

УДК 057.087.1:621.391.26

ВПЛИВ ПРОЦЕДУР ОБРОБКИ ГОЛОСОВОГО СИГНАЛУ СИСТЕМИ АВТЕНТИФІКАЦІЇ НА ЯКІСТЬ ФОРМАНТНИХ ДАНИХ



М.С. ПАСТУШЕНКО

Харківський національний університет радіоелектроніки



О.М. ПАСТУШЕНКО

Збройні Сили України



Т.А. ФАЙЗУЛАЄВ, С.В. ШТАНГЕЙ

Харківський національний університет радіоелектроніки

Abstract – The current scientific task of improving the efficiency of voice authentication systems, which are widely used in modern access systems, is considered. It is known that it is possible to reduce errors of the first and second kinds in voice authentication systems by improving the procedures for the digital processing of the voice signal being analyzed, by better extraction of user features, or by improving procedures for decision-making on user admission. An important place in all voice signal processing procedures is occupied by formant data (spectral power levels, formant frequencies, spectral envelopes, and the width of the formant frequency spectrum). Based on the first two formants, speech recognition and synthesis are solved, and the next two formants enable user authentication. The purpose of this work is to outline ways to improve the quality of the formation of formant data in relation to the tasks of digital processing of speech signals. The object of the study is the process of obtaining formant data using amplitude-frequency and phase information, as well as the results of calculating the autocorrelation function of the analyzed signal. The subject of the study is methods and procedures for extracting formant data in the context of experimental research. The scientific novelty of the obtained research results lies in the fact that, for the first time, a comparative analysis of formant data obtained from different source information, namely amplitude-frequency and phase information, as well as the results of calculating the autocorrelation function of the analyzed signal, has been performed. The reliability of the research results is justified by the proper use of the known mathematical apparatus and the coincidence of formant data estimates as a result of processing the experimental user signal. The practical significance lies in the fact that the obtained results enable improvements in the quality and efficiency of voice data processing for speech recognition and synthesis, user authentication in voice systems, and several other applied tasks related to speech production.

Анотація – Розглядається актуальне наукове завдання щодо підвищення ефективності голосових систем автентифікації, які широко використовуються в сучасних системах доступу. Відомо, що знизити помилки першого і другого роду в голосових системах автентифікації можна шляхом удосконалення процедур цифрової обробки голосового сигналу, що аналізується, за рахунок більш якісного вилучення ознак користувача або удосконалення процедур прийняття рішення про допуск користувача. Важливе місце у всіх процедурах обробки голосових сигналів займають формантні дані (рівень спектральної потужності, формантні частоти, огинаюча спектру та ширина спектру формантних частот). На основі двох перших формант вирішуються завдання розпізнавання та синтезу мови, а дві наступні форманти дозволяють вирішити проблему автентифікації користувача. Мета даної роботи – намітити шляхи підвищення якості формування формантних даних стосовно задач цифрової обробки мовних сигналів. Об'єктом дослідження є процес отримання формантних даних при використанні амплітудно-частотної та фазової інформації, а також результатів розрахунку автокореляційної функції аналізованого сигналу. Предмет дослідження – методи та процедури вилучення формантних даних у процесі експериментальних досліджень. Наукова новизна отриманих результатів дослідження полягає в тому, що вперше виконано порівняльний аналіз формантних даних, отриманих за різною вихідною інформацією, а саме амплітудно-частотною та фазовою інформацією, а також результатів розрахунку автокореляційної функції аналізованого сигналу. Достовірність результатів досліджень обґрунтована коректним використанням відомого математичного апарату та збігом оцінок формантних даних в результаті обробки експериментального сигналу користувача. Практична значущість полягає в тому, що отримані результати дозволяють підвищити якість та ефективність обробки голосових даних при розпізнаванні та синтезі мови, автентифікації користувача в голосових системах та низці інших прикладних завдань, пов'язаних із мовотворенням.

Вступ

Сучасне життя окремої людини та більшості організацій немислиме без використання інфокомунікаційних систем різного масштабу та призначення. Однак, використання цих систем пов'язано з великими ризиками. Наприклад, 19 грудня 2024 року

Україна зазнала наймасштабнішої за останній час кібератаки з боку Росії. Внаслідок цілеспрямованої атаки було тимчасово зупинено роботу кількох ключових державних реєстрів. Це спричинило перебої у роботі електронних сервісів і унеможливило доступ до низки послуг [1].

У зв'язку з цим, велика увага приділяється методам і засобам мережного захисту більшості використовуваних інфокомунікаційних систем. Важливу роль для захисту відводиться засобам автентифікації користувачів цих систем. Останнім часом під час автентифікації все ширше використовуються біометричні ознаки користувача. При цьому особливе місце займають голосові системи автентифікації (ГСА), які за критерієм ефективності/вартості є найбільш перспективними та зручними для користувачів [2]. Так у травні 2022 року акціонерне товариство комерційний банк «Приватбанк» запустив можливість скористатися голосовою біометрією в застосунку Приват24 [3].

