

УДК 057.087.1:621.391.26

# ОЦІНКА ЧАСТОТИ ОСНОВНОГО ТОНУ ГОЛОСОВОГО СИГНАЛУ КОРИСТУВАЧА СИСТЕМИ АВТЕНТИФІКАЦІЇ



Є.Є. КУЦЕНКО, М.С. ПАСТУШЕНКО

Харківський національний університет радіоелектроніки

**Abstract** – The directions for improving the quality characteristics of voice authentication systems in different access systems are analyzed and investigated in the article. One of the main areas of enhancing the quality characteristics of these user authentication systems is related to the use of phase information of the voice signal. The object of the study is the process of digital signal processing in voice authentication systems. The relevant scientific task of investigating new procedures to refine estimates of pitch frequency obtained on the basis of the analysis of the amplitude-frequency spectrum is solved. The estimates were refined based on the use of phase data of a voice signal, as well as the estimation of the pitch frequency in the process of obtaining kepsral coefficients. The purpose of this work is to investigate the effect of phase data on the accuracy of estimating the pitch frequency in the process of forming a user pattern. The results have been obtained in the process of statistical analysis of the simulation results using model and experimental voice data of the user authentication system. The phase data of the voice signal allow to obtain adequate and reliable estimates in the process of spectral analysis. However, in the presence of errors associated with gross errors, such as the acceptance of the pitch of the frequencies of the first or second formants, preference should be given to the estimate obtained in the process of calculating kepsral coefficients. The influence of sample length and their overlap in the process of calculating kepsral coefficients is investigated. The presented research results should be used in voice authentication systems, improvement of speech recognition systems, as well as in solving problems of speaker identification.

**Анотація** – У статті аналізуються і досліджуються напрямки підвищення якісних характеристик систем голосової автентифікації в різних системах доступу. Один з основних напрямків підвищення якісних характеристик цих систем автентифікації користувача пов'язаний з використанням фазової інформації голосового сигналу. Об'єктом дослідження є процес цифрової обробки сигналів у системах голосової автентифікації. Вирішується актуальне наукове завдання дослідження нових процедур для уточнення оцінок частоти основного тону, отриманих на основі аналізу амплітудно-частотного спектра. Уточнення оцінок проводилося на основі використання фазових даних голосового сигналу, а також оцінки частоти основного тону в процесі отримання кепстральних коефіцієнтів. Мета даної роботи – дослідження впливу фазових даних на точність оцінки частоти основного тону в процесі формування шаблону користувача. Дані отримані в процесі статистичного аналізу результатів моделювання з використанням модельних і експериментальних голосових даних користувача системи автентифікації. Фазові дані голосового сигналу дозволяють отримувати адекватні та достовірні оцінки в процесі спектрального аналізу. Однак при наявності помилок, пов'язаних з грубими помилками, наприклад, прийняття за оцінку частоти основного тону частот першої або другої форманти, перевагу слід віддавати оцінці, отриманій у процесі розрахунку кепстральних коефіцієнтів. Досліджено вплив довжини семплів і їх перекриття в процесі розрахунку кепстральних коефіцієнтів. Представлені результати досліджень доцільно використовувати в системах голосової автентифікації, вдосконаленні систем розпізнавання мови, а також при вирішенні завдань ідентифікації диктора.

## Вступ

В останні десятиліття досягнення науки і новітні технології як ніколи раніше визначають динаміку економічного зростання, рівень добробуту населення, конкурентоспроможність держави в світовому співтоваристві, ступінь забезпечення її національної безпеки та рівноправної інтеграції в світову економіку. Стрімкий розвиток і широке використання сучасних інформаційно-телекомунікаційних систем ознаменували перехід людства від індустріального суспільства до суспільства інформаційного, в основі якого лежать новітні системи комунікацій. Кількість, технічний рівень і доступність інформаційних систем вже зараз визначають ступінь розвиненості краї-

ни та її статус у світовому співтоваристві, а в недалекому майбутньому, безсумнівно, стануть вирішальним показником цього статусу.

Разом з тим, процес інформатизації світової спільноти породжує комплекс негативних явищ. Дійсно, висока складність і одночасно вразливість всіх систем, на яких базуються регіональні, національні та світовий інформаційні простори, а також фундаментальна залежність від їх стабільності державних інфраструктур призводять до виникнення принципово нових загроз.

У зв'язку з широким розповсюдженням розподілених систем у всіх сферах людської діяльності гостро стоїть завдання забезпечення інформаційної безпеки в таких системах. Одним з основних заходів щодо захисту фінансових коштів, інформаційних даних і обчислювального ресурсу є забезпечення надійної автентифікації користувача.

