



DOI: 10.5281/zenodo.1493012

LCC - № [RZ201-999](#)

МОДЕЛЮВАННЯ МОВНОГО АНАЛІЗАТОРА

Римма Риженко¹, Всеволод Новіков¹

¹ Херсонський національний технічний університет

Address for Correspondence: Римма Риженко, студентка

E-mail: rigenko@excite.it

Анотація. Робота присвячена створенню моделей блоків різного функціонального призначення аналізатора мови й голосу мовця. Метою роботи є моделювання мовного аналізатора. Проведено аналіз фізіологічної появи мови/голосу, а також автоматизованого розпізнавання мови/голосу мовця. Для проведення експериментальних досліджень як матеріал були обрані голосні, тому що вони є своєрідними точками опори при розпізнаванні мови.

На основі проведеного аналізу побудована функціональна модель цифрового аналізатора мови, а також фільтра й блоку сприйняття емоцій, у тому числі різних видів неправди, для даного аналізатора.

Проведено моделювання розпізнавання змін у психоемоційному поведженні людини по біосигналу при зміні голосу/мови мовця.

В ході роботи побудовано модель системи автоматичного розпізнавання голосних у потоці зливої мови, що не залежить від наявності шумів у мовному сигналі й має властивість незалежності від індивідуальних характеристик голосу мовця. Модель системи автоматичного розпізнавання мови будується на основі акустичного аналізу мовного сигналу.

Представлено результат періодично-аперіодичної моделі декомпозиції з авторегресійною параметризацією й коефіцієнтом масштабування. Проведено аналіз вибору цифрового аналізатора, а також фільтра мовного аналізатора. Розвинуто часовий метод аналізу голосових сигналів, який враховує не тільки гармонічний склад та спектральний розподіл потужності сигналу, але й часову зміну його властивостей. Обґрунтовано використання синфазного методу аналізу періодично корельованих випадкових процесів для визначення змін періоду корельованості, пов'язаних зі змінами РС. Для підтвердження слушності припущення про період корельованості як

інформативну характеристику стохастичної моделі сигналу та доцільності застосування розроблених алгоритмів проведено експериментальну верифікацію моделі.

Результати верифікації підтвердили адекватність вибраної моделі – ПКВП – реальним сигналам: статистичні оцінки (математичне сподівання, дисперсія) флюктують біля періодичних функцій.

Ключові слова: вокодер, голос, автоматизована система.

Abstract. The work is devoted to the creation of models of blocks of various functional purposes of the speaker speech and voice analyzer. The purpose of the work is to simulate a language analyzer. The analysis of the physiological appearance of speech / voice, as well as the automated voice / speech recognition of the speaker has been carried out. For conducting experimental studies as a material, vowels were chosen because they are peculiar points of support in recognition of the language. On the basis of the conducted analysis, a functional model of the digital language analyzer, as well as a filter and block of emotion perception, including various types of lies, for this analyzer is constructed. The simulation of the recognition of changes in human psycho-emotional behavior at the biosignal when changing the voice/speech of the speaker is carried out. In the course of work, the model of the system of automatic recognition of vowels in the stream of the fused language is constructed that does not depend on the presence of noise in the speech signal and has the property of independence from the individual characteristics of the voice of the speaker. The automatic speech recognition system model is based on the acoustic analysis of the speech signal.

The result of the periodic-aperiodic decomposition model with autoregressive parametrization and scaling factor is presented. The analysis of the choice of the digital analyzer, as well as the language analyzer filter is carried out.

Keywords: vocoder, voice, automated system.

Introduction. Основними завданнями біометрії є верифікація й ідентифікація людини по унікальним, властивим тільки йому біологічним ознакам. Останнім часом це завдання стало особливо актуальним при проведенні антитерористичних заходів.

До основних біометричних технологій відносять розпізнавання особистості по відбитках пальців, геометрії обличчя або кисті руки, райдужній оболонці або сітківці ока, геометрії малюнка вен або по підпису. Голос також є унікальною біометричною характеристикою людини.

Перевагами голосової біометрії є безконтактність методу, використання стандартного встаткування, наприклад стільникового телефону, і можливість одержання додаткової інформації про людину, наприклад про його психоемоційний стан і антропометричні параметри.

Велике значення має фізичне й математичне моделювання; дані важливі для мовознавства, логопедії, техніки зв'язку, конструювання систем автоматичного розпізнавання й синтезу мови.

Зміна голосу мовця відображається на його мові, емоціях, жестах, але також можуть бути видні зміни на його біосигналах.

Математична модель сигналу вирішальною мірою визначає ефективність функціонування інформаційно-вимірювальних систем, які забезпечують отримання відомостей про стан досліджуваного об'єкта. Тому центральним завданням є обґрунтування вибору такої математичної моделі, яка адекватна як об'єктові дослідження, так і класові розв'язуваних задач.

Моделювання мовного аналізатора на основі сприйняття емоцій і розшифровки змін серцевого ритму на основі змін голосу вважаємо актуальною темою.

Objective. Метою роботи є моделювання мовного аналізатора. Для досягнення мети роботи, необхідно вирішити наступні завдання:

1. Провести експериментальні дослідження з розпізнавання голосу й мови мовця. На основі отриманих даних побудувати модель цифрового аналізатора мови.
2. Проаналізувати й розробити модель блоку розпізнавання емоцій мовця.
3. Побудувати модель голосового сигналу по сигналу ритміки серця.