Разом з тим, сучасні ГСА не завжди задовольняють користувачів за своїми основними характеристиками і, в першу чергу, помилками першого і другого роду, а також можуть виникнути проблеми у користувачів з хворим голосовим трактом [4]. Зазначені проблеми ГСА можна вирішувати за допомогою удосконалення процедур цифрової обробки голосового сигналу як на етапі формування ознак шаблону користувача, так і на етапі його допуску (ухвалення рішення) до ресурсів інфокомунікаційної системи. На жаль, у сучасних ГСА цифрова обробка базується на амплітудно-частотній інформації реєстрованого сигналу [4]. Одночасно, в більшості випадків ігнорується фазова інформація [5], а також не використовуються властивості автокореляції, які застосовуються у приймачах сигналів і для підвищення відношення сигнал/шум в радіолокації та радіозв'язку [6].

Особливу роль серед ознак користувача ГСА займають формантні дані, які включають формантні частоти, рівень спектральної потужності, ширину спектру формант та їхню огинаючу спектру для аналізованого сигналу. Термін форманта позначає певну частотну область, в якій внаслідок резонансу посилюється деяке число гармонік частоти основного тону, виробленого голосовими зв'язками, тобто в спектрі звуку форманта є областю посилення частот, що досить виразно виділяється [4].

Таким чином, під формантами розуміються частотні резонанси (полюси передавальної функції) мовної акустичної системи (мовного тракту) користувача. При цьому найбільш важливий параметр – частота форманти, що визначається геометричною конфігурацією мовного тракту. У процесі промови тривалість мовного тракту змінюється, тому змінюються і формантні дані [4].

У цій роботі представлені результати досліджень формантних даних голосового сигналу на основі його амплітудно-частотної та фазової інформації, а також при використанні розрахунків автокореляційної функції. Об'єктом досліджень є процес оцінки формантних даних щодо різної вихідної інформації. Предмет досліджень: методи та процедури оцінки формантних даних стосовно ГСА. Мета досліджень – порівняльний аналіз оцінки формантних даних щодо різної вихідної інформації та розробити рекомендації щодо їх застосування.

Теоретичні та експериментальні дослідження базуються на використанні апарату математичної статистики, теорії та методів обчислювальної математики, теорії кіл та сигналів, методів математичного моделювання.

I. Аналіз робіт, пов'язаних з темою досліджень

Фундаментальні положення акустичної спектральної теорії мови були розроблені в XIX столітті. Значну роль при цьому зіграв видатний німецький вчений Г. Гельмгольц. Отримані Г. Гельмгольцем результати широко використовуються сьогодні. Слід звернути увагу і на працю відомого шведського акустика Г. Фанта, який серед іншого, запропонував теорію розрізнявальних ознак – універсальну акустичну класифікацію звуків. У цій класифікації значну роль відіграє формантна інформація. Г. Фант найбільш повно та всебічно вивчив формантну структуру звуків мови. Результати дослідження лягли основою акустичної теорії мовотворення, у якій стверджується, що кожному звуку мови відповідає своя форма акустичного мовного тракту з властивою йому структурою формант.

Усі сучасні роботи у сфері формантного аналізу голосових сигналів можна розділити на кілька класів [6]. Перший клас робіт включає методи формантного аналізу, що базуються на узагальненому математичному апараті та зводяться до аналізу локальних піків в амплітудному спектрі мовного сигналу. В цьому випадку в роботах відсутній фонетичний підхід і не враховані фізіологічні особливості мовного тракту людини. Основна увага приділяється процедурам цифрової обробки нестационарного полігармонічного (полічастотного) сигналу. Процедури здебільшого базуються на швидкому перетворенні Фур'є. Останнім часом для цих цілей використовується кепстральний аналіз, вейвлет-перетворення тощо. Тут же слід виділити роботу [7], в якій запропоновано модифікований метод аналізу амплітудного спектра з використанням переваг групової затримки (group delay). При цьому запропоновано використовувати фазову інформацію, яка дозволяє отримати більш високу роздільну здатність. Останнім часом у ряді робіт пропонується для виділення ознак користувача використовувати дробове перетворення Фур'є (fractional Fourier transform, FRFT) [8]. Вилучення ознак і аналіз мовних сигналів на основі FRFT в різних шумових середовищах, наприклад, коли мова зашумлена з відношенням сигнал/шум 0 дБ і 10 дБ відповідно, швидкість розпізнавання цього методу вище, ніж у традиційного параметричного методу. У [9] пропонуються два підходи на основі вейвлетів для отримання частоти основного тону і формант за допомогою банку фільтрів. Порівнюються результати представлених методів вилучення ознак на декількох мовних сигналах. Встановлено, що вейвлет-перетворення має хорошу точність порівняно з методом кепстру та нечутливе до шуму.

