УДК 004.934.1’1

О.А. КРИВОДУБСКИЙ, Е.Е. ФЕДОРОВ

МОДЕЛИРОВАНИЕ ОСОБЕННОСТЕЙ РЕЧИ ДИКТОРА

Abstract: *For creation of a system of verification of the speaker in the article the technique of verification on the basis of unvoiced fricative sounds was offered which uses author's methods executing a generalized classification of sounds by means of discrete and continuous wavelet-transformation. For unvoiced fricative sounds chosen by these methods, the quantitative analysis of systems of indications based on a linear prediction, normalized quantity of pulses of equal length and mel-frequency cepstral coefficients (MFCC) is conducted. The obtained indications are used in a method of verification based on the algorithm DTW.*
Key words: *verification of the speaker, generalized classification of sounds, wavelet-transformation, linear prediction, normalized quantity of pulses of equal length, mel-frequency cepstral coefficients, algorithm DTW.*

Анотація: *Для створення системи верифікації диктора у статті була запропонована методика верифікації на підставі шиплячих звуків, що використовує авторські методи, які здійснюють узагальнену класифікацію звуків за допомогою дискретного та безперервного вейвлет-перетворення. Для шиплячих звуків, виділених цими методами, проведений кількісний аналіз систем ознак, заснованих на лінійному пророкуванні, нормованої кількості імпульсів рівної довжини та мел-частотних кепстральних коефіцієнтах (MFCC). Отримані ознаки використовуються в методі верифікації, заснованому на алгоритмі DTW.*
Ключові слова: *верифікація диктора, узагальнена класифікація звуків, вейвлет-перетворення, лінійне пророкування, нормована кількість імпульсів рівної довжини, мел-частотні кепстральні коефіцієнти, алгоритм DTW.*

Аннотация: *Для создания системы верификации диктора в статье была предложена методика верификации на основе шипящих звуков, которая использует авторские методы, осуществляющие обобщенную классификацию звуков посредством дискретного и непрерывного вейвлет-преобразования. Для шипящих звуков, выделенных этими методами, проведен количественный анализ систем признаков, основанных на линейном предсказании, нормированном количестве импульсов равной длины и мел-частотных кепстральных коэффициентах (MFCC). Полученные признаки используются в методе верификации, основанном на алгоритме DTW.*
Ключевые слова: *верификация диктора, обобщенная классификация звуков, вейвлет-преобразование, линейное предсказание, нормированное количество импульсов равной длины, мел-частотные кепстральные коэффициенты, алгоритм DTW.*

1. Введение

Постановка проблемы. В настоящее время актуальной является разработка систем, предназначенных для верификации диктора. Эти системы имеют широкую область применения: криминалистика, криптография, охранные системы и др. При разработке подобных систем важную роль играет выбор системы признаков и методов верификации, использующих данные признаки.

Анализ исследований. В работах [1–2] приведены системы верификации, дающие в большинстве случаев вероятность распознавания ниже 90%. Верификация обычно проводится на целых словах, получение которых не всегда возможно.

Цель и задачи исследования. Для повышения надежности верификации необходимо разработать методику верификации диктора на основе шипящих звуков.

2. Методика верификации диктора и методы классификации звуков

В статье рассматриваются:

– методика верификации дикторов;

– авторские методы ФЕОК-ДВП и ФЕОК-НВП, осуществляющие обобщенную классификацию звуков посредством дискретного и непрерывного вейвлет-преобразования;

– проведенный количественный анализ для выбора системы признаков шипящих звуков.

Методика верификации диктора на основе шипящих звуков включает три этапа.

На первом этапе с целью выделения шипящих звуков осуществляется обобщенная классификация звуков посредством авторских методов ФЕОК-ДВП и ФЕОК-НВП.

Авторский метод ФЕОК-ДВП

Метод ФЕОК-ДВП осуществляет обобщенную классификацию звуков посредством дискретного вейвлет-преобразования и заключается в следующем.

Производится декомпозиция сигнала $x(n)$ на P уровней с вычислением высоко- (d_{im}) и низкочастотных (c_{im}) составляющих (субполос) посредством свертки на текущем i-том уровне ($i \in \overline{1,P}$) сигнала с полосовыми фильтрами с коэффициентами g_n, h_n [3–5]:

$$d_{im} = 2^{1/2} \sum_{n=0}^{N/2^{i-1}-1} c_{i-1,n} g_{n+2m}\ ,\ c_{im} = 2^{1/2} \sum_{n=0}^{N/2^{i-1}-1} c_{i-1,n} h_{n+2m}\ ,$$

где $c_{0n} = s(n)$, $m \in \overline{0, N/2^{i-1}-1}$.

