

досліджень, метою яких є синтез нових аналітичних моделей діяльності та процесів захисту інформації у рамках системодіяльничної методології захисту інформації. Отримані результати мають важливе значення для удосконалення організаційно-технічних форм та методів захисту інформації. У роботах [2, 12 – 14] викладаються основні положення системодіяльничної методології захисту інформації як складової частини системного підходу до вирішення проблем забезпечення безпеки інформації в складних організаційно-технічних системах.

У даній роботі по суті запропонована нормативна модель мікроструктури діяльності із захисту інформації на основі застосування методів морфологічного аналізу нормативних множин заходів безпеки, застосування методології SADT та методів функціонального моделювання IDEF0 і ARIS. Модель отримана вперше і являє собою нові знання відносно структури діяльності із захисту інформації. Результати носять методичний і теоретичний характер та можуть використовуватися на практиці для визначення організаційної системи захисту інформації. Отримані результати є елементами теоретичних та науково-методичних основ процесного підходу до захисту інформації.

Література: 1. Потій А. В. Формалізована модель діяльності // *Радіоелектронні і комп'ютерні системи. Науково-технічний журнал.* - № , 2007. – С. XX-XX 2. Потій А. В. Эталонная модель системы процессов защиты информации. Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. Науково-технічний збірник – Вип. 12 – Київ, 2006. – С. 17-31. 3. ISO/IEC 21827: 2002 Information technology - Systems Security Engineering - Capability Maturity Model 4. НД ТЗІ 1.4-001-2000. Типове положення про службу захисту інформації в автоматизованій системі. 5. ISO/IEC 17779:2000 Code of practice for information security management 6. NIST SP 800-26. Security Self-Assessment Guide for Information Technology Systems 7. Bundesamt fur Sicherheit in der Informationstechnik. IT Baseline Protection Manual, 1998. 8. ДСТУ ISO/IEC TR 13335-2:2003. Інформаційні технології. Настанови з керування безпекою інформаційних технологій. Частина 2: Керування та планування безпеки інформаційних технологій. 9. ISO/IEC 27001:2005 (BS 7799-2:2005) Information technology Security techniques– Information Security Management Systems. 10. NIST SP 800-53. Recommended Security Controls for Federal Information Systems. R. Ross, S. Katzke, A. Johnson, M. Swanson, G. Stoneburner, G. Rogers, A. Lee – 2005. 11. РД IDEF 0. Методология функционального моделирования IDEF0. Руководящий документ. – Госстандарт России, Москва.- 2000. 12. Потій А. В. Формальная модель процесса защиты информации // *Радіоелектронні і комп'ютерні системи. Науково-технічний журнал.* - №5, 2006. – С. 75-80 13. Бондаренко М. Ф., Потій О. В. Визначення та обґрунтування суті політики інформаційної безпеки // *Радиотехника. Всеукраїнський міжвед. Научн.-техн. Сб.* – 2003. – Вип. 134. – С. 9-25 14. Потій О. В. Процесний підхід до управління безпекою інформації//VIII Международная научно-практическая конференция "Безопасность информации в информационно-телекоммуникационных системах", 11-13 мая 2005. Тезисы докладов. – К.: НИЦ "Тезис", 2005. – С. 35-36.

УДК 621.396

МЕТОДИКА ОТРИМАННЯ ОБРАЗУ МОВНОГО СИГНАЛУ ДЛЯ ВИРІШЕННЯ ЗАВДАННЯ ЩОДО ІДЕНТИФІКАЦІЇ ДИКТОРА

Максим Кузнецов

Національна академія СБ України

Анотація: Розглянуто новітню методику спектральної обробки мовних сигналів, застосування якої дозволяє значно підвищити ефективність систем ідентифікації джерела сигналу.

Summary: This article is about the new speech-processing method, which allows to improving of the speech source identification systems efficiency greatly.

Ключові слова: Ідентифікація мови диктора, спектральна обробка, порядкові статистики, статистичні критерії згоди, голосовий статистичний образ.