Другий великий клас робіт виходить з обліку особливостей мовного тракту людини, саме використання результатів досліджень Г. Фанта. Наприклад, в [10] враховується зміна частоти форманти через зміну площі поперечного перерізу мовного тракту. У [11] використовується зважене лінійне передбачення, що враховує внесок дже-

рела гортані в оптимізацію моделі та дозволяє підвищити точність визначення формантних частот. На основі характеристик мовного тракту та формантних частот в даний час вимірюються фізичні характеристики дорослих та дітей [12].

Останнім часом дуже актуальними стали проблеми обробки мови та отримання ознак користувача на основі нейронних мереж та штучного інтелекту, які складають окремий клас робіт. Питання моделювання та оцінки параметрів голосового тракту та гортанного джерела як голосних, так і назалізованих звуків з меншою кількістю помилок та за більш короткий час розглянуті у [13]. При цьому поряд з моделлю Лільєн-кранца-Фанта широко використовуються глибокі нейронні мережі, які мають нелінійну здатність підігнати параметри моделі, що розглядається. У [14] описується система розпізнавання голосних з використанням формантного аналізу та нейронної мережі. Процедури розпізнавання включають набір даних, попередню обробку набору даних, оцінку спектральної щільності потужності, витяг формант, навчання та тестування нейронної мережі. Показано, що голосні можуть бути класифіковані та розпізнані з великою точністю за допомогою перших трьох формантних частот та відповідною архітектурою нейронної мережі. Цікавою є робота [15]. У ній об'єднані три механізми оцінки формантних частот: оцінки формант на основі лінійного передбачення з використанням звичайного коваріаційного аналізу та нещодавно запропонований квазізакритий фазовий аналіз вперед-назад, а також нейронні мережі. Для обробки мови, як і в попередній роботі, використовуються перші три форманти.

Загальний висновок щодо виконаного аналізу наступний. Щонайменше використовується фазова інформація голосового сигналу, а також ігноруються можливості автокореляції.

II. Загальне формулювання проблеми

Сучасні ГСА базуються в основному на цифровій обробці амплітудно-частотних матеріалів, які отримані після реєстрації за допомогою мікрофона. Як впливає з проведеного аналізу літератури, у більшості випадків ігнорується фазова інформація голосового сигналу і не застосовуються процедури автокореляції, які ефективно та широко використовуються у радіолокації та радіозв'язку.

Раніше, при обмеженнях на обчислювальний ресурс і оперативну пам'ять, що використовується, такий підхід був виправданий. У нових умовах, коли зазначені обмеження не існують і можуть застосовуватися сучасні процедури цифрової обробки сигналів, ігнорування фазової інформації та автокореляції стає неприйнятним.

Тому в цій роботі вирішується наукове завдання щодо порівняльного аналізу впливу амплітудно-частотної, фазочастотної інформації та автокореляції на прикладі оцінки формантних даних експериментального голосового сигналу.

Відомо, що у сучасних ГСА фазова інформація не реєструється і не використовується. Тому перше завдання пов'язане з формуванням фазової інформації за матеріалами реєстрації $X(t)$ голосового сигналу $S(t)$. В даний час фазова інформація може

бути отримана алгоритмічно під час реалізації перетворення Гільберта. Це перетворення дає можливість сформувати уявну частину $S_m(t)$ аналітичного сигналу [16]

$$S_m(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{S(\tau) d\tau}{t - \xi}, \quad (1)$$

де ξ – змінна інтегрування.

Знаючи складові аналітичного сигналу, можемо отримати фазову інформацію аналізованого сигналу за допомогою відомого співвідношення

$$\varphi(t) = \arctg \frac{S_m(t)}{S(t)}. \quad (2)$$

Після відповідної обробки $\varphi(t)$ приймає форму пилоподібного сигналу, який змінюється в межах від 0 до 2π і має невідому тривалість. Далі отримані фазові дані можуть опрацьовуватися за допомогою алгоритмів, які застосовуються для цифрової обробки амплітудно-частотної інформації. Зазначені процедури будуть використані для обробки експериментального голосового сигналу при формуванні фазових даних.

Тепер необхідно уточнити процедури оцінки автокореляції, які широко використовуються в радіолокації та радіозв'язку. Як відомо, сигнал, що реєструється, містить суміш голосового сигналу і шуму, який заважає. При адитивній моделі матеріали реєстрації мають такий вигляд

$$X(t) = S(t) + n(t), \quad (3)$$

де $n(t)$ – гаусівський шум.

Для підвищення відношення сигнал/шум в радіолокації рекомендують розраховувати автокореляційну функцію суміші, що обробляється. Це виходить з визначення автокореляційної функції [16]

$$C_{XX}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) \cdot X(t - \tau) dt, \quad (4)$$

де T – період реєстрації сигналу; τ – час затримки.