На даний момент існує безліч підходів до автентифікації і ще більше реалізацій цих підходів. При цьому не всі класичні рішення завдання автентифікації підходять для використання в розподілених системах. А різні типи систем пред'являють свої 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. У зв'язку з цим ведеться безперервна робота в галузі дослідження і розробки методів автентифікації. Постійно з'являються нові і удосконалюються існуючі алгоритми, спрямовані на забезпечення захищеної автентифікації користувачів.

Тому все більш актуальною стає проблема автентифікації користувачів, що мають доступ до громадських та особистих інформаційних ресурсів. Особливо важлива ця проблема для відкритих, масових телекомунікаційних та інформаційних систем. Одним з найбільш перспективних напрямків захисту подібних систем від несанкціонованих впливів є біометричні методи ідентифікації користувачів. Однак, незважаючи на всю привабливість, даний підхід пов'язаний з низкою серйозних проблем.

Тому останнім часом багато досліджень проводиться в області застосування біометричних систем автентифікації. Серед біометричних систем особливе місце займає голосова автентифікація, яка відрізняється простотою та зручністю. Але, як і всі біометричні системи, голосова автентифікація має низькі якісні характеристики. У зв'язку з цим в області голосової автентифікації проводяться інтенсивні дослідження, про що свідчать роботи [1-4].

У сучасних системах голосової автентифікації (СГА) реєструється амплітудна інформація полігармонічного нестационарного голосового сигналу користувача. Автентифікація користувача здійснюється в процесі аналізу амплітудно-частотного спектра матеріалів реєстрації [2]. Основні зусилля дослідників при цьому зосереджені на пошуку нових або удосконалення існуючих процедур оцінки шаблонів (набору ознак – частоти основного тону, формантних даних, коефіцієнтів кепстру, мел-частотних кепстральних коефіцієнтів, коефіцієнтів лінійного передбачення та їхні динамічні характеристики й ін.) користувача, а також на розробці вирішальних правил. Найбільш популярні серед останніх такі процедури прийняття рішень – методи

гаусових сумішей (Gaussian Mixture Model, GMM) і опорних векторів (Support Vector Machine, SVM). Для таких цілей також використовуються штучні нейронні мережі і приховані Марковські моделі (Hidden Markov Models, HMM).

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 спектра алгоритмами врахування фазових даних голосових сигналів, у тому числі з метою підвищення відношення сигнал/шум.

Таким чином, у даний час існує інший шлях підвищення якісних показників СГА, який базується на використанні фазової інформації голосового сигналу користувача. Давно відомо [5], що фаза є більш інформативним параметром сигналу, проте в СГА вона традиційно ігнорується [2].

Обумовлено це тим, що для отримання фазової інформації необхідні додаткові обчислювальні й алгоритмічні ресурси, які не завжди були доступні в зазначених додатках. Крім цього, є деякі особливості оцінки попередньої обробки та використання фазових даних. Слід зазначити, що в даний час відсутній досвід і практика використання фази сигналу стосовно завдань голосової автентифікації.

Підтвердженням цьому є те, що відомо лише обмежена кількість робіт, де фазові дані використовувалися при обробці мовних сигналів. Наприклад, у [6] зазначена актуальність використання фазової інформації при обробці мовних даних, а в [7] використовувалася фаза для уточнення частотних характеристик оброблюваних голосових даних. У [8] виконано порівняльний аналіз процедур оцінки фазових співвідношень між коливаннями основного тону й обертонів мовних сигналів, які автори пропонують використовувати для вирішення завдань розпізнавання звуків мови та ідентифікації дикторів.

Зазначене вище підкреслює актуальність досліджень оцінки впливу фазових даних на якісні характеристики процедур голосової автентифікації. Фазові дані в голосовій автентифікації можуть використовуватися за кількома напрямками:

- підвищення відношення сигнал/шум матеріалів реєстрації (відомий напрямок використання фази в радіолокації та радіозв'язку);
- підвищення якості формування ознак для традиційно використовуваних шаблонів, наприклад, частоти основного тону, формантної інформації і т.д.;
- розробка нових процедур формування елементів шаблонів на основі фазових даних [9].

Мета даної роботи – дослідження впливу фазових даних на точність оцінки частоти основного тону в процесі формування шаблону користувача. Об'єктом дослідження є процес цифрової обробки голосових сигналів.