Вступ

На сьогоднішній день повною мірою невіршеним та актуальним залишається загальнонаукове завдання щодо ідентифікації мови диктора. Розв'язанню цієї проблеми існуючими засобами заважає значна кількість перешкод, існування яких пов'язано із природою мовного сигналу, який є випадковим нестационарним процесом зі змінною дисперсією і складною формою поточної спектральної щільності потужності. Деякі з цих перешкод: неможливість встановлення точного переліку інваріантних ознак, які є характерними для

диктора, мовний сигнал якого досліджується, обмежений об'єм баз даних голосових статистичних образів, малоінформативне та складне для розуміння подання голосових статистичних образів, обчислювальна складність отримання голосового статистичного образу, неоднозначність висновків різних експертів відносно результатів дослідження однієї і тієї ж фонограми, та ін. [1 – 4].

В основі всіх методів ідентифікації диктора (джерела мовного сигналу) лежить задача обробки мовного сигналу. Мета рішення цієї задачі – отримання сукупності даних (голосового статистичного образу джерела мовного сигналу) про досліджуваний сигнал, які будуть однозначно характеризувати джерело мовного сигналу та демонструвати вплив характерних особливостей будови мовного тракту диктора, які відображаються на спектрі його голосового сигналу. Голосовий статистичний образ джерела мовного сигналу має бути правдоподібним описом досліджуваного сигналу, зручним для подальшого застосування (зіставлення з голосовим статистичним образом еталонного джерела сигналу) і зрозумілим для вивчення.

В статті розглядається методика спектральної обробки мовних сигналів, яка є новітньою, та заснована на використуванні залежності від частоти спектру порядкових статистик – квантилей $\hat{x}_\alpha(f_k)$ (квантилей, децилей, п'ятилей, процентилей) модуля спектру Фур'є (Хартлі) мовних сигналів, робастною до аномальних шумів природного і штучного походження, перешкод в каналах передачі, реєстрації і відтворення мовних сигналів в умовах негауссовости розподілів статистичних характеристик спектру [5– 8].

I Постановка завдання

Ознайомлення спеціалістів галузі обробки мовних сигналів з методикою спектральної обробки мовних сигналів, застосування якої є можливим при вирішенні завдання щодо ідентифікації джерела мовних сигналів.

II Основна частина

Спектральна ідентифікація мовних сигналів, як складова задачі розпізнавання мови, виробляється шляхом перевірки статистичних критеріїв згоди розподілів параметрів поточного спектру мовного сигналу з розподілами однотипних параметрів спектрів еталонних мовних реалізацій [1, 2, 9]. Основна причина зниження ефективності відомих систем спектральної ідентифікації мовних сигналів обумовлена природою мовного сигналу, який є випадковим, нестационарним процесом, з повільно змінною дисперсією і складною формою поточної спектральної щільності потужності сигналу. В результаті цього багатовимірні функції щільностей розподілу ймовірностей параметрів поточних спектрів Фур'є мовного сигналу є негаусовськими і полімодальними, що суттєво ускладнює застосування відомих статистичних критеріїв згоди і знижує ймовірність правильної ідентифікації [2 – 4, 9].

В сучасних системах спектральної ідентифікації використовуються цифрові методи обробки мовних сигналів. У всьому діапазоні частот спектру мовного сигналу, що ідентифікується, виконують наступні взаємозв'язані операції аналого-цифрової обробки: низькочастотну фільтрацію в смузі $3 \div 10$ кГц; поділ сигналу на сегменти однакової тривалості $10 \div 30$ мс; стандартне $9 \div 12$ – бітове аналого-цифрове перетворення з частотою дискретизації $6 \div 20$ кГц на кожному сегменті аналізу; формування на кожному сегменті аналізу дискретних відліків спектру сигналу шляхом застосування стандартного дискретного перетворення Фур'є. В деяких випадках використовують дискретне косинус-перетворення Фур'є.

Для ідентифікації мовного сигналу за всією сукупністю відліків спектральної щільності Фур'є визначають статистичні характеристики параметрів спектру для кожної частоти дискретного спектру мовного сигналу за всіма сегментами аналізу. Отримані статистичні характеристики параметрів спектру порівнюють за відомими критеріями згоди з відповідними статистичними характеристиками еталонних мовних реалізацій бази даних [2, 4, 10].

Природно, що на практиці мають місце суттєві відхилення розподілів даних параметрів від нормальних розподілів, спостерігається нестационарний характер та великий динамічний діапазон зміни значень інтенсивностей спектральних компонент самого мовного сигналу. Все це обумовлює низьку перешкодозахисну здібність систем спектральної ідентифікації в умовах шумів природного та штучного походження, перешкод в каналах передачі, реєстрації і відтворення мовних сигналів, наявність аномальних перешкод.