В силу дистрибутивності кореляційного оператора $C_{XX}(\tau)$ можемо подати у вигляді наступних доданків

$$\overline{C}_{XX}(\tau) = \overline{C}_{ss}(\tau) + \overline{C}_{nn}(\tau) + \overline{C}_{sn}(\tau) + \overline{C}_{ns}(\tau). \quad (5)$$

Проаналізуємо складові автокореляційної функції. Аналіз почнемо з другого доданку $\overline{C}_{nn}(\tau)$, яке з урахуванням стохастичності шуму прагне до нуля, а точці $\tau = 0$ прагне до дисперсії шуму. Третій і четвертий доданки ($\overline{C}_{sn}(\tau)$, $\overline{C}_{ns}(\tau)$) прагнуть до нуля,

оскільки оброблювані сигнали центровані. Складова $\bar{C}_{ss}(\tau)$ матиме більш високе відношення сигнал/шум [17]. В результаті отримаємо

$$X'(\tau) = \bar{C}_{ss}(\tau) + \varepsilon(\tau), \quad (6)$$

де $\varepsilon(\tau)$ – деяка помилка, яка зменшується зі збільшенням часу інтегрування T .

Тут звернемо увагу, що автокореляційна функція суми періодичних сигналів є періодична функція з такою самою частотою [16, 17], тобто вона може містити усі частотні складові сигналів формант. Результати розрахунку автокореляційної функції надалі обробляються за алгоритмами, що застосовуються для обробки амплітудно-частотної інформації. Особливістю голосового сигналу стосовно сигналів радіолокації полягають у тому, що голосовий сигнал є полігармонічним, а саме може включати до 10 формантних частот.

III. Результати виділення формантних даних із експериментального голосового сигналу

Як експериментальний голосовий сигнал будемо використовувати український звук «о» (жіночий голос). Частота дискретизації становила 64 кГц. Зауважимо, що тривалість аналізованого звуку становить приблизно 200 мс. У цьому полягає відмінність від традиційних процедур, які використовують фрагменти звуку 20-30 мс. Експериментальний голосовий сигнал подано на рис. 1.

Для оцінки залежності формантних частот від аналізованих даних будемо розраховувати спектр за допомогою швидкого перетворення Фур'є і подальшу порогову обробку для виявлення максимумів.

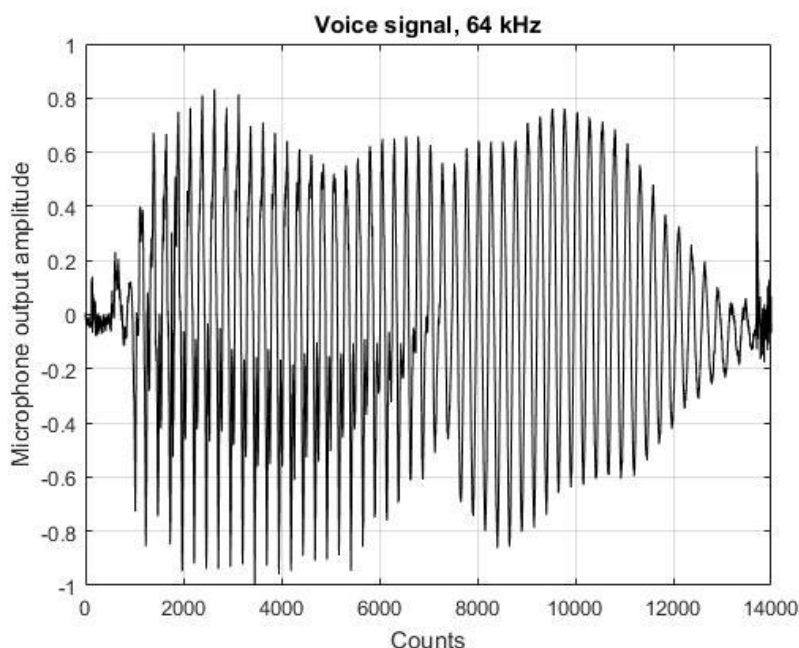


Рис. 1. Досліджуваний сигнал

Спочатку розрахуємо амплітудно-частотний спектр (АЧС) від матеріалів реєстрації та оцифрування українського звуку «о» за допомогою швидкого перетворення Фур'є. Отриманий спектр, обмежений частотою до 3,5 кГц, представлений на рис. 2.

