

УДК 004.934:681.391

DOI: 10.18372/2310-5461.35.11840

О. К. Юдін, д-р техн. наук, проф.
Національний авіаційний університет
orcid.org/0000-0001-5098-7796
e-mail: kszi@ukr.net;

Р. В. Зюбіна, аспірант
Національний авіаційний університет
orcid.org/0000-0002-8654-6981
e-mail: sbt_nau@ukr.net

ОЦІНКА ЕФЕКТИВНОСТІ МЕТОДІВ ЕФЕКТИВНОЇ ШИРИНИ СПЕКТРУ ТА НАЙБІЛЬШОЇ ІНФОРМАЦІЙНОЇ ВАГИ ОСНОВНОГО ТОНУ В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ ТА АВТЕНТИФІКАЦІЇ АУДІОСИГНАЛІВ

Вступ

Розпізнавання або голосу людини — це біометричний метод в основі якого лежить аналіз індивідуальних характеристик голосу людини. Процес розпізнавання диктора включає особливості будови голосового тракту та характер поведінки індивідуума. В останні роки використання біометричного захисту отримало значний розвиток через зростаючу небезпеку кібератак на інформаційні структури. В Україні особливого статусу набула біометрія на основі відбитків пальців з введенням біометричних паспортів, а ряд атак на підприємства державного призначення змусив задуматися над існуючим рівнем захисту критичних інформаційних ресурсів. Використання систем автентифікації в умовах віддаленого зв'язку дасть можливість отримувати доступ до секретної інформації з великою економією часу на транспортування до потрібного місця. Біометрія на основі голосу відрізняється тим, що може мати безліч паролів, або взагалі від них не залежати, а збір даних може відбуватися динамічно, безпосередньо під час взаємодії з системою або за визначений інтервал часу.

Розвиток технологій на основі мовленнєвого аналізу отримав свій розвиток паралельно з розвитком технологій розпізнавання та синтезу мови в силу схожості характеристик, що використовувались та проблем пов'язаних з їх аналізом. У свій час голосовою біометрією займалися Дуглас А. Рейнолдс, Патрік Дж. Кенні, Маркел Дж.Д., Грей А. Х., Анн К. Сурдал, Ерік Келлер, Фредерік Джелинек, Ловвер Б. Т. Б., серед вітчизняних учених Шарій Т. В., Зірнєєва Г. В., Риндич Є. В. [1; 2].

Використання ідентифікації за голосом сприяє не тільки підвищенню надійності використання систем захисту критичної інформації, а й може допомогти в роботі криміналістів, банківській, військово-оборонній сферах. Саме тому

розвиток методів ідентифікації диктора за інформативними характеристиками голосу є дуже актуальною задачею.

Мета роботи — аналіз ефективності методів ефективної ширини спектру та найбільшої інформаційної ваги основного тону в задачах ідентифікації та автентифікації аудіосигналів шляхом аналізу інформативних характеристик голосу в умовах високого шуму.

Основна частина

Системи ідентифікації голосу поділяються на текстозалежні та текстонезалежні. Текстозалежні системи покликані використовувати завчасно зафіксовані фрази. В свою чергу, текстонезалежні системи ідентифікації голосу не залежать від визначених фраз і можуть працювати з вільною мовою людини, що робить їх реалізацію значно складнішою.

Розроблені методи відрізняються тим, що може працювати з будь-яким текстом. В їх основі лежить аналіз кожної інформативної складової частотного представлення вхідного сигналу, що дає можливість говорити про індивідуальні характеристики голосу кожної людини. Однією з основних задач ідентифікації голосу є визначення простору ознак, за якими буде проводитися аналіз та ідентифікація. Методи ефективної ширини спектру та найбільшої інформаційної ваги основного тону в задачах ідентифікації та автентифікації аудіосигналів лежать в амплітудно-частотній області і передбачають використання таких інформативних характеристик:

- інтенсивність, амплітуда;
- енергія;
- частота основного тону (ЧОТ);
- формантні частоти [3].