В цих умовах підвищення ефективності спектральної ідентифікації може бути отримано на основі використування порядкових статистик – квантилей (квантилей, децилей, п'ятилей, процентилей) – спектральних параметрів мовних сигналів, які є робастними до розглянутих дестабілізуючих чинників, в умовах негауссовости розподілів статистичних характеристик спектру. Розглянемо метод спектральної ідентифікації, заснований на порівнянні порядкових статистик спектральних параметрів мовних сигналів, що дозволяє зменшити вплив чинників, які є дестабілізуючими, негауссовости статистичних характеристик

спектру сигналу на ймовірність правильної ідентифікації [8].

Останнім часом в системах цифрової обробки сигналів широко застосовують аналізатори спектру, в яких ядро перетворення сигналу побудовано на основі дискретного перетворення Хартлі. Це обумовлено речовинним характером спектральної щільності Хартлі речовинного мовного сигналу; ідентичністю прямого і зворотного перетворень Хартлі та їх взаємодозначним зв'язком з дискретним перетворенням Фур'є, і, також, спроможністю щодо розподілу спектральних компонент парних та непарних складових фонем мовного сигналу [3 – 5, 11].

В результаті виконання розглянутих вище операцій аналого-цифрового та часо-частотного перетворення мовного сигналу, що ідентифікується, для кожного дискретного значення частоти його поточного спектру Фур'є (Хартлі) може бути отримана сукупність значень модулів $x_i, i = 1 \dots n$, за всіма сегментами аналізу. Ця сукупність незалежних випадкових величин (бо тривалість інтервалу сегментації суттєво перевершує величину інтервалу кореляції мовного сигналу [9]) може вважатися вибіркою об'єму n (кількість сегментів аналізу) з генеральної сукупності, що характеризується функцією розподілу (ФР) $G(x)$ [11].

В даних умовах проведення експерименту ідентифікації, зміна початку відліку (використання правила помилкового нуля), масштабу і ладу (перемішування, перенумерації) значень $x_i, i = 1 \dots n$, ніяк не повинне позначатися на результатах статистичного аналізу. Тому, до проведення будь якої обробки значень $[x_i]$, та оцінювання статистик вибірки, можна сформувати варіаційний ряд $[\hat{x}_{i/n}]$ значень $[x_i]$:

$$\hat{x}_{1/n} \leq \hat{x}_{2/n} \leq \dots \leq \hat{x}_{i/n} \leq \dots \leq \hat{x}_{n-1/n} \leq \hat{x}_{n/n},$$

розташували їх в міру зростання, де індекс елементів варіаційного ряду $[\hat{x}_{i/n}]$ запишемо у вигляді i/n , тобто $\hat{x}_{1/n} = \min_i [x_i]$, $\hat{x}_{n/n} = \max_i [x_i]$.

Варіаційний ряд $[\hat{x}_{i/n}]$ містить всю інформацію про початкову вибірку $[x_i]$ і тому він називається тривіальною достатньою статистикою всієї сукупності незалежних випадкових величин $[x_i]$. Значення $x_i, i = 1 \dots n$ реалізації $[x_i]$ випадкові та, в загальному випадку, не передбачувані з розумною для практики точністю. Але якщо від вибірки $X_k = [x_i]_k$ до вибірки $X_m = [x_i]_m$, від однієї реалізації вектора X_k до іншої реалізації X_m значення $x_i, i = 1 \dots n$ є цілком випадковими величинами і незалежні між собою, то в значеннях елементів варіаційного ряду $[\hat{x}_{i/n}]$ розпочинають виявлятися статистичні зв'язки і моменти стійкості передбачення значень цих елементів з допустимою на практиці точністю [11].

Розглянемо основні характеристики і параметри елементів варіаційного ряду $[\hat{x}_{i/n}]$. Якщо довільне значення $x_i, i = 1 \dots n$, може розглядатися як вибіркоче значення безперервної випадкової величини x з генеральної сукупності, що характеризується функцією розподілу $G(x)$ або функцією щільності розподілу ймовірності (ПРВ) $g(x) = \frac{d}{dx} G(x)$, то значення $\hat{x}_{i/n}, i = 1 \dots n$, елементів варіаційного ряду $[\hat{x}_{i/n}]$ є вибіркочими значеннями безперервних випадкових величин $x_{i/n}$ генеральних сукупностей. В загальному випадку $G(x_{i/n})$ випадкової величини $x_{i/n}$ має вигляд [11]:

$$G(x_{i/n}) = \sum_{k=i}^n C_n^k G^k(x) [1 - G(x)]^{n-k},$$

де $C_n^k = \frac{n!}{k!(n-k)!}$ – біноміальні коефіцієнти, обчислення яких спрощується через властивості:

$$C_n^k = C_n^{n-k}; C_{n+1}^{k+1} = C_n^{k+1} + C_n^k, k < n.$$

Якщо $n=1$, тобто отримано одне значення x_1 і $x_{1/1} = x_1$, то $G(x_{1/1}) = G(x)$. Якщо $n=2$ і варіаційний ряд складається з двох елементів $\hat{x}_{1/2} \leq \hat{x}_{2/2}$, то $G(x_{1/2}) = 1 - [1 - G(x)]^2$,

$$G(x_{2/2}) = G^2(x).$$

При $n=3$ елементи варіаційного ряду

$$\widehat{x}_{1/3} \leq \widehat{x}_{2/3} \leq \widehat{x}_{3/3}$$

характеризуються функціями розподілу:

$$G(x_{1/3}) = G^3(x) - 3G^2(x) + 3G(x) = 1 - [1 - G(x)]^3;$$

$$G(x_{2/3}) = -2G^3(x) - 3G^2(x);$$

$$G(x_{3/3}) = G^3(x).$$

В загальному випадку, функція щільності розподілу ймовірності $g(x_{i/n})$ випадкової величини $x_{i/n}$ має вигляд [11]:

$$g(x_{i/n}) = \frac{C_n^i [1 - G(x)]^{n-i} d}{dx G^i(x)};$$

$$G(x_{i/n}) = C_n^i \int_{-\infty}^x [1 - G(x)]^{n-i} dx G^i(x).$$

З урахуванням того, що випадкова величина x з функцією розподілу $G(x)$ при функціональному перетворенні $y = G(x)$ трансформується у випадкову величину y , рівномірно розподілену на інтервалі $[0, 1]$, можна записати:

$$G(x_{i/n}) = C_n^i \int_0^{G(x)} [1 - t]^{n-i} dt^i$$

для довільної функції розподілу $G(x)$ генеральної сукупності вибірки $[x_i]$.

В загальному випадку статистичні характеристики елементів $\widehat{x}_{i/n}$ варіаційного ряду $[\widehat{x}_{i/n}]$ вибірки $[x_i]$ об'єму n можуть суттєво відрізнятися від статистичних характеристик самої вибірки [11]. Це обумовлено тим, що операція формування впорядкованої сукупності елементів варіаційного ряду $[\widehat{x}_{i/n}]$ з вибірки незалежних значень $[x_i]$ безперервної випадкової величини x (ранжирування вибірки) не є лінійною операцією. Проте, при асимптотичному збільшенні об'єму n вибірки (при $n \rightarrow \infty$) розподіл значень елементів варіаційного ряду $\widehat{x}_{k/n}, \widehat{x}_{m/n}$, які не є екстремальними $1/n < k/n < m/n < n/n$, та прагне нормального розподілу:

$$g(x_{k/n}) = c_{k/n} \exp[-|x_{k/n} - \bar{x}_{k/n}|^2 / 2\delta_{k/n}^2];$$

$$g(x_{m/n}) = c_{m/n} \exp\left[-\frac{|x_{m/n} - \bar{x}_{m/n}|^2}{2\delta_{m/n}^2}\right].$$

Середні значення $\bar{x}_{k/n}, \bar{x}_{m/n}$ (математичні очікування), задовольняють рівнянням:

$$G(\bar{x}_{k/n}) = k/n;$$

$$G(\bar{x}_{m/n}) = m/n, \quad (1)$$

а середньквдратичні відхилення $\delta_{k/n}, \delta_{m/n}$ описуються виразами:

$$\delta_{k/n} = \sqrt{\frac{k(n-k)}{ng(\bar{x}_{k/n})}};$$

$$\delta_{m/n} = \sqrt{\frac{m(n-m)}{ng(\bar{x}_{m/n})}}, \quad (2)$$

при цьому випадкові величини $x_{k/n}$, $x_{m/n}$ є корельованими випадковими величинами, хоча випадкові величини $[x_i]$, з яких формувався варіаційний ряд $[\hat{x}_{i/n}]$, були незалежні. Коефіцієнт кореляції $\rho_{km/n}$ елементів $x_{k/n}$, $x_{m/n}$ записується так

$$\rho_{km/n} = \sqrt{\frac{\frac{k}{m} - \frac{1 - \frac{m}{n}}{1 - \frac{k}{n}}}{1 - \frac{k}{n}}} = \sqrt{\frac{k(n-m)}{m(n-k)}},$$

незалежно від виду розподілу $G(x)$ (функції щільності розподілу ймовірностей $g(x)$ генеральної сукупності).