Materials and methods. Для проведення експериментальних досліджень як матеріал були обрані голосні, тому що вони є своєрідними точками опори при розпізнаванні мови.

Описуване дослідження складалося із трьох етапів.

На початку було проведено два аудиторських експерименти, метою яких був відбір матеріалу для подальшої обробки змодельованою системою аналізу мовного сигналу.

Для того щоб подивитися наскільки точно зберігається інформативність мовного сигналу при обробці, результати обробки були запропоновані для прослуховування групі аудиторів.

Було проведено два аудиторських експерименти, метою яких було відібрати слова, які надалі будуть використовуватися як матеріал для акустичного аналізу. Відбиралися лише правильно розпізнані слова.

Матеріалом для експерименту послужили логатоми (псевдослова, що не мають змісту).

Результати експерименту представлені на Рис.1.

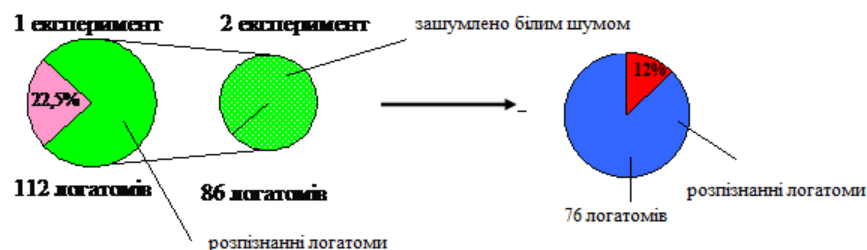


Рис. 1. Розпізнавання аудиторями логатомів

Під час першого експерименту, як матеріал було подано для прослуховування 112 логатомів. З них не були розпізнані хоча б одним з аудиторів 26 логатомів (22,5 %). Всі інші логатоми стали матеріалом для наступного аудиторського експерименту.

Для прослуховування під час другого експерименту подавалися 86 логатомів. Слова були зашумленні білим шумом. (зі співвідношенням 6 дБ (тобто інтенсивність корисного сигналу перевищувала інтенсивність шуму приблизно в 4 рази)).

(Таке співвідношення було обрано тому, що дослідження сприйняття людьми зашумленої мови показали, що шум такого рівня є границею, після якої розуміння мови людиною починає порушуватися. Білий шум був обраний з тієї причини, що він сильніше інших типів шумів порушує розуміння мови й з ним складніше всього боротися.[1])

У ході експерименту аудиторам не вдалося розпізнати лише 11 логатомів (12 %).

Всі інші логатоми (76 слів) були розпізнані абсолютно вірно й піддалися акустичній обробці.

Для акустичної обробки використовувався полосний вокодер з бінарною системою кодування.

Результати аудиторського експерименту показали, що правильно було розпізнано 76 логатомів (близько 91 %).

Експериментальні дані можуть далі служити вхідними сигналами для навчання нейронної мережі, на якій буде ґрунтуватися блок розпізнавання голосних. Загальна схема вибудованої моделі системи розпізнавання голосних у злитій мові показана на Рис.2.

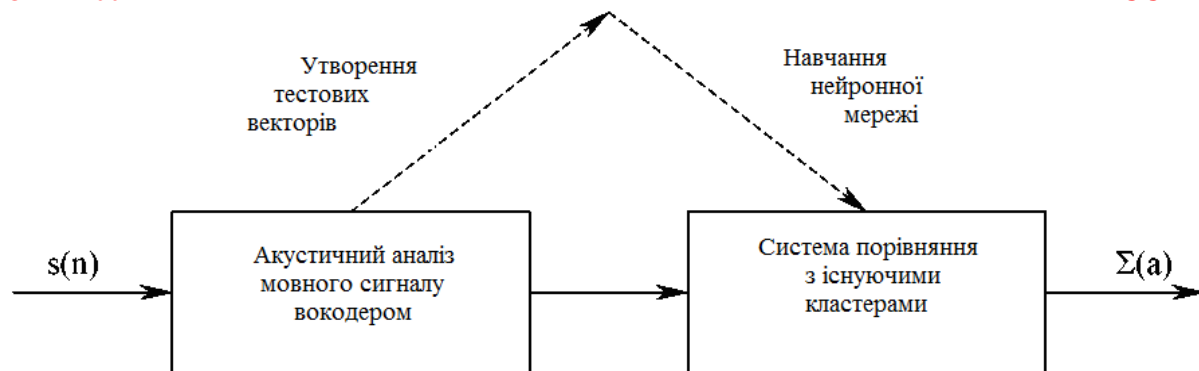


Рис. 2. Загальна схема вибудованої моделі системи розпізнавання голосних у злитій мові.

По отриманим експериментальним даним і на основі літературного аналізу[2-4] проводиться моделювання мовного аналізатора. Моделювання включає наступні етапи:

1. Моделювання голосу й мови мовця;
2. Аналіз блоку сприйняття й розпізнавання голосу/мови;
3. Моделювання блоку розпізнавання емоцій;
4. Моделювання фільтра аналізатора;
5. Моделювання розпізнавання голосу/мови по аналізу ритму серця, як одного з біосигналів.

