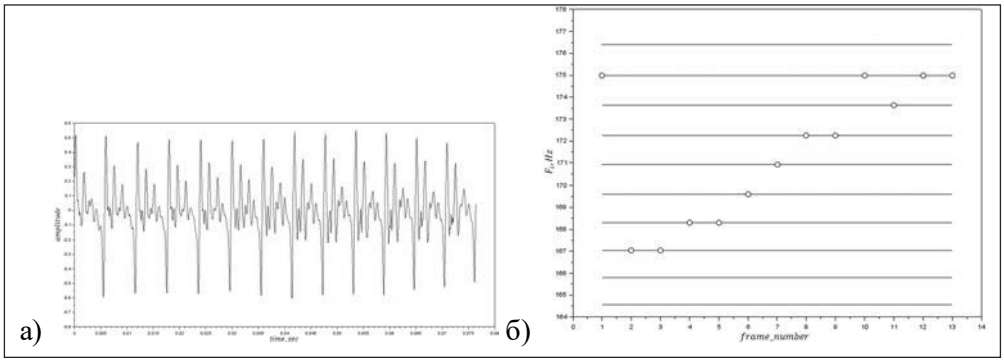

Рис. 1. Приклад вокалізованої ділянки голосу людини (а) та динаміки її основної частоти (б)

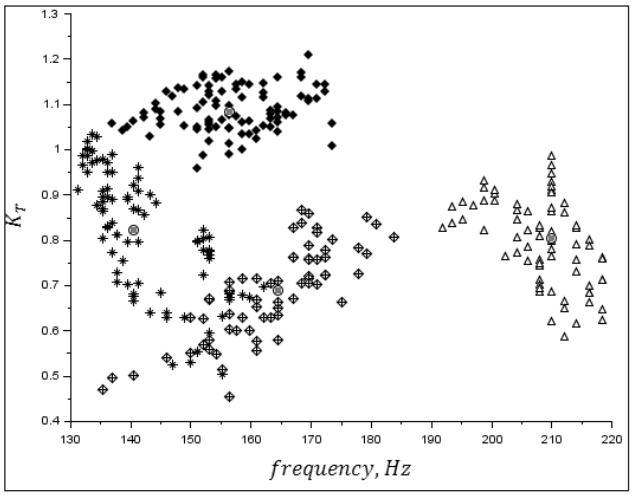

Рис. 2. Приклад формування локалізованих структур голосового сигналу

Залежно від того, яку парольну фразу використовує диктор для аутентифікації, може змінюватись конфігурація локалізованих структур, утворених у просторі характеристик голосового сигналу. Насамперед, це пов'язано з кількістю голосних звуків у фразі, яка визначатиме кількість цих структур. Також це пов'язано з комбінаціями приголосних та голосних звуків у словах. Річ у тому, що залежно від характеру супроводжуючих звуків один і той самий голосний звук може мати різне звучання з точки зору сприйняття та різну структуру з точки зору аналізу. Тому підбір парольної фрази є важливим моментом у формуванні персоніфікованого голосового еталону людини.

У роботі використовувались фрази з різною кількістю вокалізованих ділянок для ряду дикторів різної статі, віку та антропометрії. Приклади конфігурації локалізованих структур для групи з 20 дикторів, отримані при цьому, подано на рис. 3. Усі вокалізовані ділянки виділялись з акустичного запису голосового сигналу за допомогою розробленого методу локальних максимумів із подальшим подвійним кепстральних уточненням.

З огляду на отримані локалізовані структури, варто зауважити, що як парольна фраза однозначно не може бути використана фраза, в якій наявні виключно вокалізовані ділянки (рис. 3а). Причиною цього є те, що за відсутності приголосних звуків фраза «АУ» вимовляється протяжно. Як наслідок, у просторі характеристик вона формує ланцюгові структури, що відповідає значній динаміці як частоти, так і структури сигналу. Аналогічна ситуація спостерігається і у разі, коли дві вокалізовані ділянки динамічно переходять одна в одну, як у слові «Океан» (рис. 3в).