Відзначимо, що при збільшенні об'єму вибірки в ρ раз таким чином, що відносини k/n і m/n залишаються незмінними, тобто $k/n = \rho k / \rho n$, а $\rho m / \rho n = m/n$, значення математичних очікувань $\bar{x}_{k/n}$, $\bar{x}_{m/n}$ не змінюються

$$\bar{x}_{k/n} = \bar{x}_{\rho k / \rho n}, \quad \bar{x}_{m/n} = \bar{x}_{\rho m / \rho n}$$

через особливості рівняння зв'язку (1). При цьому значення дисперсії $\delta_{k/n}^2, \delta_{m/n}^2$ зменшуються в ρ раз:

$$\delta_{\rho k / \rho n}^2 = \frac{\delta_{k/n}^2}{\rho}; \quad \delta_{\rho m / \rho n}^2 = \frac{\delta_{m/n}^2}{\rho},$$

а значення коефіцієнтів кореляції $\rho_{km/n}$ також не змінюються:

$$\rho_{\rho k \rho m / \rho n} = \rho_{km/n}.$$

У тому випадку, коли $\frac{m}{n} = 1 - \frac{k}{n}$, $k < m < n$, справедливі співвідношення: $(\bar{x}_{k/n} + \bar{x}_{m/n}) / 2 = x_{1/2}$; $\delta_{k/n} = \delta_{m/n}$; $\rho_{km/n} = k/m = k/(n-k)$, де $x_{1/2}$ – медіана функції щільності розподілу ймовірностей $g(x)$, для якої:

$$\int_{-\infty}^{x_{1/2}} g(x) dx = \int_{x_{1/2}}^{\infty} g(x) dx = 0,5.$$

Таким чином, вектор з елементів варіаційного ряду $[\hat{x}_{i/n}]$ модулів спектральної щільності Фур'є (Хартлі) для кожного значення частоти спектру (за винятком екстремальних значень $\hat{x}_{1/n}, \hat{x}_{n/n}$) є асимптотично (при $n \rightarrow \infty$) випадковим гаусовським вектором при довільному законі $G(x)$ розподілу елементів генеральної сукупності, якщо похідна $g'(x)$ функції щільності розподілу ймовірностей безперервна [11].

За значеннями номерів i/n елементів варіаційного ряду $[\hat{x}_{i/n}]$ при достатньому значенні об'єму вибірки n визначають вибіркові квантілі \hat{x}_α : $\hat{x}_{j/4}$, $j = 1, 2, 3$ – вибіркові кватилі ($\hat{x}_{0,25}, \hat{x}_{0,5}, \hat{x}_{0,75}$); $\hat{x}_{j/10}$, $j = 1 \dots 9$ – вибіркові децилі ($\hat{x}_{0,1}, \hat{x}_{0,2}, \dots, \hat{x}_{0,9}$); $\hat{x}_{j/20}$, $j = 1 \dots 19$ – вибіркові п'ятилі ($\hat{x}_{0,05}, \hat{x}_{0,1}, \hat{x}_{0,15}, \hat{x}_{0,20}, \dots, \hat{x}_{0,95}$); $\hat{x}_{j/100}$, $j = 1 \dots 99$ – вибіркові процентилі ($\hat{x}_{0,01}, \hat{x}_{0,02}, \dots, \hat{x}_{0,99}$), які ділять весь діапазон $[\hat{x}_{1/n}, \hat{x}_{n/n}]$ зміни значень варіаційного ряду $[\hat{x}_{i/n}]$ відповідно на 4, 10, 20 і 100 нерівних за величиною підінтервалів, в які потрапляє однакова кількість значень ряду. Через випадковий характер мовних процесів значення кватилей, децилей і процентилей \hat{x}_α , що отримали в літературі загальну назву порядкових статистик, є випадковими і від вибірки до вибірки міняються цілком незалежно. В загальному випадку порядкові статистики \hat{x}_α можна розглядати як впорядковану сукупність