Результати моделювання голосу й мови мовця відображенні на Рис. 3.

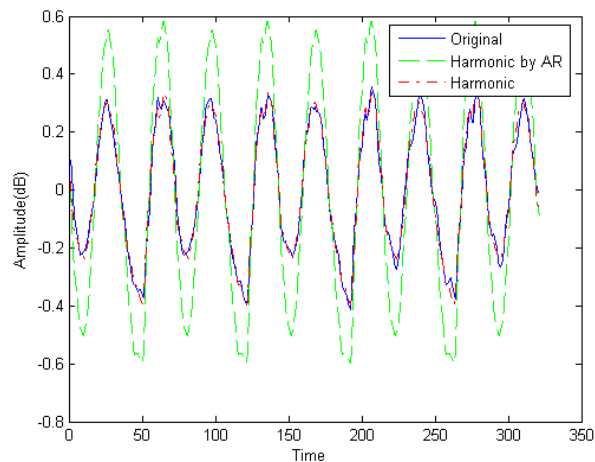


Рис. 3. Форма мовного сигналу, пересинтезованої гармоніки й форма хвилі з використанням авторегресійної моделі (безперервна лінія) і оригінальної амплітуди

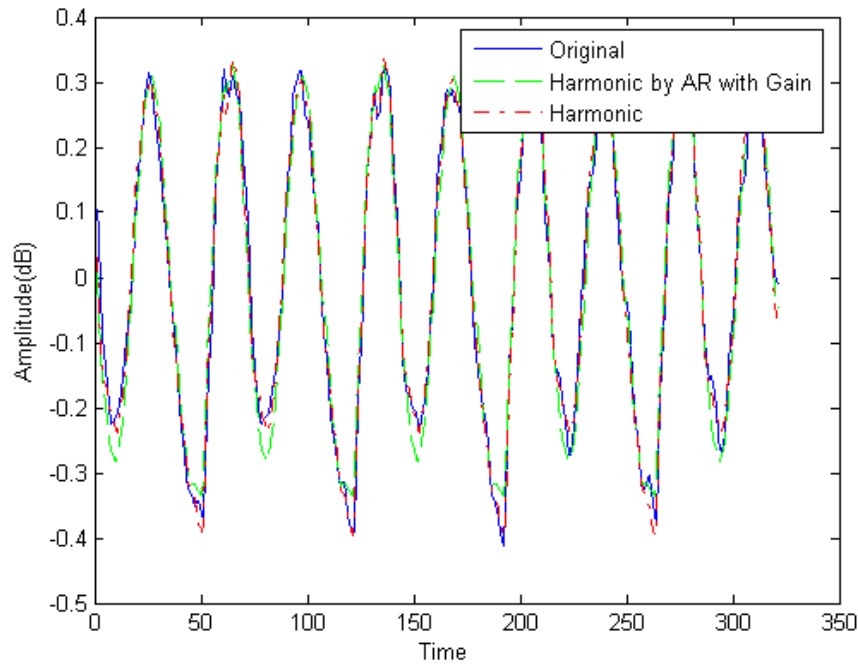


Рис. 4. Форма мовного сигналу, пересинтезованої гармоніки й форма хвилі з використанням авторегресивної моделі (безперервна лінія) з коефіцієнтом збільшення й оригінальної амплітуди

Як показано на рис. 3, форма пересинтезованої гармоніки мовного сигналу, з використанням оригінальної амплітуди збігається з вихідною формою сигналу, а форма сигналу, пересинтезованої гармоніки не збігаються (їхні амплітуди відмінні). Результат на рис. 4 показує, що форма пересинтезованої гармоніки мовного сигналу на основі авторегресивної моделі може збігатися з оригінальною формою мовного сигналу і її періодичній складовій.

Для визначення помилки або розходження між періодичними компонентами використовувалося середньквдратичне значення відхилення і логарифм спектрального перекручування. Здійснювалася оцінка трьох підходів:

-періодична частина, що синтезується рівнянням (1);

$$h(n) = \sum_{k=0}^{K(n)-1} A_n(k) \cos(k\omega_0 n + \varphi_k) = \sum_{k=0}^{K(n)-1} A_n(k) \cos(\theta(n,k) + \varphi_k), \theta(n,k) = nk\omega_0, \quad (1)$$

де $A_n(k)$ і $\varphi_k(k)$ - амплітуда й фаза для k -ї гармоніки;

ω_0 - фундаментальна частота,

$K(n)$ - номер гармоніки.

-періодична частина авторегресивної моделі:

$$P(z) = \frac{G}{1 + \sum_{k=1}^{N_A} a_k z^{-k}}, \quad (2)$$

де N_A - порядок моделі, коефіцієнт G і коефіцієнти a_k є параметрами авторегресивної моделі або параметрами лінійного проорокування.

-Періодична частина авторегресивної моделі з коефіцієнтом збільшення.

$$h_{AR}(n) = g \sum_{k=0}^{K(n)-1} A_t(k) \cos(k\omega_0 n + \varphi_k), \quad (3)$$

коефіцієнт підсилення визначається вираженням

$$g = \frac{\sqrt{\sum (H(k))^2}}{\sqrt{\sum (P(k))^2}} \quad (4)$$

Амплітуда визначається $A(k) = |P(k)|$

$$H(k) = \sum_{n=0}^{N-1} s(n) e^{-j \left(n \left(\omega_0 + \frac{\Delta\omega_0 n}{2N} \right) \right)}. \quad (5)$$

$s(n)$ періодичної (гармонійна) частини

N - довжина фрейму.