Також недоцільним є використання фраз, в яких повторюються однакові чи близькі голосні звуки. Якщо у випадку слова «Молоко» усі три звуки «О» формували різні локалізовані структури, то у випадку фрази «Усе добре» комбінація двох звуків «Е» з різних частин фрази мала однакову локалізацію (рис. 3 б, е). При подальшому аналізі вони можуть бути сприйняті як єдиний звук, що не є дійсним.

Наявність йотованих звуків на початку фрази також формує динамічну ланцюгову структуру, наприклад у слові «Їжачок» (рис. 3г). Водночас наявність йотованого голосного в середині слова, як у «Поїхали» (рис. 3д) спричинює появу динамічної ланцюгової структури не в самому йотованому звуці, а у голосному, який йому передує. Тому наявність йотованого голосного у парольній фразі також не є доцільною.

При застосуванні описаних процедур досить часто серед виділених локальних максимумів отримуються ті, що відповідають сонорним приголосним, які людина вимовляла з особливою чіткістю, зокрема «Р» на початку фрази «Справи ідуть добре». Усі ж інші локальні максимуми для даної фрази відповідають голосним звукам. Від-

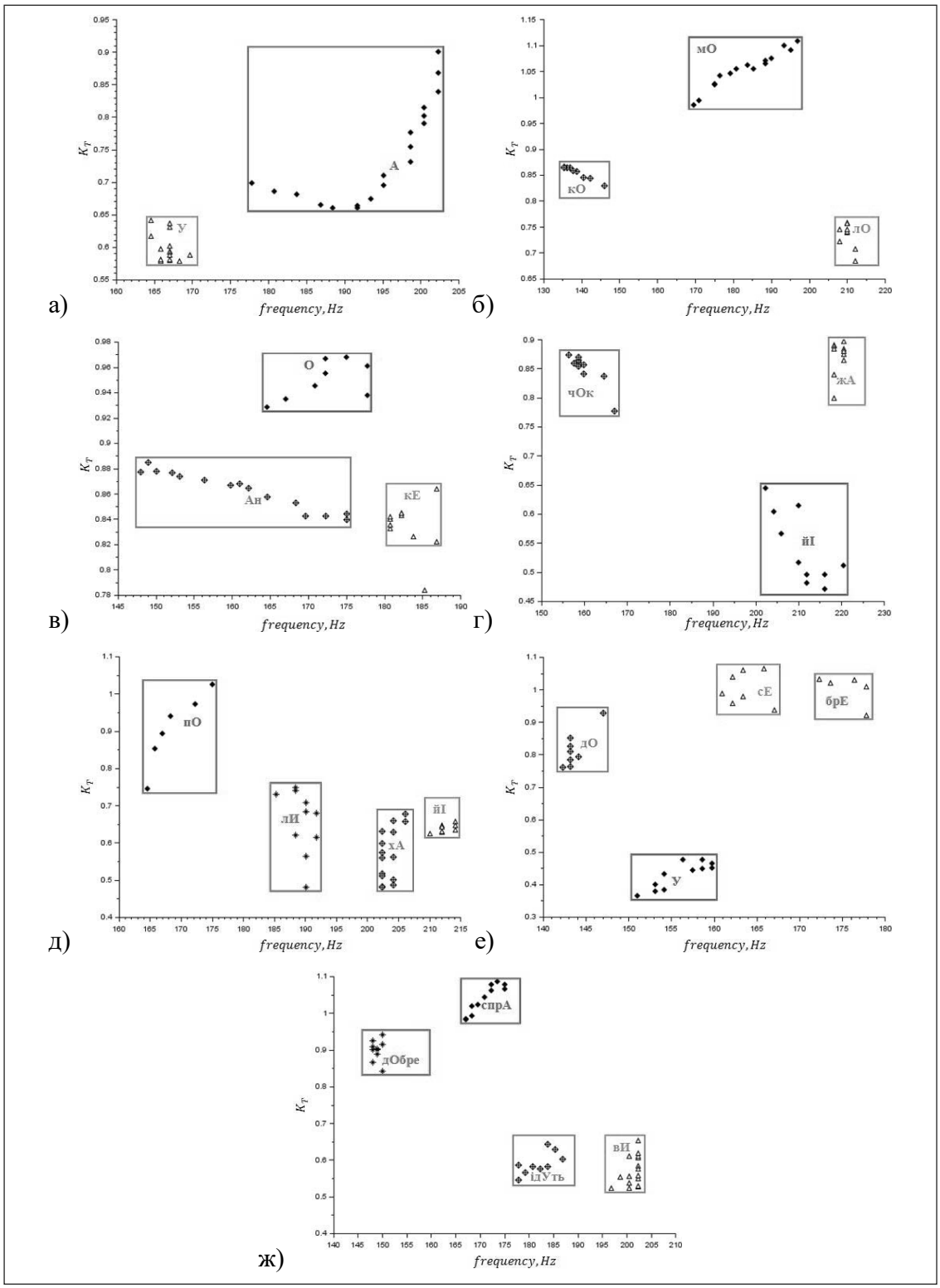

Рис. 3. Конфігурації локалізованих структур різних фраз різних дикторі

мінність спостерігається лише для звуків «І» та «Е». У випадку звуку «І» це може пояснюватись тим, що досить часто при вимові парольної фрази людина вимовляє більш зручний сонорний приголосний «Й» на місті звуку «І» – отримується фраза «Справи йдуть добре». За таких умов для отриманого логарифмованого масиву ця ділянка сприймається як сонорний приголосний.

Аналогічна ситуація спостерігається для голосного звуку «Е», який слідує після двох сонорних приголосних – «Б» та «Р». Ще однією причиною низьких показників останнього звуку може бути

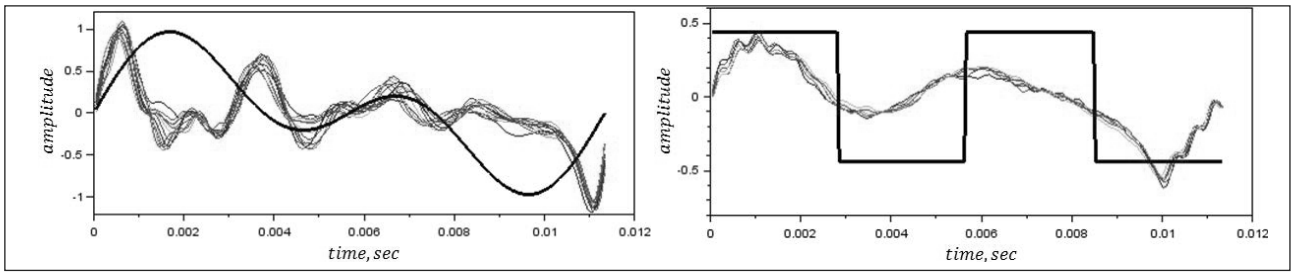

Рис. 4. Приклади базових сигналів, що аналізувались у процесі дослідження, накладені на різні звуки голосу людини (подвійна гармоніка та відповідний їй імпульсний сигнал)