нових випадкових величин (випадковий вектор), статистичні характеристики якої залежать від виду функції щільності розподілу ймовірностей $g(x)$ і об'єму вибірки n . Разом з тим, при $n \rightarrow \infty$ розподіл кожної порядкової статистики \hat{x}_α варіаційного ряду $[\hat{x}_{i/n}]$ є асимптотично нормальним

$$g(x_\alpha) = c_\alpha \exp \left[\frac{-|x - \bar{x}_\alpha|^2}{2\delta_\alpha^2} \right],$$

з центром (математичним очікуванням \bar{x}_α), що задовольняє рівнянню $G(\bar{x}_\alpha) = \alpha$, та середньоквадратичним відхиленням: $\delta_\alpha = \sqrt{\alpha(1-\alpha)/n/g(\bar{x}_\alpha)}$, якщо лише статистика x_α генеральної сукупності єдина і похідна функції щільності розподілу ймовірностей $g'(x)$ існує і безперервна в околі $x = x_\alpha$.

Коефіцієнт кореляції $\rho_{\alpha,\beta}$ порядкових статистик x_α, x_β описується виразом $\rho_{\alpha,\beta} = \sqrt{\frac{\alpha(1-\beta)}{\beta(1-\alpha)}}$ незалежно від типу розподілу $G(x)$ (див. вирази 1 та 2). При $n \rightarrow \infty$ значення кватилей, децилей, п'ятилей і процентилей \hat{x}_α вибірки $[x_i]$ потребують відповідних значень параметрів x_α функцій $g(x)$ і $G(x)$ генеральної сукупності, для яких:

$$\int_{-\infty}^{x_\alpha} dG(x) = \int_{-\infty}^{x_\alpha} g(x) dx = \alpha.$$

Таким чином, при переході від аналізу вибірки $[x_i]$ з генеральної сукупності, що характеризується функціями $g(x)$ і $G(x)$ (вигляд і параметри яких на практиці звичайно невідомий), до аналізу варіаційного ряду $[\hat{x}_{i/n}]$ суттєво спрощується отримання оцінок параметрів нових розподілів. Це обумовлено тим, що за винятком екстремальних значень $[\hat{x}_{1/n}, \hat{x}_{n/n}]$ сукупність нових випадкових величин є асимптотично нормальною. Крім того, вибіркові кватилі, децилі, п'ятилі, процентілі є робастними до аномальних значень спектральних компонент $[x_i]$, які виходять з ряду і розташовані на "хвостах" варіаційного ряду $[\hat{x}_{i/n}]$.

Приведені результати показують, що варіаційний ряд $[\hat{x}_{i/n}]$ є статистичним еквівалентом сукупності незалежних значень $[x_i]$. При цьому, з допустимим для практики точністю, асимптотичні властивості кватилей і децилей варіаційного ряду розпочинають виявлятися при числі вимірів $n > 50$, а стійкість результатів їх статистичної обробки виявляється практично при $n > 100$. Все це підкреслює перевагу аналізу кватилей, децилей, п'ятилей, процентилей варіаційного ряду порівняно з аналізом статистичних характеристик вибірки прямих вимірювань при $n > 1000$.

В загальному випадку визначеність асимптотичних властивостей порядкових статистик суттєво спрощує отримання оцінок максимальної правдоподібності і перевірку критеріїв згоди параметрів розподілів в умовах статистичної невизначеності характеристик шумів і помилок вимірювань [6].

Для нормування елементів варіаційного ряду і приведення його до стандартизованого вигляду використовують наступні оцінки параметра положення:

$$m_1 = \text{med}(\hat{x}_{i/n}) = \hat{x}_{0,5}$$

– медіану ряду;

$$m_2 = (\hat{x}_{0,25} + \hat{x}_{0,5} + \hat{x}_{0,75}) / 3$$

– середнеквартильне;

$$m_3 = (0,25\hat{x}_{0,25} + 0,5\hat{x}_{0,5} + 0,75\hat{x}_{0,75})$$

– усічене середнє Т'юки.