Як показано на рис. 5 і 6, авторегресивна модель із коефіцієнтом масштабування дає кращі результати, чим авторегресивна модель. Тестування показує, що якість синтезованої мови дуже висока, і слухачі не можуть відрізнити оригінальний сигнал від синтезованого моделлю на основі періодично-аперіодичної декомпозиції.

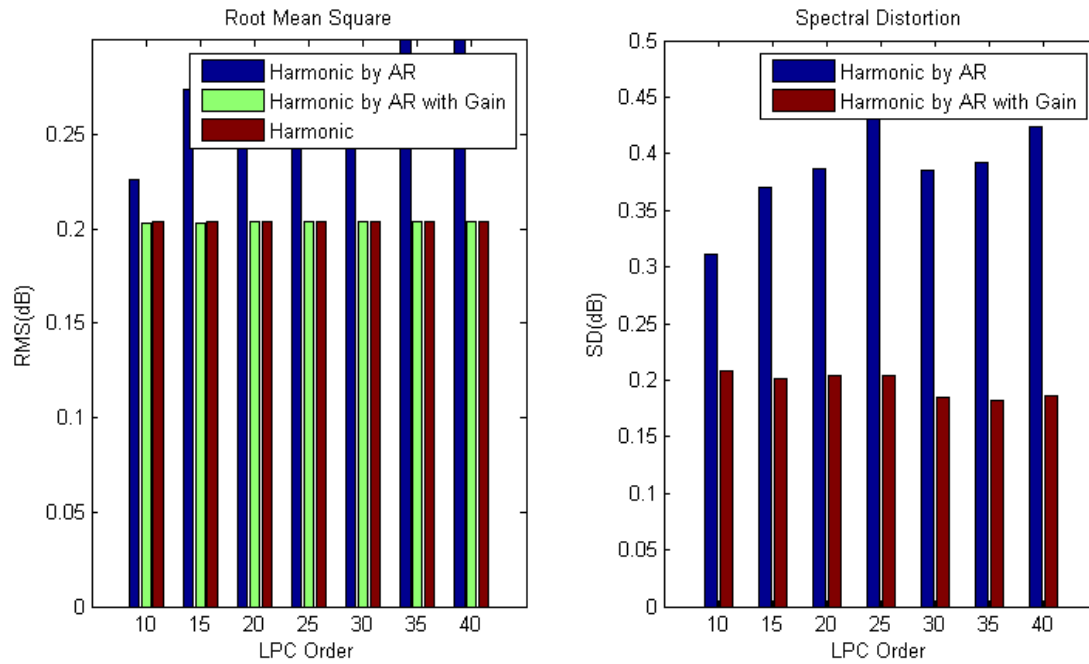


Рис. 5. RMS і логарифм спектрального перекручування для жіночого голосу

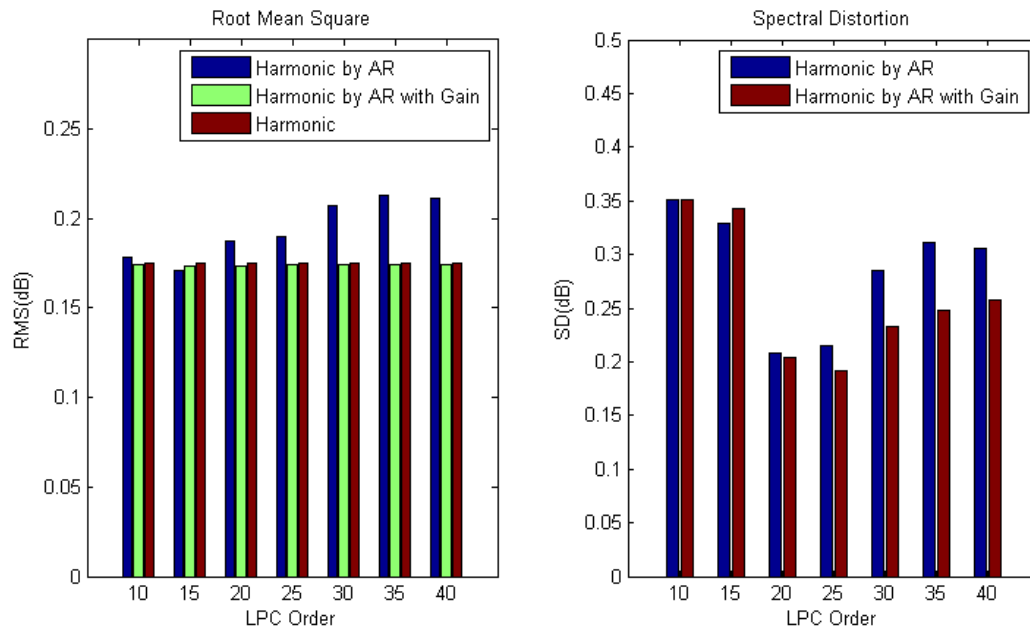


Рис. 6. RMS і логарифм спектрального перекручування для чоловічого голосу

Таким чином, результат дослідження показав, що якість синтезованої мови дуже висока, і запропонована модель збігається з періодично-аперіодичною декомпозицією, у той же час модель має менший порядок і більше стабільна.