Зважаючи на те, що методи ідентифікують голос дикторів умовах високого шуму, вплив якого вносить значні зміни на більшість характеристик, то за основну характеристику приймемо частоту основного тону та значення математич-

ного сподівання кожної складової частотної спектру сигналу. У результаті аналізу існуючих методів визначення частоти основного тону, для роботи методів ефективної ширини спектру та найбільшій інформаційній ваги основного тону в задачах ідентифікації та автентифікації аудіосигналів обрано найбільш зручний метод визначення частоти основного тону, а саме метод спектрального гармонічного аналізу. В основі роботи даного методу лежить ідея спектрального аналізу та визначення максимального значення амплітуд. Частота основного тону може бути визначена шляхом знаходження частоти гармонічних складових з максимальною амплітудою та обчислення найбільшого спільного дільника цих частот для гармонічних складових. Цей дільник визначається шляхом внесення запису в частотну гістограму для кожної гармоніки. Гармоніка з найбільшою амплітудою являє собою найбільший спільний дільник частоти, і, отже, знаходиться на частоті основного тону. До переваг використання цього методу відносяться: простота реалізації, стійкість до адитивних і мультиплікативних завад, можливість регулювання деяких характеристик [4]. Нехай існує зразок голосу жінки, що підлягає ідентифікації, $S_{\text{вх1}}(t)$ та $S_{\text{вх2}}(t)$, залежать від наступних загальновідомих параметрів t, A, φ, ω . Зразок голосу $S_{\text{вх1}}(t) = S_{\text{эф}}(t)$, а $S_{\text{вх2}}(t) \neq S_{\text{эф}}(t)$, що дає можливість провести ідентифікацію для методу ефективної ширини спектру та методу найбільшої інформаційної ваги основного тону.

Незалежно від використовуваного методу на вхід системи по чергово приходять сигнали $S_{\text{вх1}}(t)$ та $S_{\text{вх2}}(t)$, які підлягають аналізу та визначенню їх належності до множини значень $DB = \{A, B\}$, де $A = \{Sm(t)_1, Sm(t)_2, Sm(t)_3 \dots Sm(t)_n\}$ — множина

зразків чоловічих голосів, а $B = \{Sw(t)_1, Sw(t)_2, Sw(t)_3 \dots Sw(t)_m\}$ — множина зразків жіночих голосів, відповідно. В експерименті використовували дванадцять аудіозаписів, серед яких вісім жіночих і чотири чоловічих.

Специфіка будови всіх систем запису аудіо сигналів припускає наявність завад створених самою системою, звуками навколишнього середовища при записі, завад в каналі зв'язку. В такому випадку оцінити ефективність роботи системи можна використовуючи співвідношення сигнал/шум (SNR).

Нехай в базі даних існує запис, зроблений голосом тієї ж людини $S_{\text{эф}}(t)$, то будемо вважати його за еталонний.

Таким чином маємо множину жіночих голосів $B = \{S_{\text{эф}}(t), Sw(t)_1, Sw(t)_2, Sw(t)_3 \dots Sw(t)_m\}$.

У результаті використання методів ідентифікації отримані такі результати:

1) імовірність вірної ідентифікації жіночого голосу $S_{\text{вх1}}(t)$ та $S_{\text{вх2}}(t)$ в множині DB становить 100 % у випадку 5,22 дБ для методу ефективної ширини спектру і 5,26 для методу найбільшої інформаційної ваги основного тону, і тільки після того починає знижуватись (рис. 1).

2) частота основного тону та обертони використовуються як основні інформативні складові в процесі ідентифікації людини за голосом і дають можливість знехтувати іншою частиною спектру.

Використання методу найбільшої інформаційної ваги основного тону дає можливість ідентифікувати голос при співвідношенні шум/сигнал в 5,21 дБ, що зменшує ефективність текстонезалежної ідентифікації всього на 2,1 %, і зовсім не впливає на ймовірність текстозалежної ідентифікації (рис. 2).

