

М. М. Биков, канд. техн. наук, проф.;
В. В. Ковтун, канд. техн. наук, доц.;
А. Раїмі, PhD, доц.

МЕТОД ВИДІЛЕННЯ ОСНОВНОГО ТОНУ НА ОСНОВІ МОДИФІКОВАНОЇ МАТЕМАТИЧНОЇ МОДЕЛІ СЛУХОВОЇ СИСТЕМИ ЛЮДИНИ

Розпізнавання голосів є важливою складовою надійної системи біометричної ідентифікації людини, які набувають все більшої актуальності в сучасному комп'ютеризованому світі. Однією з ознак, що використовуються для розпізнавання голосів, є частота (період) основного тону голосу, інформативність якої суттєво знижується за присутності в мовному сигналі низькочастотних шумів. Запропоновано метод виділення частоти основного тону, який, на відміну від існуючих, є стійкішим до наявності в мовному сигналі низькочастотних шумів.

Вступ

Створення надійної системи ідентифікації людини за індивідуальними особливостями її голосу досі є актуальною задачею, незважаючи на зусилля багатьох наукових колективів в світі та Україні. Мовний сигнал утворюється пропусканням повітряного потоку з легенів через голосові зв'язки (мовне джерело) та артикуляторний тракт, відповідно, індивідуальні особливості мовного сигналу зумовлені, в основному, специфікою будови та функціонування цих складових мовного апарату людини. Проте, індивідуальні особливості мовного сигналу, притаманні кожній людині голосові зв'язки, менше залежать від емоційного та фізіологічного стану людини [1, 2], тому їх доцільніше використовувати в якості індивідуальних ознак для автоматизованого розпізнавання особи. Літературні джерела стверджують [1, 2], що найефективнішою з таких ознак є період (частота) основного тону голосових зв'язок людини, проте відомі методи її виділення з мовного сигналу виявилися ефективними лише за лабораторних умов запису мовного сигналу, що не завжди можливо. Отже, постала задача створення методу визначення періоду (частоти) основного тону за присутності в мовному сигналі завад.

Метою проведеної роботи було вдосконалення методу виділення такої інформативної ознаки для розпізнавання особи за індивідуальними особливостями її голосу, як частоти основного тону. Існуючі методи виділення частоти основного тону бувають неефективними, якщо в мовному сигналі з'являються низькочастотні шуми, наприклад, внаслідок роботи неякісного акустичного тракту системи автоматизованого розпізнавання особи за індивідуальними особливостями її голосу. Авторами запропоновано метод виділення частоти основного тону, який поєднує амплітудні та кореляційні методи виділення частоти основного тону [1—3], з урахуванням методу нульових перетинань [1—3] та аналізу обвідної короткочасної енергії мовного сигналу [1—4], що покращує ефективність виділення частоти основного тону з мовного сигналу на тлі низькочастотних завад.

Розробка методу виділення основного тону на основі модифікованої математичної моделі слухової системи людини

Метод оснований на таких математичних засадах. Нехай $s(n) = s(0), s(1), \dots, s(N-1)$ — стаціонарний часовий ряд з нульовим середнім значенням. Відповідно до теореми Уолда [1], коваріаційна функція такого процесу має вигляд

$$\gamma(k) = \int_{-\pi}^{\pi} \cos(k\omega) dG(\omega), \quad (1)$$

де $G(\omega)$ — спектральна щільність потужності сигналу; $k = 0, \pm 1, \pm 2, \dots$ — часове зміщення.

Автокореляційна функція — нормована величина $R(k) = \gamma(k)/\gamma(0)$, яку з урахуванням (1) можна описати таким рівнянням:

$$R(k) = \frac{\int_{-\pi}^{\pi} \cos(k\omega) dG(\omega)}{\int_{-\pi}^{\pi} dG(\omega)}, \quad k = 0, \pm 1, \pm 2, \dots \quad (2)$$

Визначимо двосторонньо-обмежений бінарний ряд

$$y(n) = \begin{cases} 1, & s(n) \geq 0; \\ 0, & s(n) < 0, \end{cases} \quad 0 \leq n \leq N-1$$

і введемо індикаторну функцію F_n для моменту часу n

$$F_n = (y(n) - y(n-1))^2.$$

Якщо $F_n = 1$, то відбувається перетинання нульового рівня сигналом в момент часу n , в протилежному випадку $F_n = 0$. Кількість перетинань нульового рівня сигналом для ряду $s(n)$, $n = 0, 1, \dots, N-1$ визначається співвідношенням

$$W = \sum_{i=0}^{N-1} F_i.$$

Для процесу, що описується розподілом Гауса з нульовим середнім, математичне сподівання кількості перетинань нульового рівня сигналом, можна визначити як

$$E\{W\}_1 = (N-1) \left(\frac{1}{2} - \frac{1}{\pi} \arcsin r(1) \right),$$

де $E\{\}$ — математичне сподівання; $r(1)$ знаходимо за формулою (2), якщо $k = 1$.