Далі згідно методики експерименту проводимо моделювання цифрової моделі мовного тракту. На підставі літературних даних[5-6] і проведених експериментальних досліджень, при моделюванні мовного аналізатора, за основний блок приймемо вокодер, як цифрову модель мовного

тракту.

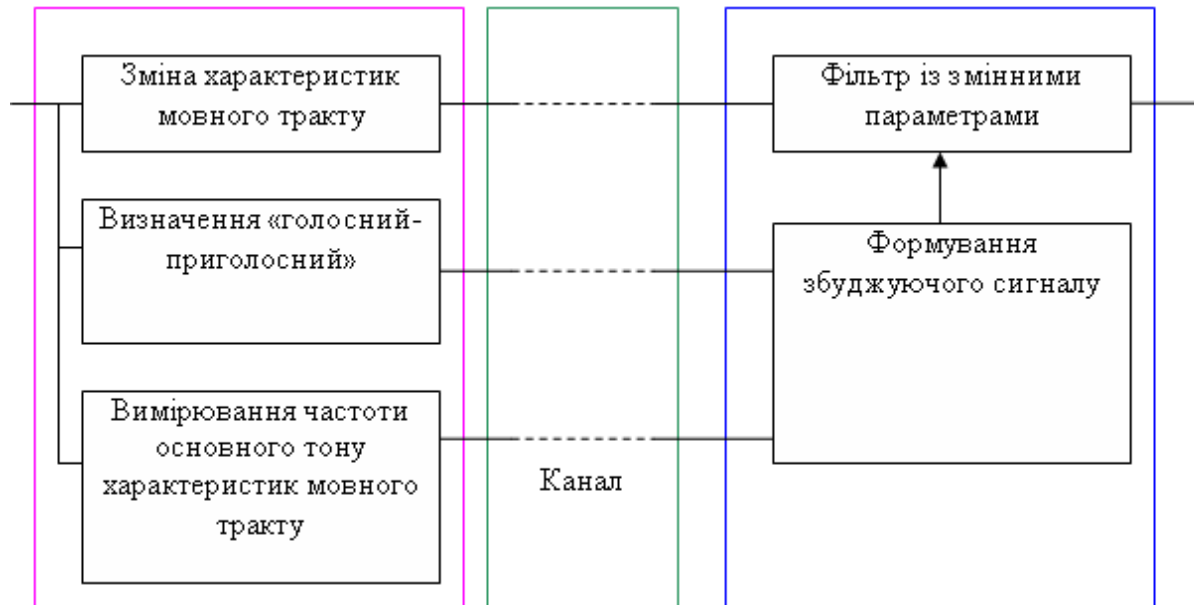


Рис. 7. Узагальнена структурна схема полосного вокодера



Рис. 8. Структура програмних модулів аналізатора й синтезатора

Мовний аналізатор, що нами моделюється, відрізняється тим, що він буде здатний розпізнавати емоційний стан людини та 5 типів неправди. Рисунок 9. відбиває передбачуване робоче вікно програми блоку розпізнавання емоцій мовного аналізатора.

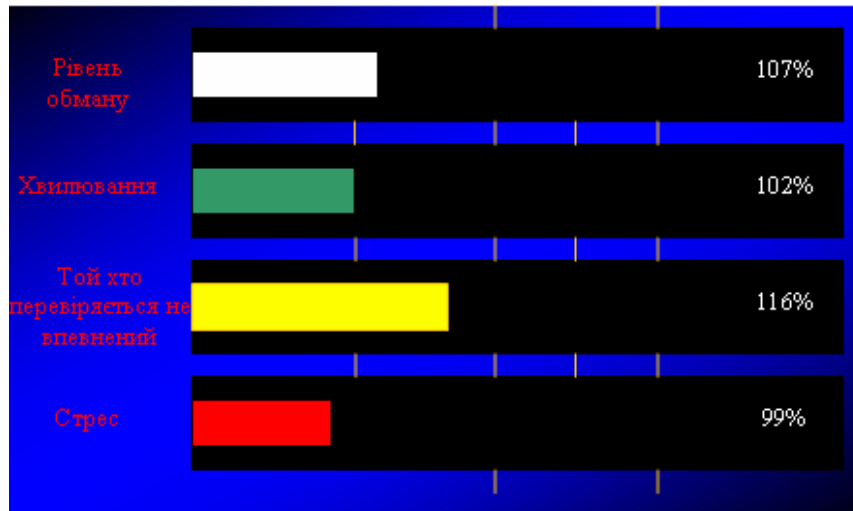


Рис. 9. Робоче вікно програми блоку емоцій

Наступним шагом роботи є моделювання мовного сигналу по біосигналу, як найінформативнішої ознаки розпізнавання психічного та емоційного стану по голосу мовця.

За означенням періодично корельований випадковий процес (ПКВП) – це такі випадкові процеси (ВП), математичні сподівання та симетричні коваріації яких є періодичними функціями часу щодо сукупності часових зсувів їх аргументів

$$m(t+T) = m(t), \quad r(t+T, s+T) = r(t, s), \quad \forall t, s \in R, \quad (6)$$

де $T > 0$ – період корельованості.