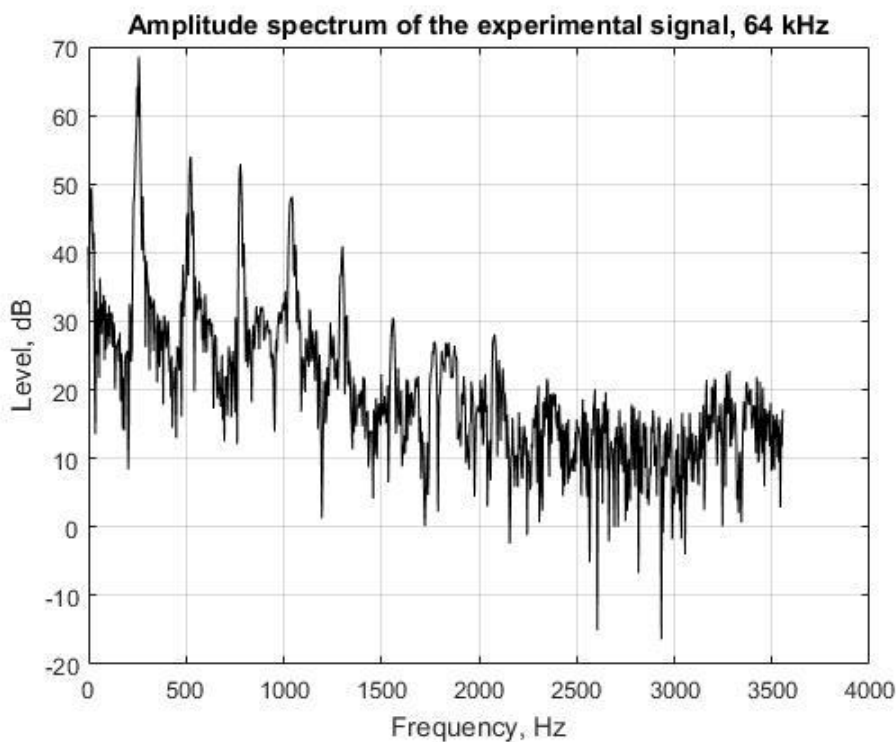


Рис. 2. АЧС спектр аналізованого звуку

Обмеження частоти обумовлено тим, що основну увагу приділимо першим чотирьом формантам. Відомо, що перші дві мають велике значення для розпізнавання мови, а третя та четверта визначають якість автентифікації (а саме, хто говорить).

Поверхневий аналіз АЧС спектру дає право зробити висновок, що перші шість формантів визначаються досить просто. При цьому є проблеми із визначенням сьомої форманти. Числові характеристики формантних даних спектра будуть представлені нижче. На рис. 3 представлений фазочастотний (ФНС) спектр аналізованого сигналу. Як і раніше, виникають складнощі з визначенням сьомої форми, оскільки в спектрі присутні два максимуми. Виділити за допомогою граничної обробки частоту сьомої форманти проблематично. Потрібно приймати додаткові процедури.

На рис. 4 представлений спектр, збудований за результатами розрахунку автокореляційної функції (АКФ) від досліджуваних даних. Зауважимо, що АКФ можна розраховувати кілька разів (ітерацій). У даному випадку представлений спектр після першої ітерації. Аналіз спектра АКФ показує, що виникають проблеми із визначенням частоти шостої форманти. Це зумовлено «дробленням» максимуму цієї форманти. Разом з тим максимум сьомої форманти є яскраво вираженим і дає можливість визначити її частоту.