Тоді

$$r(1) = \cos \left(\frac{\pi E\{W_1\}}{N-1} \right) = \frac{\int_{-\pi}^{\pi} \cos(\omega) dG(\omega)}{\int_{-\pi}^{\pi} dG(\omega)}. \quad (3)$$

Для гармонічного сигналу з частотою ω_0 спектральна щільність потужності, у разі $\omega \in [0, \pi]$, повинна задовольняти умові [2]

$$G(\omega) = \frac{A^2}{4} [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)], \quad (4)$$

де A — амплітуда гармоніки; $\delta(\omega)$ — дельта-функція.

Підставляючи (3) в (4), отримаємо рівняння

$$\frac{\pi E\{W_1\}}{N-1} = \omega_0, \quad (5)$$

тобто частота гармоніки збігається з нормованим значенням математичного сподівання кількості перетинань нульового рівня сигналом. Для дискретного сигналу і лінійної частоти, рівняння (5) має вигляд

$$\frac{E\{W_1\} f_d}{N-1} = f_0,$$

де f_d — частота дискретизації; f_0 — частота гармоніки.

У загальнішому випадку, коли в сигналі присутня деяка домінуюча частота на тлі завад, рівність (5) не виконується, але можна припустити, що нормоване значення кількості перетинань нульового рівня сигналом буде прямувати до домінуючої частоти. Якщо не враховувати операцію математичного сподівання, то можна стверджувати, що у випадку коли деяка частота стає домінуючою, величина $W_1 f_d / 2(N-1)$ буде мати значення, що відповідає або є близьким до цієї час-

тоти. Таким чином, якщо в спектрі сигналу присутня домінуюча частота, то її можна визначити оцінюванням кількості перетинань нульового рівня сигналом.

Цей підхід до виділення домінуючої частоти є ефективним, якщо в сигналі присутня одна гармоніка і відношення сигнал/шум велике. Коли в сигналі присутні кілька гармонік чи амплітуда гармонічної компоненти значно менша шуму, доцільно застосовувати фільтри, які дозволяють виділяти періодичні компоненти сигналу та оцінювати домінуючу частоту в різних частотних смугах.

Для обмеження частотного діапазону, в якому спостерігається частота основного тону, пропонується використовувати смуговий фільтр. Потім до виділеного частотного діапазону по чергово застосовуються процедури низькочастотної та високочастотної фільтрації з використанням рекурсивних цифрових фільтрів з симетричною імпульсною характеристикою, для яких різницеве рівняння має вигляд

$$y(n) = \sum_{i=0}^N h(i) s(n-i),$$

де $h(i)$ — коефіцієнти фільтра; N — порядок фільтра.

Для виділення сигналу в області нижніх частот можна застосувати операцію повторного додавання (фільтр нижніх частот), для якого перша сума запишеться у вигляді

$$\Delta s(n) = s(n) + s(n-1),$$

а k -та кінцева сума запишеться як:

$$\Delta^k s(n) = \sum_{j=0}^k C_k^j s(n-j).$$

Нехай W_K — кількість перетинів амплітудою сигналу нульового рівня для ряду $\Delta^{k-1} s(n)$. Через те, що оператор повторного додавання є фільтром нижніх частот, ряд $\Delta^k s(n)$ буде менш осцилюючим, ніж ряд $\Delta^{k-1} s(n)$, отже $W_{K+1} < W_K$.

Як фільтр нижніх частот використано кінцево-різницевий оператор ∇ , для якого перша різниця має такий вигляд

$$\nabla s(n) = s(n) - s(n-1),$$

а k -та кінцева різниця запишеться як

$$\nabla^k s(n) = \sum_{j=0}^k C_k^j (-1)^j s(n-j), \quad C_k^j = \frac{k!}{(k-j)! j!}.$$

Позначимо W_D — кількість перетинань нульового рівня сигналом для ряду $\nabla^{k-1} s(n)$.