Математичне сподівання і зсувова коваріація (функція кореляції $b(t, u) = r(t+u, t)$) є періодичними з періодом T :

$$m(t+T) = m(t), \quad b(t+T, u) = b(t, u), \quad \forall u \in R.$$

Голосні звуки при різній ритміці серця породжуються системою (серцем), що знаходиться в стані "пульсуючої" рівноваги. Оцінки, обчислені за експериментальними часовими рядами голосного звуку [а] в різні моменти часу, характеризуються відмінністю середніх значень, дисперсій, кореляційних функцій. Згідно з критеріями Фішера для дисперсій та рівня значимості $Z_{0,95}$ гіпотеза про стаціонарність випадкових процесів відкидається; процес класифікується як нестаціонарний відносно критерію.

$$\xi_C(t_j) = \{\xi(t_j + kT), k \in Z, t_j \in [0, T)\}. \quad (7)$$

де $T > 0$ – період корельованості.

Формули, виведені з цієї властивості, тобто зведенням до теорії стаціонарних векторних послідовностей:

$$\hat{m}_{k_j}(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \xi(t_0 + kT) \quad \hat{b}_\xi(t, u) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \xi(t + u + kT) \overline{\xi(t + kT)}, \quad (8)$$

дають слушні незсунуті оцінки характеристик ПКВП синфазним методом. Для скінченного відрізка ПКВП (довжиною $\theta = NT$) формули оцінок (8) мають вигляд:

$$\hat{m}_{k_j}(t) = \frac{1}{N} \sum_{k=0}^{N-1} \xi(t_0 + kT) \quad \hat{b}_\xi(t, u) = \frac{1}{N} \sum_{k=0}^{N-1} \left[\xi(t + kT) \right] \times \left[\xi(t + u + kT) \right]. \quad (9)$$

Тому критерієм вибору оцінки періоду корельованості із пробних значень вибрано міру рельєфності кривих $\hat{m}_{k_j}(t_0|T_p)$, $\hat{b}_\xi(t_0, u=0|T_p) = \hat{d}_\xi(t_0|T_p)$, $t_0 \in [0; T_p)$ – відому в теорії функцій характеристику – варіацію функції:

$$V(\hat{m}_\xi; [T_1; T_2]) = \sup_{\forall D_n([T_1; T_2])} \sum_{k=0, n-1} |\hat{m}_\xi(t_{k+1}) - \hat{m}_\xi(t_k)|, \quad (10)$$

$$V(\hat{d}_\xi; [T_1; T_2]) = \sup_{\forall D_n([T_1; T_2])} \sum_{k=0, n-1} |\hat{d}_\xi(t_{k+1}) - \hat{d}_\xi(t_k)|, \quad (11)$$

де $D_n([T_1; T_2])$ – розбиття інтервалу $[T_1; T_2]$. Ця величина є невід'ємною і адитивною, тобто вона справді є мірою в математичному сенсі. Для підсилення рельєфності використано інші аналогічні характеристики – Φ -варіації $V_\Phi(f; [a; b])$ ($u \geq 0, \Phi(u) = 0$) за означенням:

$$V_{\Phi}(f; [a; b]) \stackrel{\Delta}{=} \sup_{\forall D_n([a; b])} \sum_{k=0; n-1} \Phi(|f(t_{k+1}) - f(t_k)|). \quad (12)$$

метрика Бора-Безиковича

На рис. 10 зображено залежності варіацій лінійної ($\Phi(u) = U$), квадратичної ($\Phi(u) = U^2$) та кубічної ($\Phi(u) = U^3$) функцій статистичних оцінок при різних пробних значеннях базових частот як значень відповідних функціоналів.

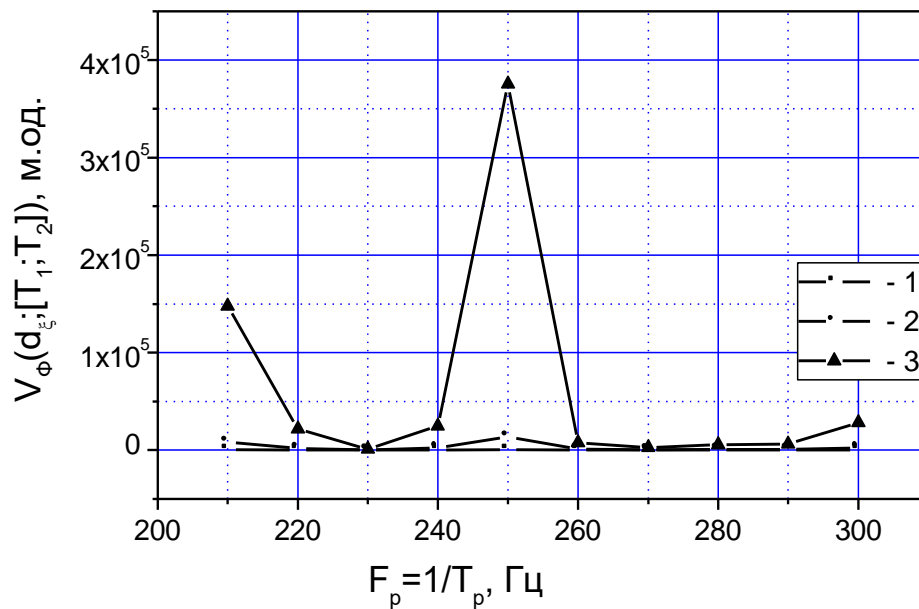
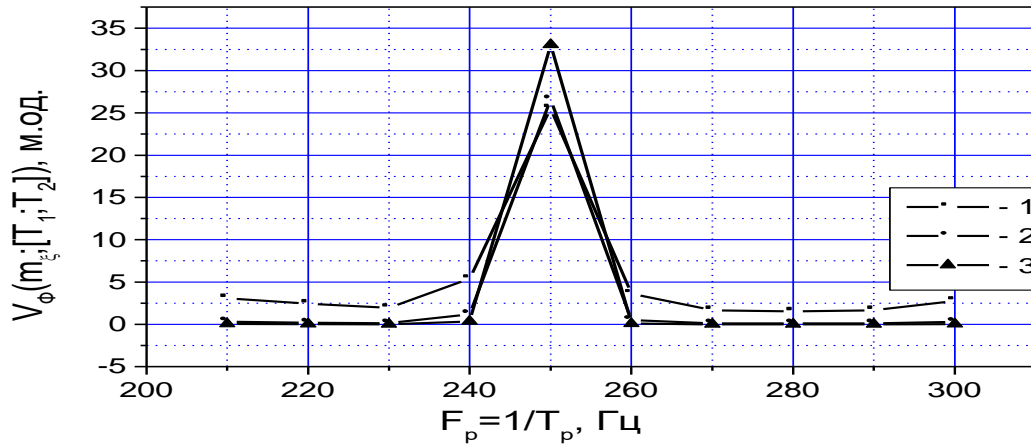