## **I. Постановка задачі**

Відомо [2, 10], що найбільш важливий фактор індивідуальності голосу – це частота основного тону  $F_0$ , за нею йдуть формантні частоти, розмір флуктуацій  $F_0$  і на-

хил спектру. Прийнято вважати, що на ділянках вокалізованих звуків мовний тракт людини порушується періодичним коливанням зв'язок. Період цього коливання називають частотою (періодом) основного тону. Ця величина є індивідуальною характеристикою диктора. Вона може змінюватися в залежності від емоційного забарвлення мови, але в досить вузьких межах. При параметричному кодуванні мови припускають, що частота основного тону людини лежить в межах 80 - 400 Гц.

Деякі автори стверджують, що ознаки, пов'язані з  $F_0$ , забезпечують найкращу роздільність голосів, а за ними слідує енергія сигналу та тривалість сегментів. Логарифмічні подання  $F_0$  більш інформативні, ніж сама частота основного тону.

Методи визначення (оцінки) частоти основного тону можна розділити на наступні групи: амплітудна (часова) селекція, кореляційні методи, частотна селекція (спектральний аналіз) і кепстральний аналіз. Точніші результати дає спектральний аналіз, який пов'язаний з розрахунком амплітудно-частотного спектру й оцінкою максимуму спектральної щільності. Остання відповідає частоті основного тону. Однак у деяких випадках максимум спектральної щільності зміщується на частоту формант [11]. Тому актуальною є наукова задача дослідження нових процедур для уточнення оцінок частоти основного тону, отриманих на основі аналізу амплітудно-частотного спектра.

## II. Методика і результати досліджень

У даній роботі будемо досліджувати оцінки частоти основного тону на основі аналізу амплітудно-частотного та фазочастотного спектрів, а також кепстрального аналізу.

Спочатку проведемо аналіз модельного сигналу, який містить послідовність двох гармонік різної амплітуди та частоти. При цьому перша гармоніка мала несуттєву частоту 100 Гц, а друга – 500 Гц. Фрагмент аналізованого сигналу наведений на рис. 1.

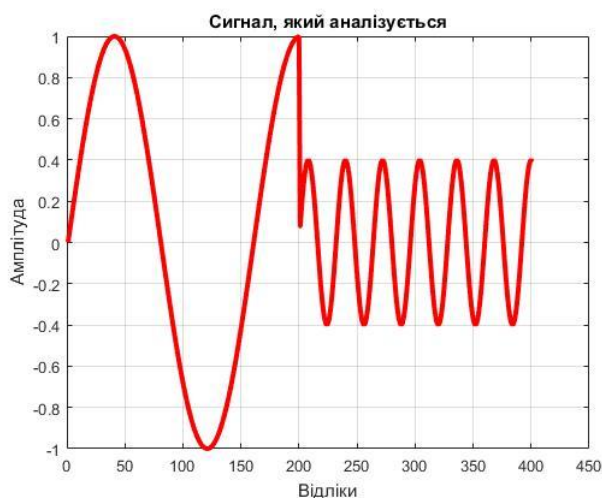


Рис. 1. Фрагмент аналізованого сигналу

Аналіз амплітудного (рис. 2) та фазового (рис. 3) спектрів свідчить, що вони не дозволяють виділити перший максимум (аналог частоти основного тону).