Для отримання гнучкіших і точніших результатів для оцінювання властивостей сигналу варто одночасно використовувати величини W_D і W_K , послідовно застосовуючи до сигналу процедури фільтрації $\nabla^{k-1} \Delta^{j-1} s(n)$, підраховуючи кількість нульових перетинань.

Отже, для знаходження частоти основного тону за кількістю нульових перетинань можна використати формулу

$$\omega(i) = \frac{f_d}{2M} W(i, j), \tag{6}$$

$$W_{K_D}(i) = \begin{cases} W_{K_D}(i) + 1, & \text{sgn}[z_{K_D}(n)] \neq \text{sgn}[z_{K_D}(n-1)]; \\ W_{K_D}(i), & \text{sgn}[z_{K_D}(n)] = \text{sgn}[z_{K_D}(n-1)]; \end{cases}$$

$$W_{K_S}(j) = \begin{cases} W_{K_S}(j) + 1, & \text{sgn}[z_{K_S}(n)] \neq \text{sgn}[z_{K_S}(n-1)]; \\ W_{K_S}(j), & \text{sgn}[z_{K_S}(n)] = \text{sgn}[z_{K_S}(n-1)], \end{cases}$$

якщо

$$W(i, j) = \sum_{i, j=1}^M W_{K_D}(i) \cdot W_{K_S}(j),$$

де $\omega(i)$ — домінуюча частота; i — порядковий номер K_D в діапазоні $0 \leq K_D \leq K_{D_{\max}}$; j — порядковий номер K_S у діапазоні $0 \leq K_S \leq K_{S_{\max}}$; M — кількість відліків;

$$f_d \text{ — частота дискретизації; } W_{K_S}(j) = \begin{cases} W_{K_S}(j) + 1, & \text{sgn}[z_{K_S}(n)] \neq \text{sgn}[z_{K_S}(n-1)]; \\ W_{K_S}(j), & \text{sgn}[z_{K_S}(n)] = \text{sgn}[z_{K_S}(n-1)] \end{cases} \text{ — кількість}$$

нульових перетинань повторно-різницевої вибірки;

$$W_{K_D}(i) = \begin{cases} W_{K_D}(i) + 1, & \text{sgn}[z_{K_D}(n)] \neq \text{sgn}[z_{K_D}(n-1)]; \\ W_{K_D}(i), & \text{sgn}[z_{K_D}(n)] = \text{sgn}[z_{K_D}(n-1)] \end{cases} \text{ — кількість нульових перетинань вибі-$$

рки повторного додавання; $z_{K_D}(n) = \nabla^{K_D}(z_s(n)) = \nabla(\nabla^{K_D-1}(z_s)(n))$; $n = 1, \dots, M_D$;

$M_D = M_S - K_D$; $K_D = 0, \dots, K_{\max}$ — сформована повторно-різницева вибірка;

$z_s(n) = \Delta^{K_S} s(n) = \Delta(\Delta^{K_S-1} s(n))$, $n = 1, \dots, M_S$; $M_S = M - K_S$ — сформована вибірка повторного додавання значень $s(n)$.

Алгоритм виділення кореляційних функцій частоти основного тону диктора

На основі запропонованого методу розроблено алгоритм виділення кореляційних функцій частоти основного тону диктора на основі запропонованого методу складається із таких кроків:

1. Формуються вибірки повторного додавання з індексом K_S (реалізація процедури низькочастотної фільтрації мовного сигналу). Підраховується кількість перетинань нульового рівня сигналом.

2. Формуються повторно-різницеві вибірки з індексом K_D (реалізація процедури високочастотної фільтрації мовного сигналу). Підраховується кількість перетинань нульового рівня сигналом.

На основі (6), з урахуванням даних, отриманих на 1, 2 кроках алгоритму, будується матриця домінуючих частот розміром (4×4) , де стовпці визначаються порядком повторно-різницевої вибірки, а рядки — порядком вибірки повторного додавання.