Рис. 10. Значення варіацій при різних пробних частотах $F_p = 1/T_p$: 1 – $\Phi(u) = U$; 2 – $\Phi(u) = U^2$; 3 – $\Phi(u) = U^3$.

Отже, за оцінку істинного значення періоду корельованості – прийнято те значення, яке

забезпечує максимум функціоналів варіацій:

$$\hat{T} = \arg \max_{\{T_p\}} V_{\Phi}(\hat{m}_{\xi}; [T_1; T_2]), \quad \hat{T} = \arg \max_{\{T_p\}} V_{\Phi}(\hat{d}_{\xi}; [T_1; T_2]) \quad (13)$$

Значення \hat{T} для медичної норми РС однієї і тієї ж людини є однаковими. Тому його використано як часово-інваріантну характеристику.

Для перевірки припущення про зміну цієї характеристики при відхиленнях РС протягом перших 2-х секунд після ФП було зареєстровано голосні звуки [a].

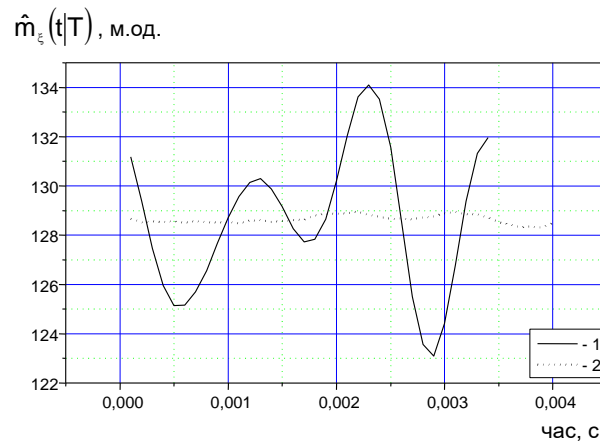
Для них пораховано статистики (8) із тим самим значенням \hat{T} , яке визначено для медичної норми ритму серця. Виявлено, що відповідні графіки мають на порядок менший розмах. Функціонали (9)-(11) досягли максимумів при інших значеннях T_p , ніж для медичної норми ритму для тієї ж людини.

Для підтвердження слушності припущення про період корельованості як інформативну характеристику стохастичної моделі сигналу та доцільності застосування розроблених алгоритмів проведено експериментальну верифікацію моделі.

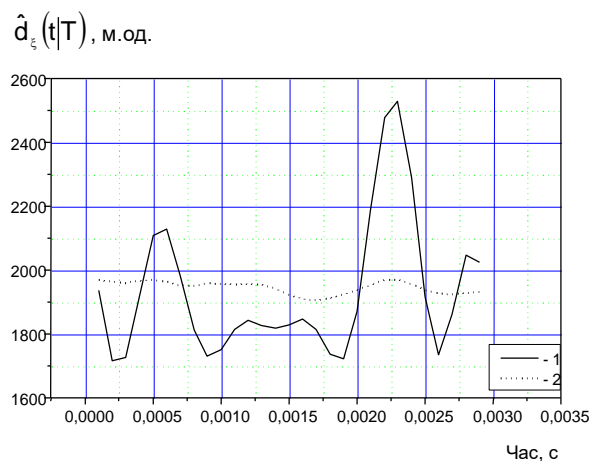
Результати верифікації підтвердили адекватність вибраної моделі –полівузькосмугового ПКВП – реальним сигналам: статистичні оцінки (математичне сподівання, дисперсія) флюктують біля періодичних функцій (рис. 11).