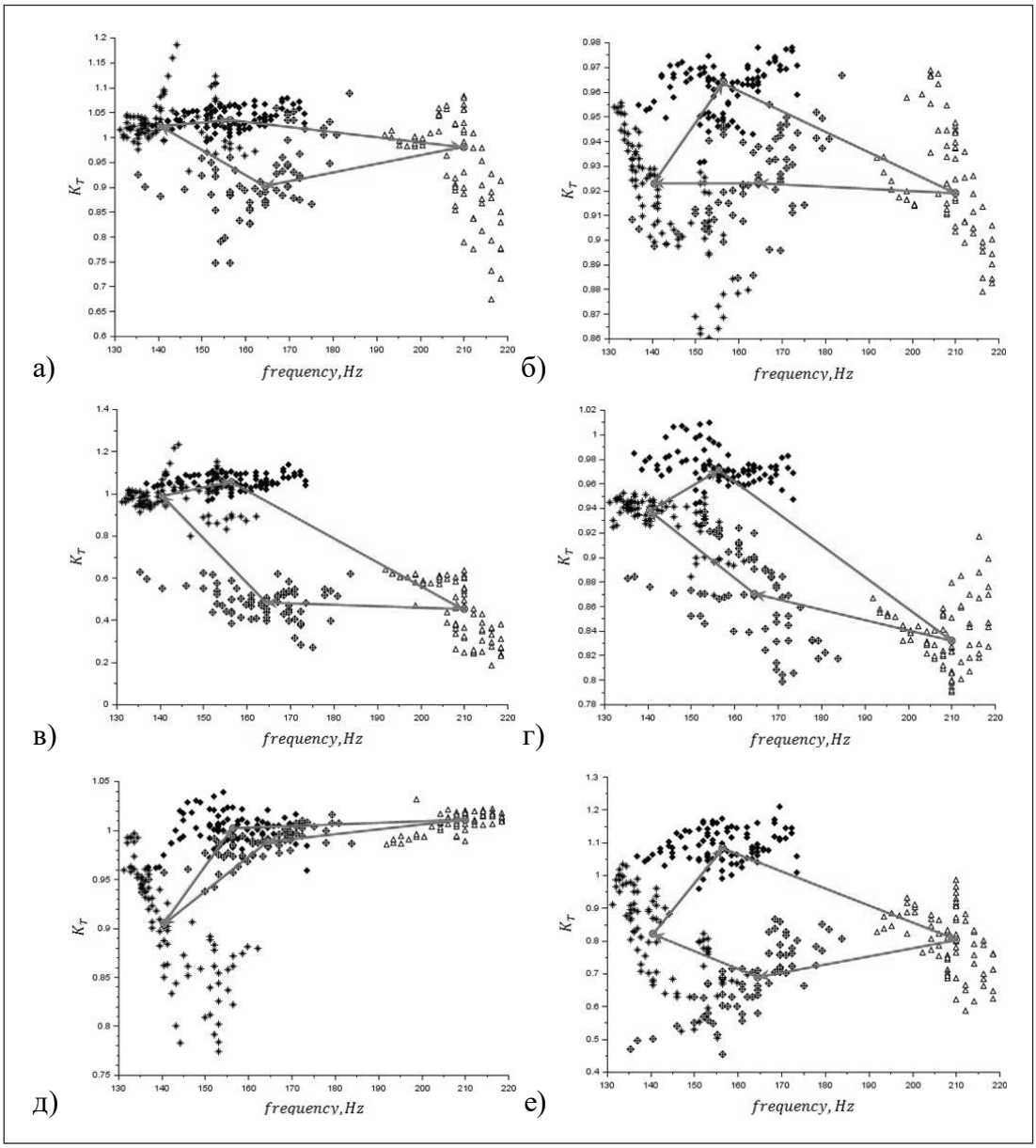

Рис. 5. Типова форма локалізованих структур одного диктора, утворених на основі різних базових сигналів

те, що наприкінці фрази звук починає переходити у фазу гасіння, голосовий апарат використовує на нього менше енергії і він за характеристиками стає подібним до перехідного процесу. З іншого боку, якщо людина спеціально робить логічний наголос на цьому звуці, характеристики його максимуму будуть на рівні інших голосних звуків парольної фрази. При цьому здебільшого голо-