Кроки 1—4 повторюються для кожної паролної фрази кожного з дикторів, які приймають участь в процесі ідентифікації і для отриманих функцій зміни частоти основного тону $x_{ki}(nT)$ в часі будується кореляційна функція за формулою

$$R_{x,y}(mT) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x_{ki}(nT) x_{k(i+1)}(nT + mT),$$

де $k = \{k_1, k_2, \dots, k_D\}$ — множина дикторів, які приймають участь в процесі ідентифікації,

$$i = \begin{cases} i1_{k1} & i2_{k1} & \dots & i\Pi_{k1} \\ i1_{k2} & i2_{k2} & \dots & i\Pi_{k2} \\ \dots & \dots & \dots & \dots \\ i1_{kD} & i2_{kD} & \dots & i\Pi_{kD} \end{cases} \text{ — множина функцій частоти основного тону } k_i \text{ диктора, де}$$

$i = 1, 2, \dots, D$ — кількість дикторів, що приймають участь в експерименті, а розмірність множини $i[D, \Pi]$, де Π — кількість отриманих функцій зміни частоти основного тону в часі для кожного з дикторів, що приймають участь в експерименті.

На рис. 1, як приклад роботи алгоритму, наведено кореляційні функції частоти основного тону для двох мовців, для однакової паролльної фрази.

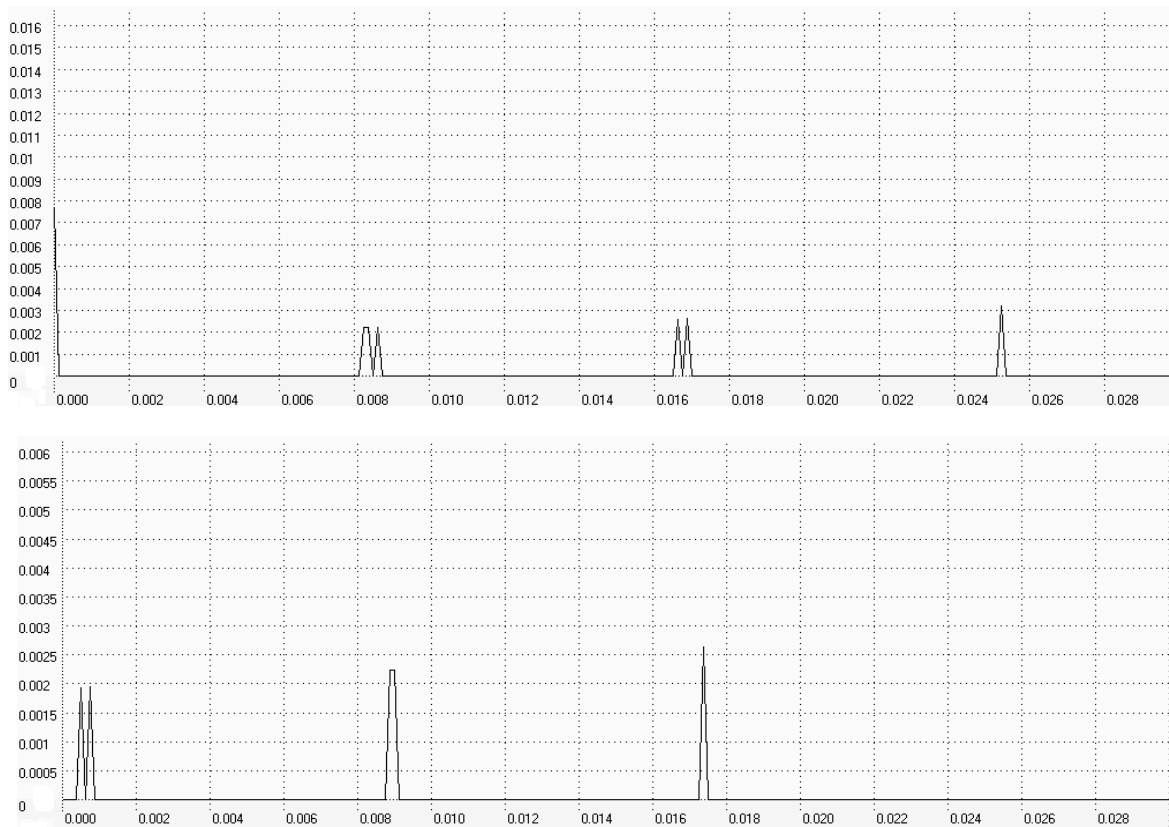


Рис. 1. Кореляційні функції частоти основного тону для двох мовців

Для цієї ознаки викликає дослідницький інтерес виявлення залежності достовірності розпізнавання голосів за запропонованою ознакою від рівня низькочастотних шумів присутніх в мовному сигналі, адже сучасні методи виділення частоти основного тону чутливі до такого типу завад. Отже, із записів паролльних фраз мовців виділялись частотні діапазони від 0, 20, 40, ..., 500 Гц, до яких додавалася шумова компонента, після чого записи паролльних фраз оброблялися за допомогою наведеного вище алгоритму та проводилося розпізнавання голосів з використанням нейромережевого класифікатора. Результати дослідження показані на рис. 2.