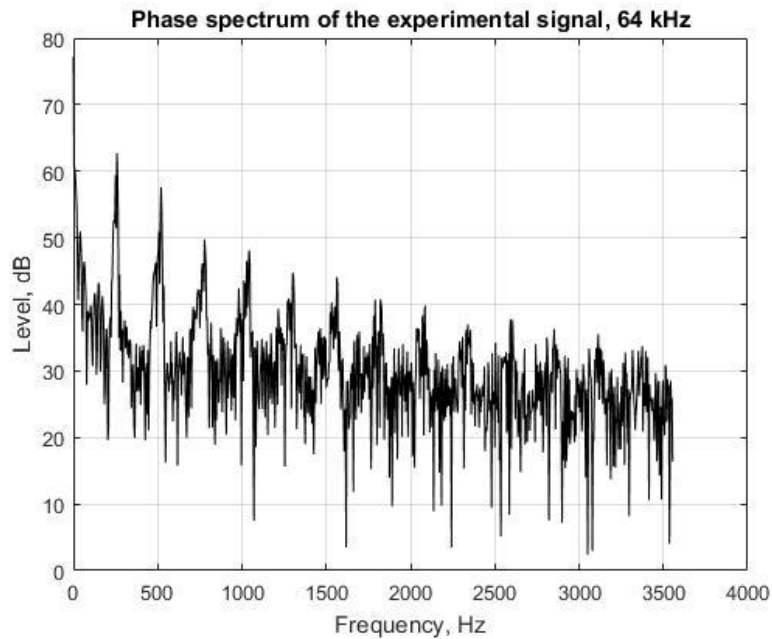


Рис. 3. ФНС спектр звуку, який аналізується

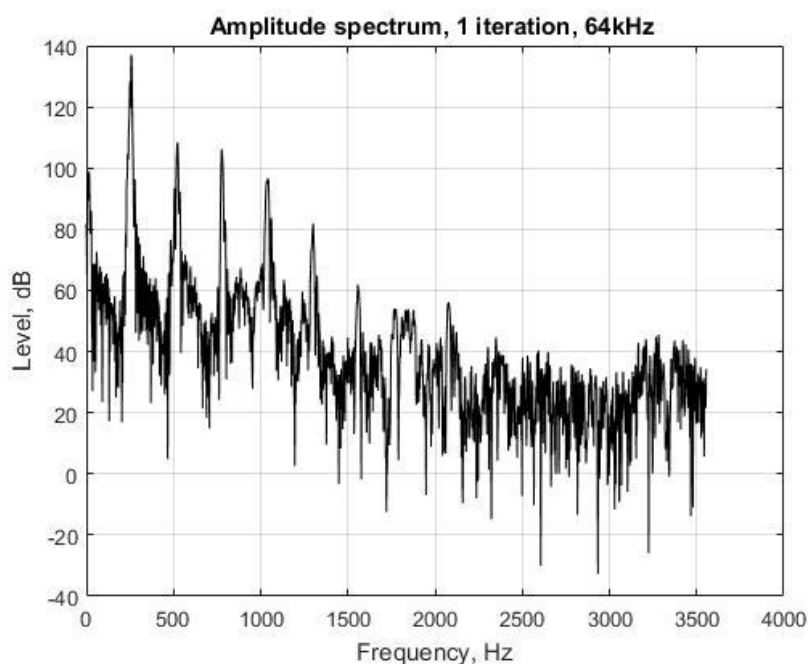


Рис. 4. Спектр, розрахований з урахуванням АКФ досліджуваного сигналу (1 ітерація)

Проаналізуємо вплив ітераційного розрахунку АКФ на досліджуваний спектр. З цією метою рис. 5 представлений спектр АКФ після третьої ітерації. Аналіз цього спектра показує суттєве збільшення спектральної густини досліджуваних формант. Яскраво вираженим став максимум шостої форманти. Більш складно виділяти наступні формантні частоти.

Тепер здійснимо числовий аналіз досліджуваних формантних частот. Зауважимо, що точність визначення формантних частот становить 4 Гц. Це обумовлено роздільною здатністю спектрального аналізу. Зведені дані представлені у табл. 1. У табл. 1 прийнято такі позначення АКФ₁ – спектр розрахований після однієї ітерації оцінки автокореляційної функції, АКФ₃ – три ітерації розрахунку АКФ відповідно.

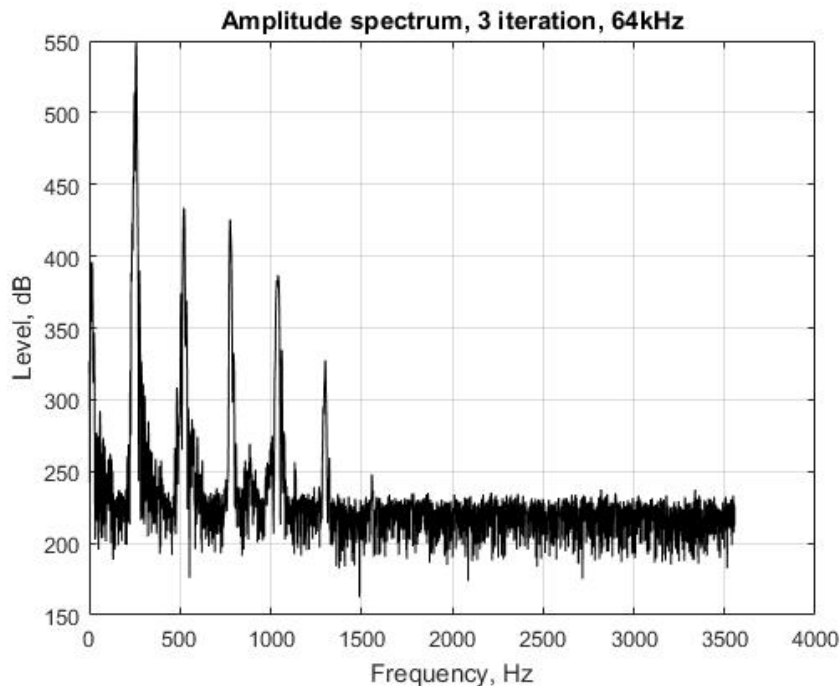


Рис. 5. Спектр, розрахований з урахуванням АКФ досліджуваного сигналу (3 ітерація)