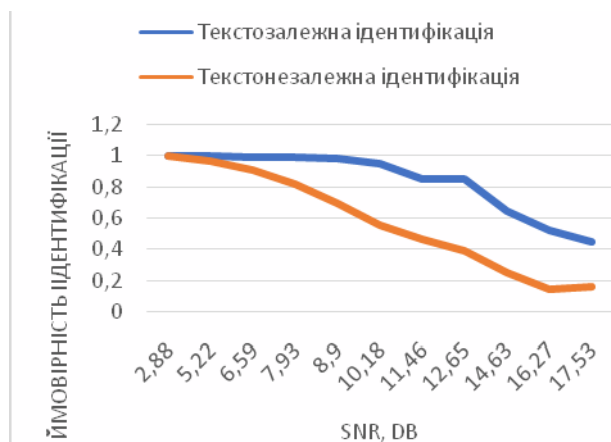


Рис. 1. Залежність ймовірності вірної ідентифікації жіночого голосу в множині DB залежно від рівня SNR методом ефективної ширини спектру

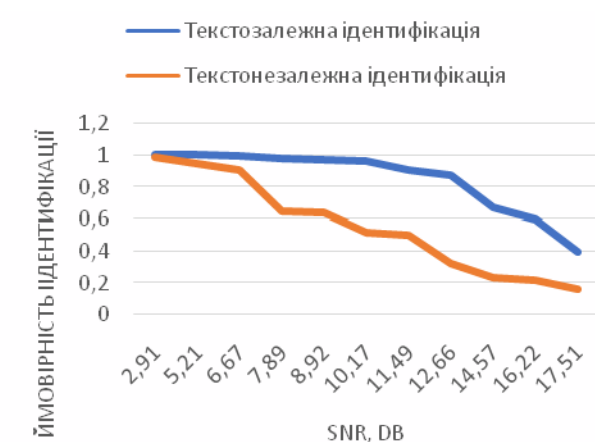


Рис. 2. Залежність ймовірності вірної ідентифікації жіночого голосу в множині DB залежно від рівня SNR при використанні методу найбільшої інформаційної ваги основного тону

Таблиця 1

Показники ідентифікації диктора методом ефективної ширини спектру

$DB = \{A, B\}$					
№ з/п	$\sigma, 10^3$	SNR, дБ	Шум перевищує сигнал (раз)	Текстозалежна ідентифікація	Текстонезалежна ідентифікація
1	13	-2,88	0,83	1	1
2	17	-5,22	1,096	1	0,97
3	20	-6,59	1,27	0,99	0,91
4	23	-7,93	1,43	0,99	0,82
5	26	-8,9	1,67	0,98	0,69
6	30	-10,18	1,91	0,95	0,55
7	35	-11,46	2,25	0,85	0,46
8	40	-12,65	2,59	0,85	0,39
9	50	-14,63	3,16	0,65	0,25
10	60	-16,27	3,82	0,52	0,14
11	70	-17,53	4,41	0,45	0,16

Таблиця 2

Показники ідентифікації диктора методом найбільшої інформаційної ваги основного тону

$DB = \{A, B\}$					
№ з/п	$\sigma, 10^3$	SNR, дБ	Шум перевищує сигнал (раз)	Текстозалежна ідентифікація	Текстонезалежна ідентифікація
1	13	-2,91	0,83	1	0,99
2	17	-5,21	1,096	1	0,95
3	20	-6,67	1,27	0,99	0,91
4	23	-7,89	1,43	0,98	0,65
5	26	-8,92	1,67	0,97	0,64
6	30	-10,17	1,91	0,96	0,52
7	35	-11,49	2,25	0,9	0,5
8	40	-12,66	2,59	0,87	0,32
9	50	-14,57	3,16	0,67	0,23
10	60	-16,22	3,82	0,6	0,22
11	70	-17,51	4,41	0,4	0,16