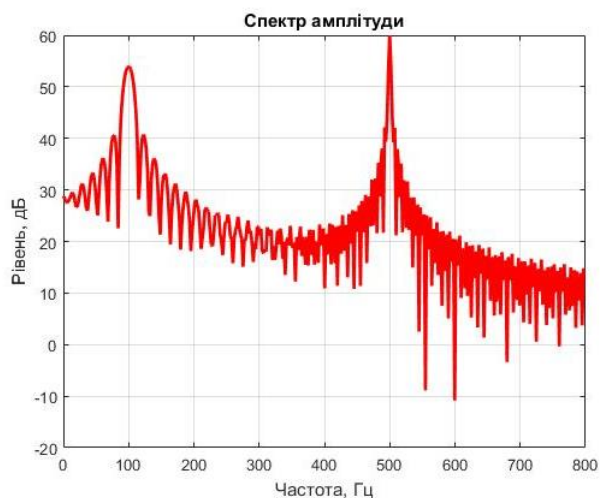


Рис. 2. Амплітудний спектр аналізованого сигналу

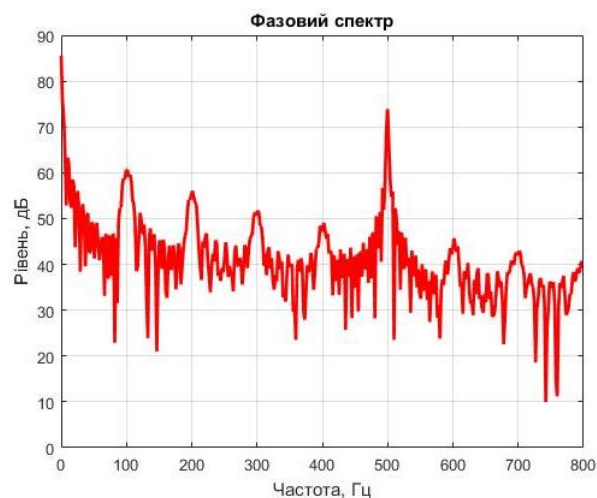


Рис. 3. Фазовий спектр аналізованого сигналу

Інший шлях отримання наближеної оцінки частоти основного тону може бути пов'язаний з розрахунком кепстральних або мел-частотних кепстральних коефіцієнтів (Mel Frequency Cepstral Coefficient - MFCC), які, як правило, входять ознаками до шаблону користувача. Як відомо, кепстральні коефіцієнти визначаються відповідно до схеми, представленій на рис. 4. На даному рисунку прийняті наступні позначення: FFT – блок швидкого перетворення Фур'є сигналу; LOG – блок логарифмування спектру; IFFT – блок зворотного швидкого перетворення Фур'є.



Рис. 4. Загальна схема кепстрального аналізу сигналу

Таким чином, кепстральні коефіцієнти є результатом застосування зворотного перетворення Фур'є до логарифму енергетичного спектру. Розрахунок цих коефіцієнтів здійснюється на семплах сигналу, які по тривалості становлять кілька десятків мілісекунд. При цьому семпли вибираються з деяким перекриттям. Результат кожного зворотного перетворення дозволяє отримати оцінку максимуму частоти в семпли. Як правило, розраховується приблизно 40 коефіцієнтів, а значить і стільки ж оцінок частоти. Шляхом усереднення результатів можемо отримати більш точну оцінку максимуму частоти. Зазначені процедури обробки модельного сигналу дозволили отримати оцінку приблизно 200 Гц, що ближче до першого максимуму.

Тепер зробимо обробку реального голосового сигналу, який аналізувався в [10]. Спектральний аналіз амплітудних і фазових спектрів реального голосового сигналу

користувача дозволив отримати оцінку частоти основного тону в 240 Гц [10]. Результати обробки цього сигналу за допомогою запропонованих процедур оцінки частоти основного тону представлені на рис. 5 і 6. Коефіцієнт перекриття дорівнював 0.75. На рис. 5 представлені оцінки математичного очікування (МО). При цьому суцільною лінією показані результати обробки амплітудних даних, а штриховою – фазових. На рис. 6 представлені оцінки середнього квадратичного відхилення (СКВ) для амплітудного та фазового кепстрів.