Таблиця 1. Числові значення формантних даних

В и д	Параметр, який аналізувався	Номер форманти			
		1	2	3	4
А Ч С	Рівень, дБ	69	54	53	48
	Частота, Гц	258	524	778	1043
	Ширина, Гц	61	78	66	129
Ф Ч С	Рівень, дБ	62	58	50	48
	Частота, Гц	258	520	774	1043
	Ширина, Гц	59	90	105	106
А К Ф1	Рівень, дБ	137	108	106	97
	Частота, Гц	258	522	776	1042
	Ширина, Гц	45	32	41	31
А К Ф3	Рівень, дБ	514	434	425	383
	Частота, Гц	250	522	777	1034
	Ширина, Гц	63	54	36	38

Для аналізу рівня спектральної щільності доцільно побудувати графіки, які дозволять оцінити форму огинаючої піків формантних частот. У зв'язку із значним розкидом абсолютних значень спектральних рівнів, виконаємо їхнє нормування на максимальне значення для кожного виду даних. Результати подано на рис. 6.

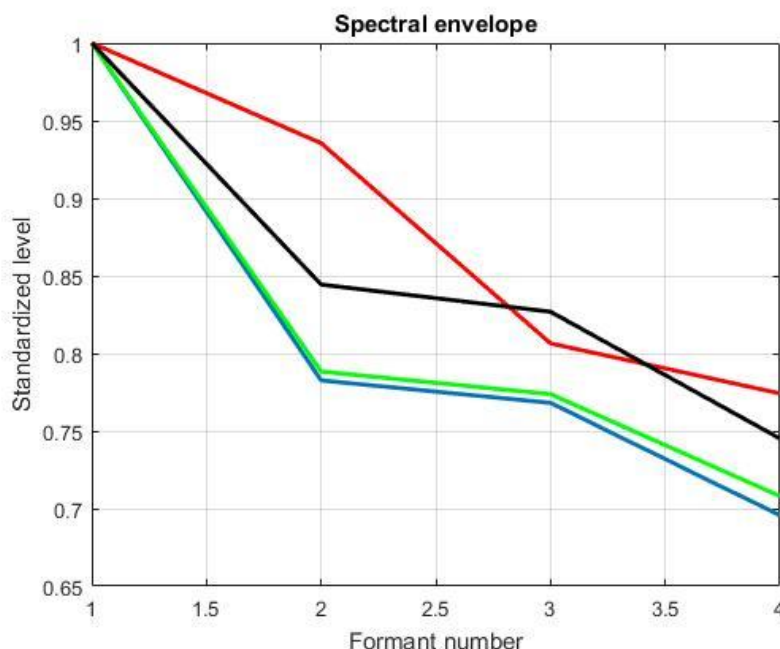


Рис. 6. Нормована огинаюча спектральних рівнів формантних частот

При цьому синім кольором на цьому рис. 6 показано огинаючу для АЧС, червоним – ФНС, зеленим – АКФ (1 ітерація), чорним – АКФ (3 ітерації) відповідно. Аналіз рис. 6 показує, що огинаючі АЧС і АКФ (1 ітерація) практично збігаються. Огинаючі ФНС і АКФ (3 ітерації) відрізняються і практично перебувають у «протифазі».

Тепер проведемо аналіз формантних частот. Вони трохи відрізняються, максимум до 9 Гц. Тому для практичного використання можна рекомендувати використовувати усереднену оцінку відповідної формантної частоти.

Ширина спектру формантних частот: гірші значення для АЧС, а найкращі для АКФ. Останнє зумовлено тим, що АКФ забезпечує значне підвищення відношення сигнал/шум.

Висновки

Розглядається актуальне наукове завдання щодо підвищення ефективності голосових систем автентифікації за рахунок удосконалення процедур оцінки формантних даних аналізованого сигналу користувача. Формантні дані широко та ефективно використовуються при всіх видах обробки голосового сигналу, а саме, розпізнаванні мови, автентифікації користувачів, виявленні хвороб голосового тракту, синтезі голосових даних. Особливе місце приділяється першим чотирьом формантам: дві перші важливі для розпізнавання мови, а дві наступні – для автентифікації користувача.

У цій роботі досліджувався вплив оброблюваних даних на якість формування формантних даних. Вперше виконано порівняльний аналіз формантних даних, отриманих на основі амплітудно-частотної та фазової інформації, а також після розрахунку автокореляційної функції досліджуваного сигналу користувача.

Достовірність отриманих результатів підтверджується коректністю використовуваного математичного апарату і близькими оцінками, отриманими в процесі модельного експерименту з обробки голосового сигналу користувача системи автентифікації. Отримані результати можуть знайти практичне застосування в системах розпізнавання мови, сучасних системах автентифікації та інших додатках, пов'язаних з обробкою і синтезом мови. Подальші дослідження доцільно орієнтувати на розробку процедур узагальнення отриманих оцінок з метою підвищення їхньої точності.