Таблиця 3

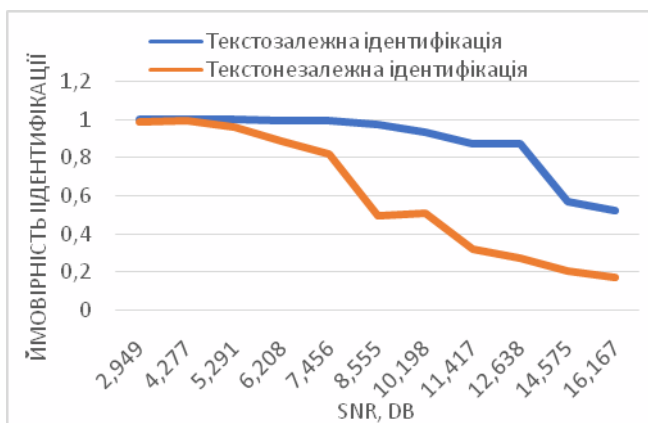
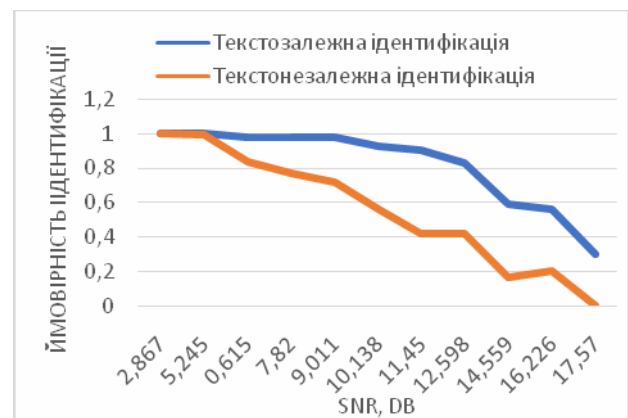
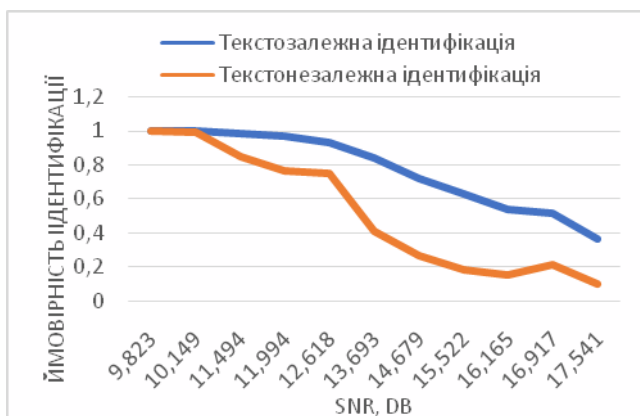
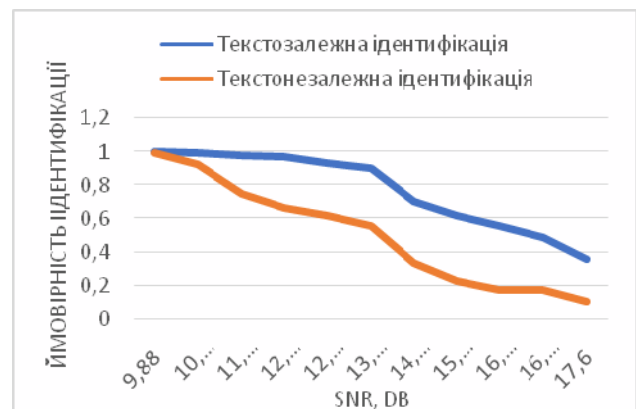
Показники ідентифікації диктора методом ефективної ширини спектру

$DB = \{B\}$					
№ з/п	$\sigma, 10^3$	SNR, дБ	Шум перевищує сигнал (раз)	Текстозалежна ідентифікація	Текстонезалежна ідентифікація
1	13	-2,949	0,83	1	0,99
2	15	-4,277	0,96	1	0,996
3	17	-5,291	1,096	1	0,962
4	19	-6,208	1,25	0,99	0,889
5	22	-7,456	1,43	0,99	0,82
6	25	-8,555	1,62	0,97	0,492
7	30	-10,198	1,91	0,93	0,508
8	35	-11,417	2,25	0,87	0,32
9	40	-12,638	2,59	0,87	0,27
10	50	-14,575	3,16	0,565	0,203
11	60	-16,167	3,82	0,52	0,17