Таблиця 1

Розрахунковий критерій оптимальності базового сигналу

Диктор	Одинична гармоніка	Одиничний імпульс	Дві гармоніки	Подвійний імпульс	Три гармоніки	Потрійний імпульс
1	2,280725	2,147323	1,703720	1,627341	2,124992	2,072514
2	2,717681	2,672270	1,409259	1,765154	2,938353	2,818793
3	2,174144	2,218394	1,768177	1,593370	2,007627	2,076515
4	1,157025	1,598765	1,362184	1,624920	2,495515	2,281771
5	2,696152	2,369970	2,551286	2,534524	1,126795	1,146160
6	1,135752	1,566920	1,482608	1,488141	2,588811	2,687620
7	1,042327	1,105082	1,969038	1,831148	1,848160	1,694886
8	1,036834	1,668583	1,780740	1,535735	2,718052	2,739350
9	2,583295	2,292492	1,735588	1,841599	2,022537	2,075440
10	1,096441	1,418680	2,733767	2,988863	2,121742	1,879209
11	2,417704	2,206978	1,707254	1,788234	1,136529	1,014703
12	1,501837	1,766988	2,296648	2,750775	1,691384	1,738740
13	2,105675	2,131133	1,075725	1,358237	1,590708	1,607094
14	2,126002	2,062242	1,668374	1,544751	2,710007	2,478667
15	2,745217	2,748215	1,172995	1,331794	1,950807	1,679071
16	2,461466	2,981093	2,276191	2,460632	1,977408	1,931138
17	1,154109	1,287459	1,850514	1,550112	2,885515	2,337230
18	1,315872	1,234402	2,640302	2,422446	1,884617	1,756798
19	1,219759	1,019483	2,995053	2,535778	1,768235	1,512491
20	2,831745	2,863388	2,621534	2,689634	2,000080	1,729218

сний «Е» наприкінці фрази не потрапляв до виділених вокалізованих ділянок. Усі ж інші вокалізовані ділянки, які відповідають голосним звукам «А», «И», «У» та «О», мають чіткі локальні максимуми. Вони формують у просторі характеристик компактні структури, які у більшості дикторів не мають ланцюгового характеру. З огляду на це для подальшого дослідження було прийнято рішення використовувати саме цю фразу.

У той же час для різних задач аутентифікації чи діагностики стану людини цілком можливим є використання інших парольних фраз чи звукових комбінацій, які як наявні у цьому дослідженні, так і відсутні у ньому.

Таким чином, за допомогою розробленого методу локальних максимумів із парольної фрази «Справи ідуть добре» виділяється чотири вокалізовані ділянки, які відповідають голосним звукам «А», «И», «У» та «О». Однак для формування оптимальних локалізованих структур у просторі характеристик «частота–структура» необхідно підібрати такий базовий сигнал, який дасть змогу чітко розмежувати отримані чотири структури у просторі. З цією метою було розглянуто можливість використання низки сигналів. Дослідження проводилось на гармонійних сигналах з однією, двома та трьома гармоніками та відповідних їм імпульсних сигналах з одиничною скважністю. Амплітуда кожного сигналу приводилась до максимального рівня аналізованого сигналу (рис. 4).

Для кожного аналізованого звуку у межах однієї фрази використовувався один і той самий базовий сигнал. Завдяки цьому у просторі основних характеристик голосового сигналу «частота–структура» формувались локалізовані структури різної форми та локалізації (рис. 5). Для формування кожної такої структури використовувалось від 5 до 20 записів парольної фрази.

На рис. 5. вказано середньозважені центри локалізованих структур, які утворюються у просторі характеристик, а також послідовність їх вимови.

Як критерій оптимальності базового сигналу було обрано довжину траєкторії голосу у цьому просторі характеристик. Під довжиною траєкторії голосу у цьому разі розуміється сума відстаней між двома послідовними локалізованими структурами у порядку їх вимови (А-И-У-О). При цьому значення частоти сигналу ділилось на 100, щоб звести вплив кожної характеристики до одного рівня. Результати розрахунків цього критерію для кожного з аналізованих базових сигналів для групи з 20 дикторів подано у таблиці 1.

На основі аналізу цієї характеристики очевидним є той факт, що не можна визначити один єди-

ний базовий сигнал, який був би оптимальним для усіх людей. Причиною цього є, насамперед, саме індивідуальність людського голосу. Якщо для розмежування локалізованих структур однієї людини оптимальною є одинична гармоніка, то для іншої людини це цілком може досягатись використанням подвійного імпульсу чи взагалі іншого сигналу, який не було розглянуто у процесі цього дослідження.

У той же час за допомогою програмних засобів цілком можливо реалізувати гнучке використання базового сигналу. При формуванні бази цих акустичних записів кожної людини програмний продукт може розраховувати довжину траєкторії голосу у просторі характеристик для кожного із використовуваних базових сигналів. На основі розрахункового значення цього критерію система для кожного диктора підбиратиме оптимальний базовий сигнал і використовуватиме саме його для формування структур.