Показано, що період корельованості дійсно можна використати як інформативну ознаку. Порухення ритмічності серцевих скорочень внаслідок проведеної проби спричинює зміну періоду статистичних характеристик (періоду корельованості) моделі голосних звуків у вигляді корельованих випадкових процесів.

Спостерігається порушення ритміки статистичних характеристик голосового сигналу, яке можна пояснити відхиленням періоду корельованості після проведеної проби від його значення у нормі.



а)



б)

Рис. 11. Графіки оцінок математичного сподівання (а) та дисперсії (б) голосного звуку [а] для медичної норми ритміки серця (криві 1) та після ФП (криві 2)

Conclusions. У ході роботи зроблено наступні висновки:

1. Побудовано модель системи автоматичного розпізнавання голосних у потоці зливої мови, що не залежить від наявності шумів у мовному сигналі й має властивість незалежності від індивідуальних характеристик голосу мовця. Модель системи автоматичного розпізнавання мови будується на основі акустичного аналізу мовного сигналу.

2. Представлено результат періодично-аперіодичної моделі декомпозиції з авторегресійною параметризацією й коефіцієнтом масштабування. Результат дослідження показав, що якість синтезованої мови дуже висока, і запропонована модель збігається з періодично-аперіодичною

декомпозицією, у той же час модель має менший порядок і більш стабільна. Крім того, запропонована модель може бути застосована в системі конверсії голосу.

3. Проведено аналіз вибору цифрового аналізатора, а також фільтра мовного аналізатора. Методом математичного моделювання показано, що оптимальна форма імпульсу збудження сходового фільтра для різних голосних практично збігається й може бути апроксимована трапецієвидним імпульсом. Проведено моделювання блоку розпізнавання емоцій мовця.

4. Розвинуто часовий метод аналізу голосових сигналів, який враховує не тільки гармонічний склад та спектральний розподіл потужності сигналу, але й часову зміну його властивостей. Обґрунтовано використання синфазного методу аналізу періодично корельованих випадкових процесів для визначення змін періоду корельованості, пов'язаних зі змінами РС. Показано ефективність застосування запропонованих методів діагностики РС у криміналістиці, об'єктивній експрес діагностиці.

Conflict of interest statement: The authors state that there are no conflicts of interest regarding the publication of this article.

REFERENCES:

1. Бабский Е, Зубков А. Физиология человека. 1966;:655.
2. Мирошников Д. Современные решения беспроводного доступа. Электросвязь. 1998;4:33-37.
3. Карпов О. Технология построения устройств распознавания речи. Изд-во ДНУ; 2001.
4. Коротаяев Г. Системы анализа и синтеза речевого сигнала с линейным предсказанием. Зарубежная радиоэлектроника. 2018;10:3-14.
5. Галкин В. Цифровая мобильная радиосвязь. Москва: Телеком; 2007.
6. Новікова А, Головець Н. ПОБУДОВА МОДЕЛІ ЗВ'ЯЗКУ АНТРОПОМЕТРИЧНИХ ТА ДЕРМАТОГЛІФІЧНИХ ПАРАМЕТРІВ ОБЛИЧЧЯ. Биомедицинская инженерия и электроника [Internet]. 2017 [cited 15 October 2018];5(19):38-41. Available from: <http://biofbe.esrae.ru/pdf/2017/5/1136.pdf>

100% Unique

Total 15033 chars (**2000 limit exceeded**) , 222 words, 15 unique sentence(s).

Essay Writing Service - Paper writing service you can trust. Your assignment is our priority! Papers ready in 3 hours! Proficient writing: top academic writers at your service 24/7! Receive a premium level paper!

Results	Query	Domains (original links)
Unique	робота присвячена створенню моделей блоків різного функціонального призначення аналізатора мови й голосу мовця	-
Unique	Проведено аналіз фізіологічної появи мови/голосу, а також автоматизованого розпізнавання мови/голосу мовця	-
Unique	Проведено моделювання розпізнавання змін у психоемоційному поведженні людини по біосигналу при зміні голосу/мови мовця	-
Unique	Keywords: вокодер, голос, автоматизована система	-
Unique	Останнім часом це завдання стало особливо актуальним при проведенні антитерористичних заходів	-
Unique	Голос також є унікальною біометричною характеристикою людини	-
Unique	Велике значення має фізичне й математичне моделювання	-
Unique	дані важливі для мовознавства, логопедії, техніки зв'язку, конструювання систем автоматичного розпізнавання й синтезу мови	-
Unique	Математична модель сигналу вирішальною мірою визначає ефективність функціонування інформаційно-вимірвальних систем, які забезпечують отримання	-
Unique	DOI: LCC - № RZ201-999 Моделювання мовного аналізатора Римма Риженко1, Всеволод Новіков11 Херсонський національний технічний університет Address for	-
Unique	На основі проведеного аналізу побудована функціональна модель цифрового аналізатора мови, а також фільтра	-
Unique	Основними завданнями біометрії є верифікація й ідентифікація людини по унікальним, властивим тільки йому біологічним	-
Unique	До основних біометричних технологій відносять розпізнавання особистості по відбитках пальців, геометрії обличчя або кисті руки.	-
Unique	Перевагами голосової біометрії є безконтактність методу, використання стандартного встаткування, наприклад стільникового телефону, і можливість одержання	-
Unique	Зміна голосу мовця відображається на його мові, емоціях, жестах, але також можуть бути видні зміни	-