Таблиця 4

Показники ідентифікації диктора методом найбільшої інформаційної ваги основного тону

$DB = \{B\}$					
№ з/п	$\sigma, 10^3$	SNR, дБ	Шум перевищує сигнал (раз)	Текстозалежна ідентифікація	Текстонезалежна ідентифікація
1	13	-2,867	0,83	1	1
2	17	-5,245	1,096	1	0,991
3	20	-6,15	1,27	0,98	0,836
4	23	-7,82	1,43	0,98	0,768
5	26	-9,011	1,67	0,98	0,72
6	30	-10,138	1,91	0,929	0,56
7	35	-11,45	2,25	0,9	0,415
8	40	-12,598	2,59	0,83	0,416
9	50	-14,559	3,16	0,59	0,167
10	60	-16,226	3,82	0,562	0,199
11	70	-17,57	4,41	0,3	0

Рис. 3. Залежність імовірності вірної ідентифікації жіночого голосу в множині B залежно від рівня SNR методом ефективної ширини спектруРис. 4. Залежність імовірності вірної ідентифікації жіночого голосу в множині B залежно від рівня SNR при використанні методу найбільшої інформаційної ваги основного тонуРис. 5. Залежність імовірності вірної ідентифікації жіночого голосу в множині A залежно від рівня SNR методом ефективної ширини спектруРис. 6. Залежність імовірності вірної ідентифікації жіночого голосу в множині A залежно від рівня SNR при використанні методу найбільшої інформаційної ваги основного тону

Зразки чоловічих та жіночих голосів знаходяться в різних частотних діапазонах, тому доцільно було розділити їх на дві підмножини A та B . Результат експерименту показав, що точна ідентифікація голосу можлива коли SNR становить 5,29 дБ для текстозалежного випадку і 4,27 дБ для текстонезалежного для методу ефективної ширини спектру. Визначена динаміка прослідковується і для методу найбільшої інформаційної ваги основного тону, відповідно показники становлять 5,7 та 5,24 дБ. Отже, виходячи із вищесказаного, можна зробити висновок, що розділення бази даних на дві множини з різними групами зразків дасть можливість підвищити ефективність роботи системи в середньому на 27 % для різних значень SNR.

Перевірка роботи системи для множини B підтвердила вірність ідентифікації і необхідність розділення множини DB на дві підмножини. Ідентифікація жіночого голосу серед більш низьких чоловічих можлива навіть в умовах коли

шум перевищує сигнал більш ніж у 10 дБ, як для текстозалежних, так і для текстонезалежних методів.

Такі показники ідентифікації мовленнєвого сигналу демонструють універсальність роботи розроблених методів ефективної ширини спектру і найбільшої інформаційної ваги основного тону, та дозволяють обирати між текстозалежною і текстонезалежною ідентифікацією залежно від обраного порогу прийняття рішення.

Важливим показником ефективності системи є час затрачений на проведення аналізу, обробки та прийняття рішення. В основі експерименту використовується сім дванадцять сигналів довжиною від 400 до 750 мс.

Час затрачений на зчитування, аналіз, обробку і прийняття рішення займає в середньому 2,46 с для методу ефективної ширини спектру, а для методу найбільшої інформаційної ваги основного тону, в середньому 2,29 дБ, що дає вииграш у часі розміром 7 %.