Для оцінки параметра розсіяння використовують кватильне розширення

$$s_1 = \hat{x}_{0,75} - \hat{x}_{0,25}$$

та медіану абсолютних відхилень значень елементів варіаційного ряду $[\hat{x}_{i/n}]$ від медіани ряду $\hat{x}_{0,5}$ [11]

$$s_2 = med(|\hat{x}_{i/n} - \hat{x}_{0,5}|).$$

Таким чином, в результаті виконання розглянутих вище операцій аналого-цифрового перетворення мовного сигналу, що ідентифікується, для кожного дискретного значення частоти спектру Фур'є (Хартлі) $f_k, k=1...K$, можуть бути отримані вибіркові порядкові статистики $\hat{x}_\alpha(f_k)$ (квартилі, децилі, п'ятилі, процентилі) модульних значень спектральної щільності. Сформована залежність $\hat{x}_\alpha(f_k), k=1...K$, є статистичним еквівалентом всіх модулів поточних спектрів Фур'є (Хартлі) сегментів аналізу мовного сигналу, що ідентифікується. Характер цієї залежності відображає індивідуальні особливості зміни спектральної щільності потужності з частотою в усьому динамічному діапазоні мовного сигналу. Для ілюстрації, на рис. 1 та на рис. 2 приведені графіки реалізацій залежностей квартилей $\hat{x}_{0,25}(f_k), \hat{x}_{0,5}(f_k), \hat{x}_{0,75}(f_k), k=1...K$ модуля спектра Хартлі мовного сигналу від значень частоти спектру $f_k, k=1...K$, нормовані до максимуму залежність медіани $\hat{x}_{0,5}(f_k)$ для двох дикторів – чоловіка (рис. 1) та жінки (рис. 2). Просте зіставлення цих залежностей показує істотну відмінність спектральних характеристик мовних сигналів двох дикторів.

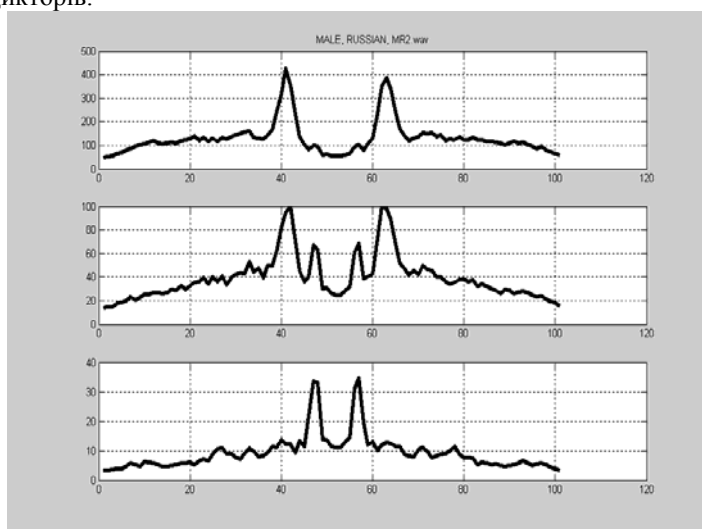


Рисунок 1 – Реалізація залежностей квартилей модуля спектра Хартлі мовного сигналу, Диктор – чоловік

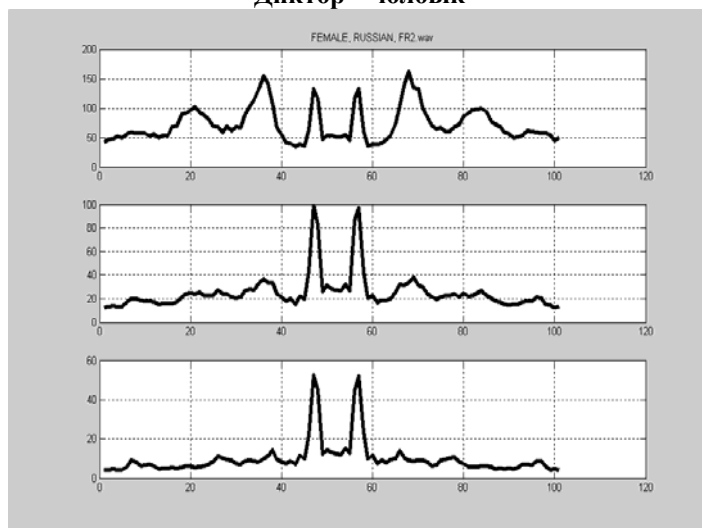


Рисунок 2 – Реалізація залежностей квартилей модуля спектра Хартлі мовного сигналу, диктор – жінка

Висновки

Отриманий в результаті застосування розглянутої методики голосовий статистичний образ диктора є правдоподібним описом сигналу, який досліджується; образ є зручним для подальшого використання та зрозумілий для вивчення. Передбачена деталізація представлення голосового статистичного образу диктора. Мовний сигнал, який піддається процедурі обробки за пропонованою методикою, не підлягає усередненню. В цьому випадку втрачається інформаційна повнота відображення характерних особливостей будови мовотворчого тракту диктора немає. Для вирішення завдання щодо ідентифікування джерела мовного сигналу усі сегменти аналізу мовної реалізації можуть бути задіяними. Суттєва асиметрія речовинного Хартлі спектру сигналу надає додаткові риси відмінності до голосових статистичних образів дикторів порівняно із комплексним спектром Фур'є сигналу, який досліджується.