Список літератури

1. Стефанішина повідомила про масштабну кібератаку на держреєстри Мін'юсту, їхню роботу зупинено. URL: <https://www.radiosvoboda.org/a/news-stefanishyna-khakerska-ataka-reyestry/33246251.html>.
2. Пастушенко, М. О., Пастушенко, М. С., Петраченко, М. О. (2023), "До питання оцінки ефективності біометричних систем", Проблеми телекомунікацій», No. 1(32), С. 37–44. DOI: <https://doi.org/10.30837/pt.2023.1.03>.
3. Захищаємо та спрощуємо ваші дзвінки до клієнтської підтримки. URL: <https://privatbank.ua/voice-biometrics>.
4. Beigi, H. (2011), Fundamentals of Speaker Recognition, NY: Springer, 942p. DOI: <https://link.springer.com/book/10.1007/978-0-387-77592-0>.
5. Pastushenko, M., Krasnozheniuk, Ya., Zaika, M. (2020), "Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User", Proceedings of the 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, P. 467–472. DOI: <https://doi.org/10.1109/PICST51311.2020.9468083>.
6. Pastushenko, M. S., Pastushenko, M. A., Faizulaiev, T. A. (2024), "Estimation of formant information using autocorrelation function of voice signal", Radio Electronics, Computer Science, Control, No. 3(70), P. 144–150. DOI: <https://doi.org/10.15588/1607-3274-2024-3-12>.
7. Chowdhury, H.A., Rahman, M.S. (2021), "Formant estimation from speech signal using the magnitude spectrum modified with group delay spectrum", Acoustical Science and Technology, No. 42(2), P. 93–102. DOI: <https://doi.org/10.1250/ast.42.93>.
8. Zhai, Y. (2023), "Feature Extraction and Analysis of Speech Signal Based on Fractional Fourier Transform", Academic Journal of Computing & Information Science, No. 6(5), P. 107–112. DOI: <https://doi.org/10.25236/AJCIS.2023.060515>.
9. Goki, S. H., Ghazvini, M., Hamzenejadi, S. (2022), "A Wavelet Transform Based Scheme to Extract Speech Pitch and Formant Frequencies", arXiv preprint arXiv:2209.00733, P. 1–16. DOI: <https://doi.org/10.48550/arXiv.2209.00733>.
10. Kaburagi, T., Takano, T., Sakamoto, Y. (2013), "Estimating area function of the vocal tract from formants using a sensitivity function and least-squares", Acoustical Science and Technology, No. 34(5), P. 301–310, DOI: <https://doi.org/10.1250/ast.34.301>.

11. Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A. M., Story, D. (2012), "Improved formant frequency estimation from high-pitched vowels by downgrading the contribution of the glottal source with weighted linear prediction", Proceedings of the INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association: Portland, Oregon, USA, September 9-13, 2012. International Speech Communication Association, P. 1612–1615. DOI: <https://doi.org/10.21437/Interspeech.2012-443>.
12. Lehman, J. F., Singh, R. (2016), "Estimation of children's physical characteristics from their voices", Proceedings of the INTERSPEECH 2016, September 8-12, San Francisco, USA, P. 1417–1421. DOI: <https://doi.org/10.21437/Interspeech.2016-146>.
13. Li, K., Akagi, M., Li, Y., Unoki, M. (2024), "Modeling and Estimation of Vocal Tract and Glottal Source Parameters Using ARMAX-LF Model", arXiv preprint arXiv:2410.04704, P. 1–15. DOI: <https://doi.org/10.48550/arXiv.2410.04704>.
14. Kisić, E., Dikić, G., Petrović, V., (2020), "Vowel recognition using formant analysis and neural networks", in Proceedings of the 7th International Conference on Electrical, Electronic and Computing Engineering IcETRAN 2020 and 64th National Conference on Electronics, Telecommunication, Computing, Automatic Control and Nuclear Engineering ETRAN 2020, (ETRAN Society, Belgrade, 2020), P. 169–173. URL: https://www.etrans.rs/2020/ZBORNIK_RADOVA/Radovi_prikazani_na_konferenciji/031_AUI1.4.pdf
15. Alku, P., Kadiri, S. R., Gowda, D. (2023), "Refining a Deep Learning-based Formant Tracker using Linear Prediction Methods", Computer Speech & Language, No. 81, p.101515. DOI: <https://doi.org/10.1016/j.csl.2023.101515>.
16. Van Trees, H., Bell K. L. (2007), "Excerpts from Part I of Detection, Estimation, and Modulation Theory", Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking, IEEE, P. 89–109. DOI: <https://doi.org/10.1109/9780470544198.ch1>.
17. Ifeachor, E.C., Jervis, B. W. (1993), Digital Signal Processing: A Practical Approach Second Edition, Addison-Wesley Publishing Company, 779p. URL: <https://seemabaji1.wordpress.com/wp-content/uploads/2019/01/jervis-book.pdf>.