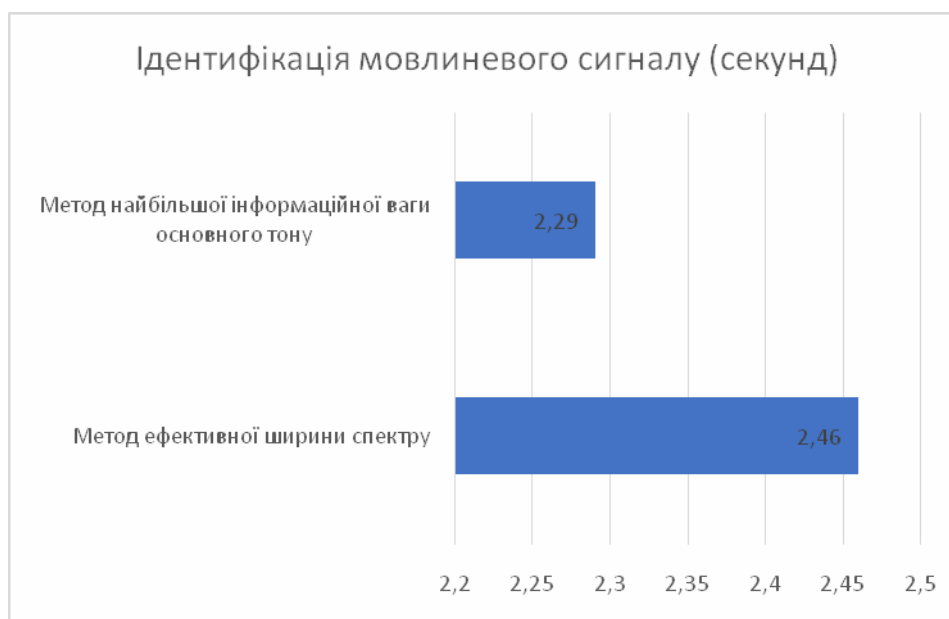


Рис. 7. Час роботи методу ефективної ширини спектру та методу найбільшої інформаційної ваги основного тону

Висновки

Проведено оцінку ефективності розроблених методів ефективної ширини спектру та найбільшої інформаційної ваги основного тону в задачах ідентифікації аудіосигналів. У результаті експерименту проведеного для трьох груп мовних сигналів було визначено, що метод ефективної ширини спектру дозволяє визначити диктора з імовірністю 100 %: при значенні SNR 4,27 дБ для текстонезалежної ідентифікації та 5,29 дБ для текстозалежної в умовах коли $S_{\text{вх1}}(t) \in B$; якщо $S_{\text{вх1}}(t) \in DB$, то значення SNR становлять 5,22 дБ

і 5,26 дБ відповідно. Отже, розподіл голосів на групи дає можливість підвищити ефективність роботи системи в середньому на 27 % для різних значень SNR, а використання методу найбільшої інформаційної ваги основного тону дає вииграш по часу на 110 мс порівняно з методом ефективної ширини спектру, але при цьому зменшується поріг прийняття рішення.

ЛІТЕРАТУРА

1. Yudin O., Frolov O., Ziubina R. Quantitative quality indicators of the invariant spatial method of compressing video data // Problems of

Infocommunications Science and Technology (PIC S&T), 2015 Second International Scientific-Practical Conference. — IEEE, 2015. — С. 227–229. doi.org/10.1109/infocommst.2015.7357320

2. Юдін О. К., Фролов О. В., Бойко Ю. П. Організація систем підтримки прийняття рішень для управління в кризових ситуаціях // Наукоємні технології. — 2015. — Т. 27. — №. 3. — С. 244–249. doi.org/10.18372/2310-5461.27.9402.

3. Хроматиди А. Ф. Исследование психофизиологического состояния человека на основе эмоциональных признаков речи // Таганрог: диссертация на соискание уч. ст. канд. техн. наук. — 2005. — №. 2005. — С. 154.

4. Юдін О. К., Зюбіна Р. В. Класифікація методів ідентифікації частоти основного тону // Наукоємні технології. — 2017. — Т. 33. — №. 1. doi.org/10.18372/2310-5461.33.11553.

Юдін О. К., Зюбіна Р. В.