Тим самим у системі аутентифікації особистості індивідуальні особливості її голосового сигналу фактично враховуватимуться двічі – і при підборі оптимального базового сигналу, і при визначенні критерію аутентифікації особистості. Розробка та дослідження цього критерію будуть проведені у подальшому. При цьому варто зауважити, що виключно оптимальний базовий сигнал не можна вважати критерієм аутентифікації, оскільки цілком можливо, що для двох чи більше різних людей оптимально буде використовувати один і той самий вид базового сигналу.

Висновки. У роботі розроблено персоніфікований еталон особистості на основі характеристик голосового сигналу. Ці характеристики виділяються з голосового сигналу за допомогою попередньо розроблених автором алгоритмів. На основі цих характеристик будуються локалізовані структури у просторі характеристик голосового сигналу, кожна з яких відповідає окремій фонемі. Досліджено можливість використання різних парольних фраз для побудови цих локалізованих структур. Показано, що найбільш доцільно використовувати парольні фрази з найбільшою кількістю різних голосних звуків, без йотованих та комбінацій двох голосних звуків поруч. З огляду на це, парольною фразою для цього дослідження обрано фразу «Справи йдуть добре». Визначено, що найбільш чіткого розмежування локалізованих структур у просторі характеристик голосового сигналу можна досягти, використовуючи різні види базових сигналів. При цьому для кожного диктора вид цього сигналу також є індивідуальним. Тому запропоновано у розробленому персоніфікованому голосовому еталоні особистості підбирати базовий сигнал на основі розробленого критерію довжини траєкторії голосу у просторі характеристик.

Список літератури:

1. Сулавко Е.А., Еременко А.В., Борисов Р.В. Генерация криптографических ключей на основе голосовых сообщений. Прикладная информатика / Journal of Applied Informatics. 2016. № 5 (65). С. 78−91.
2. Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Фатова М.В. Сегментация речевых сигналов на основе анализа особенностей распределения долей энергии по частотным інтервалам. Вестник НТУ ХПИ. 2011. № 17. С. 44−50.
3. Первушин Е.А. Обзор основных методов распознавания дикторов. Математические структуры и моделирование. 2011. Вып. 24. С. 41−54.
4. Трубина А.Ю. Компьютерная обработка речи. Задача определения личности говорящего. Перспективы развития информационных технологий. 2013. № 12. С. 233−238.
5. Мєшков О.Ю., Новіков О.О., Злепко С.М. Метод локальних максимумів для виділення вокалізованих ділянок голосового сигналу людини. Вісник Хмельницького національного університету. 2018. № 6. С. 197−210.
6. Понизов А.Г. Устройство и методика формирования тестовых акустических сигналов эквивалентных камертону для оценки качества слуха: автореф. дис. … канд. техн. наук: спец. 05.13.05 «Элементы и устройства вычислительной техники и систем управления». Томск, 2012. 20 с.

РАЗРАБОТКА ПЕРСОНИФИЦИРОВАННОГО ГОЛОСОВОГО ЭТАЛОНА ДЛЯ ЗАДАЧИ АУТЕНТИФИКАЦИИ ЛИЧНОСТИ

В статье исследуется возможность использования характеристик голосового сигнала человека для задачи аутентификации. Разработан персонифицированный голосовой эталон личности, который предусматривает формирование локализированных структур в пространстве характеристик голосового сигнала. Исследованы особенности использования разных парольных фраз и базовых сигналов для построения данных структур.

Ключевые слова: *голосовой сигнал, аутентификация, персонифицированный голосовой эталон.*

A PERSONIFIED VOICE STANDARD DEVELOPMENT FOR THE TASK OF THE PERSONAL AUTHENTICATION

The article investigates the possibility of human voice signal characteristics for the task of the authentication. A personified voice standard that involves the formation of localized structures in human voice characteristics space is developed. The peculiarities of different passphrases and base signals using for construction these structures are explored.

Key words: *voice signal, authentication, personified voice standard.*