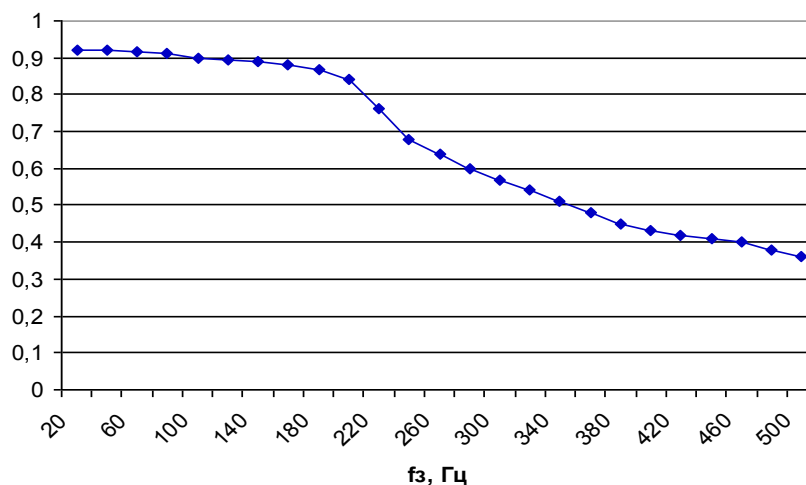


Рис. 2. Залежність достовірності розпізнавання голосів за кореляційними функціями частоти основного тону від рівня низькочастотних шумів

Тестування методу та аналіз результатів

Результати експерименту (див. рис. 2) показують, що ефективність частоти основного тону для розпізнавання голосів різко падає за присутності в мовному сигналі низькочастотних шумів, що обмежує область використання цієї ознаки, але за умови контролю рівня низькочастотних шумів запропонована ознака дозволяє досягати високого значення (94 %) достовірності розпізнавання голосів.

Висновки

Отже, авторами запропоновано метод виділення частоти основного тону на основі кореляції функцій перетинань нульового рівня сигналом за використання повторно-різницевих та повторно-додавальних вибірок на основі модифікованої математичної моделі слухової системи людини, який дозволяє покращити роздільність класів голосів мовців в просторі ознак та виділяти частоту основного тону за присутності в мовному сигналі шумів навколишнього середовища. Ефективність методу підтверджено емпірично, зокрема, проведено експеримент по розпізнаванню 20 осіб за індивідуальними особливостями їх голосів з використанням запропонованого методу, причому до запису паролльної фрази на відповідних частотах (див. рис. 2) додавалася низькочастотна завада. В результаті сумарна достовірність розпізнавання 20 осіб була вищою 90 % за наявності низькочастотної завади в діапазоні часто від 0 до 150 Гц (зауважимо, що частота основного тону осіб чоловічої статі спостерігається у частотному діапазоні 80...180 Гц.). Класичні методи виділення частоти основного тону не дозволили досягти достовірності розпізнавання більше 70 % та тому самому мовному матеріалі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Айфичер Э. С. Цифровая обработка сигналов: практический подход / Э. С. Айфичер, Б. У. Джервис. ; пер. с англ. — 2-е изд. — М. : Изд. дом «Вильямс», 2004. — 992 с.
2. Рамишвили Г. С. Автоматическое опознавание говорящего по голосу / Г. С. Рамишвили. — М. : Радио и связь, 1981. — 224 с.
3. Рабинер Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Шафер. ; пер. с англ. — М. : Радио и связь, 1981. — 496 с.
4. Сергиенко А. Б. Цифровая обработка сигналов : учеб. пос. / А. Б. Сергиенко. — 2-е изд. — СПб. : Питер, 2006. — 752 с.

Рекомендована кафедрою автоматики та інформаційно-вимірювальної техніки

Стаття надійшла до редакції 30.03.11
Рекомендована до друку 18.04.11

Биков Микола Максимович — професор, **Ковтун В'ячеслав Васильович** — доцент.

Кафедра комп'ютерних систем управління, Вінницький національний технічний університет;

Раймі Абдурахман — викладач кафедри прикладної математики і інформатики.

Університет ім. Шейха Анта Діюп, Дакар, Сенегал