ОЦІНКА ЕФЕКТИВНОСТІ МЕТОДІВ ЕФЕКТИВНОЇ ШИРИНИ СПЕКТРУ ТА НАЙБІЛЬШОЇ ІНФОРМАЦІЙНОЇ ВАГИ ОСНОВНОГО ТОНУ В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ ТА АВТЕНТИФІКАЦІЇ АУДИОСИГНАЛІВ

Визначено, що ідентифікація за голосом сприяє підвищенню надійності використання систем захисту критичної інформації. В результаті аналізу існуючих методів визначення частоти основного тону, обрано найбільш зручний метод визначення частоти основного тону, а саме метод спектрального гармонічного аналізу. Проведено оцінку ефективності розроблених методів ефективної ширини спектру та найбільшої інформаційної ваги основного тону в задачах ідентифікації аудіосигналів. У результаті експерименту проведеного для трьох груп мовних сигналів було визначено, що метод ефективної ширини спектру дозволяє визначити диктора з імовірністю 100 % : при значенні SNR 4,27 дБ для текстонезалежної ідентифікації та 5,29 дБ для текстозалежної в умовах, коли вхідний сигнал належить підмножині бази даних; якщо вхідний сигнал належить базі даних не розділеній на підмножини, то значення SNR складають 5,22 дБ і 5,26 дБ відповідно.

Ключові слова: частота основного тону; методи ідентифікації диктора; розпізнавання сигналів.

Юдин А. К., Зюбина Р. В.

ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ ЭФФЕКТИВНОЙ ШИРИНЫ СПЕКТРА И НАИБОЛЬШЕГО ИНФОРМАЦИОННОГО ВЕСА ОСНОВНОГО ТОНА В ЗАДАЧАХ ИДЕНТИФИКАЦИИ И АУТЕНТИФИКАЦИИ АУДИОСИГНАЛОВ

Определено, что идентификация по голосу способствует повышению надежности использования систем защиты критической информации. В результате анализа существующих методов определения частоты основного тона, выбран наиболее удобный метод определения частоты основного тона, а именно метод спектрального гармонического анализа. Проведена оценка эффективности разработанных методов эффективной ширины спектра и наибольшего информационного веса основного тона в задачах идентификации аудиосигналов. В результате проведенного эксперимента для трех групп речевых сигналов было определено, что метод эффективной ширины спектра позволяет определить диктора с вероятностью 100 %: при значении SNR 4,27 дБ для текстонезависимой идентификации и 5,29 дБ для текстозависимой в условиях, когда входной сигнал принадлежит одному из подмножеств базы данных; если входной сигнал принадлежит базе данных не разделенной на подмножества, то значение SNR составляют 5,22 дБ и 5,26 дБ соответственно.

Ключевые слова: частота основного тона; методы идентификации диктора; распознавание сигналов.

Yudin O. K., Ziubina R. V.

EVALUATION OF EFFECTIVENESS OF METHODS FOR EFFECTIVE SPECTRUM WIDTH INFORMATION AND THE BIGGEST WEIGHT OF THE PITCH TONE IN THE TASKS OF IDENTIFICATION AND AUTHENTICATION OF AUDIO SIGNALS

It is determined that identification by voice contributes to the reliability of the use of protection systems critical information. The analysis of existing methods of determining the fundamental tone frequency that you selected the most convenient method for the determination of the fundamental frequency, namely the method of spectral harmonic analysis. An evaluation of the effectiveness of the developed methods effective spectral width and the largest scales the information of the fundamental tone in the tasks of identification of audio signals. As a result of the experiment for three groups of speech signals, was determined that the method for effective spectrum width allows to identify the speaker with a probability of 100 %, where a value SNR of 4.27 dB for extensibility identification and of 5.29 dB for textualizing in conditions, when the input signal belongs to one of the subsets of the database; if the input signal belongs to a database, not divided into subsets, the value of SNR 5.22 dB is amount of 5.26 dB, respectively.

Keywords: pitch frequency; methods of speaker identification; signal recognition.

Стаття надійшла до редакції 28.08.2017 р.
Прийнято до друку 01.09.2017 р.
Рецензент — д-р техн. наук, проф. Бараннік В. В.