Література: 1. Сапожков М. А. Речевой сигнал в кибернетике и связи – М.: Гос. изд-во л-ры по вопр. связи и радио, 1963.-с. 368 – 377. 2. Рамишвили Г. С. Автоматическое опознавание говорящего по голосу / Москва, «Радио и связь», 1981. – 224 с. 3. Новосельский А. Ф. Измерительный аппаратно – программный комплекс для идентификации личности по голосу: Дис. Канд.техн.наук: 05. 11. 16.-Киев, 1998. - 192 с. 4. Женило В. Р. Минаев В. А. Компьютерные технологии в криминалистических фоноскопических исследованиях и экспертизах / Учебное пособие.-М.: Академия МВД РФ, 1994.– 137 с. 5. Селетков В. Л. Соотношения связи одномерных преобразований Фурье и Хартли. // Радиоэлектроника, 2002.- № 7,- с. 46 – 50. (Изв. высш. учеб. заведений). 6. Селетков В. Л. "Соответствие операций обработки сигналов при использовании преобразования Хартли" // Радиоэлектроника, 2002.– № 6 – с. 36 – 44. (Изв. высш. учебн. заведений). 7. Селетков В. Л. "Модифицированное преобразование Хартли" // Радиоэлектроника, 1997.– № 1 – с. 75 – 77. (Изв. высш. учебн. заведений). 8. Кузнецов М. В. "Метод обработки речевых сигналов в системах спектральной идентификации" // Збірник наукових праць НА СБ України, 2006. – № 14. 9. Шелухин О. И. Лукьянцев Н. Ф. Цифровая обработка и передача речи. / Под. ред. О. И. Шелухина.- М.: Радио и связь, 2000.- 456.: ил.- с. 98 – 106. 10. Журавлев В. Н. "Анализ влияния частоты дискретизации на точность цифровой обработки речевых сигналов в системах биометрической идентификации" // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні, 2005.– № 10 – с.51 – 59. 11. Дейвуд Дж. Порядковые статистики: Пер. с англ. – М.: Мир, 1989.-540 с., ил.

УДК 681.3.06

СИНТЕЗ СИСТЕМИ ЗАХИСТУ ІНФОРМАЦІЇ, ОПТИМАЛЬНОЇ ЗА РІВНЕМ РИЗИКУ

Юрій Боня, Олексій Новіков

Фізико-технічний інститут Національного технічного університету України "КПІ"

Анотація: Запропоновані алгоритми розв'язання задач математичного програмування, що виникають в процесі проектування відкритих систем захисту інформації, оптимальних за рівнем ризику. Використання методів сепарабельного та лінійного бульового програмування дозволило побудувати працездатні та ефективні алгоритми, що можуть використовуватися як складові спеціалізованих систем автоматизованого проектування систем захисту інформації.

Summary: Solution algorithms of mathematical programming tasks were proposed, which emerge during design of information protection systems with open architecture which provide optimal risk level. Use of separable and linear Boolean programming methods made it possible to develop algorithms capable of efficient working as a component of particularized computer aided design systems of information protection systems.

Ключові слова: Системи захисту інформації, ризик, загроза, відкриті системи, автоматизоване проектування, бульове програмування, сепарабельне програмування.

І Вступ

Існують декілька концепцій побудови систем забезпечення мінімуму ризику (безпеки): "нульового ризику", "ненульового ризику" та змішаний підхід [1]. Концепція "нульового ризику" має іншу назву – детерміністський підхід при забезпеченні мінімуму ризику. В її основі лежить припущення, що можна виключити будь-яку небезпеку, якщо не пожалкувати грошей на побудову системи мінімізації ризику та забезпечити високий рівень дисципліни. Ця концепція є неадекватною щодо внутрішніх законів бізнесу,