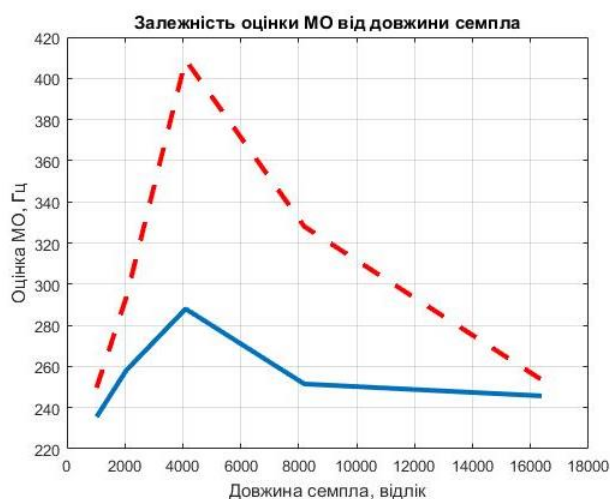


Рис. 5. Залежність оцінки МО від довжини семпла

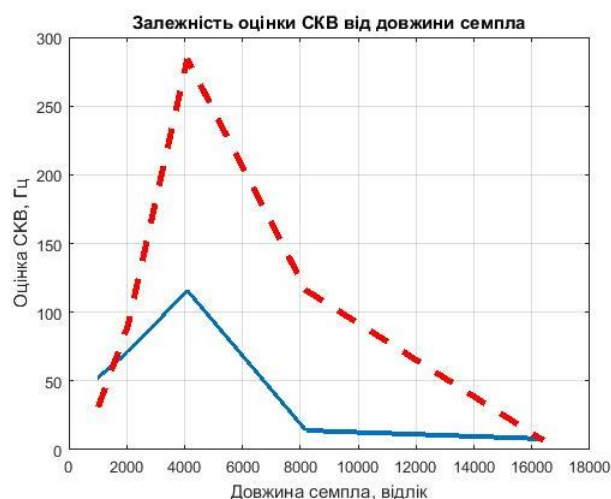


Рис. 6. Залежність оцінки СКВ від довжини семпла

Аналіз наведених рисунків дозволяє зробити висновок, що даний варіант оцінки частоти основного тону дозволяє вибрати параметри для отримання надійної оцінки. Досліджено вплив перекриття семплів у діапазоні від 0,5 до 0,85. Характер залежностей оцінок МО та СКВ відповідає результатам, представленим на рис. 5 і 6 відповідно.

## Висновки

До складу більшості шаблонів користувача систем голосової автентифікації входять в якості обов'язкових параметрів частота основного тону, кепстральні коефіцієнти і ряд інших ознак. Частота основного тону дозволяє вирішувати наступні завдання: розпізнавання емоцій, визначення статі, сегментація аудіо з декількома голосами і поділі мови на фрази.

У зв'язку з цим в роботі розглядалася актуальне наукове завдання дослідження нових процедур для уточнення оцінок частоти основного тону отриманих на основі аналізу амплітудно-частотного спектра. Уточнення оцінок проводилося на основі використання фазових даних голосового сигналу, а також оцінки частоти основного тону в процесі отримання кепстральних коефіцієнтів.

Результати отримані в процесі статистичного аналізу результатів моделювання з використанням модельних і експериментальних голосових даних користувача системи автентифікації.

Фазові дані голосового сигналу дозволяють отримувати адекватні і достовірні оцінки в процесі спектрального аналізу. Однак, при наявності помилок пов'язаних з грубими промахами, наприклад, прийняття за оцінку частоти основного тону частот першої або другої формант, перевагу слід віддавати оцінці, отриманої в процесі розрахунку кепстральних коефіцієнтів. Досліджено вплив довжини семплів і їх перекриття в процесі розрахунку кепстральних коефіцієнтів.

Подальші дослідження доцільно проводити в напрямку оцінки якості формування ознак для традиційно використовуваних шаблонів з урахуванням фази голосового сигналу, а також розробки нових процедур формування елементів шаблонів на основі фазових даних.

### Список літератури:

1. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. М: Радио и связь, 1981. 224 с.
2. Beigi H. Fundamentals of Speaker Recognition. NY: Springer, 2011. 1029 p.
3. ISO/IEC 2382-37:2012 Information technology – Vocabulary – Part 37: Biometrics. URL: <https://www.iso.org/ru/standard/55194.html>
4. Болл Р.М., Коннел Дж. Х., Панканти Ш., Ратха Н.К., Сеньор Э.У. Руководство по биометрии. М.: Техносфера, 2007. 368 с.
5. Oppenheim A.V., Lim J.S. The Importance of Phase in Signals: Article in Proceeding of the IEEE, 1981. № 69(5). P. 529-541.
6. Paliwal K. Usefulness of phase in speech processing. Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan, Feb. 2003. P. 1-6.
7. Paliwal K., Atal B. Frequency-related representation of speech. In: Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-2003), 2003. P. 65–68
8. Борисенко С.Ю., Воробьев В.И., Давыдов А.Г. Сравнение некоторых способов анализа фазовых соотношений между квазигармоническими составляющими речевых сигналов. Сборник трудов 1-ой Всероссийской акустической конференции. 2004. С. 2-7.
9. Wu Z. , Kinnunen T., Chng E., Li H., Ambikairajah E. A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case. In Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference. IEEE, 2012. P. 1-5.
10. Пастушенко Н.С., Педро В.Г., Файзулаева О.Н. Исследование информативности фазовых данных голосового сигнала пользователя системы аутентификации. Проблемы телекоммуникаций. 2018. №1(22). С. 67-74. URL: [http://pt.journal.kh.ua/2018/181\\_pastushenko\\_voice.pdf](http://pt.journal.kh.ua/2018/181_pastushenko_voice.pdf).
11. Gerhard D. Pitch extraction and fundamental frequency: History and current techniques. Regina, Canada: Department of Computer Science, University of Regina. 2003. 22 p. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.834&rep=rep1&type=pdf>