Затем сигнал разбивается на фреймы длиной ΔN. Для s-го фрейма на i-м уровне разложения вычисляется энергия

$$DE_{si} = \sum_{m=(s-1)\Delta N}^{s\Delta N} d_{im}^2\ ,\ CE_{sP} = \sum_{m=(s-1)\Delta N}^{s\Delta N} c_{Pm}^2.$$

Для s-го фрейма на i-м уровне разложения вычисляется мера контрастности

$$Contr_{si} = DE_{si} \Big/ \sum_{j=1}^{i} DE_{sj}\ .$$

Тип каждого s-го фрейма определяется следующим образом:

$$n = \begin{cases} \alpha_{1\gamma} < Contr_{s\gamma} < \alpha_{2\gamma}, & шум \\ \alpha_{1\gamma} \geq Contr_{s\gamma}, & шипящий \\ \alpha_{2\gamma} \leq Contr_{s\gamma}, & тональный, \end{cases}$$

где $\alpha_{1\gamma}$, $\alpha_{2\gamma}$ – пороги, которые автоматически вычисляются в подсистеме адаптации и представляют собой минимальное и максимальное значения контрастности шума на уровне γ.

Если $n = 2$, то определяются левая l_1 и правая l_2 границы шипящих звуков.

Достоинством метода ФЕОК-ДВП является малая вычислительная сложность. Недостатком – сложность проведения классификации в зашумленном сигнале.

Авторский метод ФЕОК-НВП

Метод ФЕОК-НВП осуществляет обобщенную классификацию звуков посредством аппроксимированного непрерывного вейвлет-преобразования и заключается в следующем.

Производится декомпозиция сигнала $x(n)$ на P уровней с вычислением вейвлет-коэффициентов (d_{im}) [4–5].

$$d_{il} = \sum_{n=0}^{N-1} x(n) \psi_{il}(n) \Delta t\ ,\ l \in \overline{0, N-1}\ ,\ i \in \overline{1,P},$$

где Δt – величина, обратная частоте дискретизации; $\psi_{il}(n) = a_0^{-i/2}\psi\left(a_0^{-i}n - b_0 l\right)$, – вейвлет, $a_0 > 1$, $b \neq 0$.

Затем сигнал разбивается на фреймы длиной ΔN. Для s-го фрейма на i-м уровне разложения вычисляется энергия

$$DE_{si} = \sum_{m=(s-1)\Delta N}^{s\Delta N} d_{im}^2.$$

Для s-го фрейма по всем уровням разложения вычисляется мера контрастности

$$Contr_{si} = DE_{si} \Big/ \sum_{j=1}^{i} DE_{sj},\ i \in \overline{1,P}.$$

Тип каждого k-го фрейма определяется следующим образом:

$$n = \arg\max_t \left\{ \sum_{\gamma \in \Theta} \delta_{tk\gamma} \right\},\ \Theta \subset \{1,...,P\},$$

$$\delta_{1k\gamma} = \begin{cases} 1, & \alpha_{1\gamma} < Contr_{k\gamma} < \alpha_{2\gamma} \\ 0, & иначе \end{cases},$$

$$\delta_{2k\gamma} = \begin{cases} 1, & \alpha_{1\gamma} > Contr_{k\gamma} \\ 0, & иначе \end{cases},$$

$$\delta_{3k\gamma} = \begin{cases} 1, & Contr_{k\gamma} > \alpha_{2\gamma} \\ 0, & иначе \end{cases},$$

где $\alpha_{1\gamma}$, $\alpha_{2\gamma}$ – пороги, которые автоматически вычисляются в подсистеме адаптации и представляют собой минимальное и максимальное значения контрастности шума на уровне γ, Θ – подмножество уровней, получаемое в результате численного исследования.

$$n = \begin{cases} 1, & шум \\ 2, & шипящий \\ 3, & тональный \end{cases}.$$

Если $n = 2$, то определяются левая l_1 и правая l_2 границы шипящих звуков.

Достоинством метода ФЕОК-НВП является возможность проведения классификации в зашумленном сигнале. Недостатком – большая вычислительная сложность.

На втором этапе вычисляются значения признаков шипящих звуков. Сигнал $s(m)$ предварительно разбивался на фреймы (участки равной длины). В качестве признаков использовались:

1. Нормированная автокорреляция [6–7], вычисляемая на n-м фрейме сигнала $s_n(m)$,

$$\hat{s}_n(m) = s_n(m)w(m),$$

где $w(n) = 0{,}54 + 0{,}46\cos\dfrac{2\pi m}{\Delta N}$ – окно Хемминга,

$$R_n(i) = \sum_{m=0}^{N-1-i} \hat{s}_n(m)\hat{s}_n(m+i),\ n \in \overline{1,p},$$

где p – порядок линейного предсказателя,

$$\left\|R_n(i)\right\| = \frac{R_n(i)}{R_n(0)},$$

$$Q_{in}^1 = (\left\|R_n(1)\right\|,...,\left\|R_n(p)\right\|),\ i \in \overline{1,\eta(\mathbf{Q}^1)},\ n \in \overline{1,L},$$

где $\eta(\mathbf{Q}^1)$ – количество эталонов шипящих фонем, L – количество фреймов.

2. Коэффициенты линейного предсказания, вычисляемые с помощью алгоритма Дарбина [6–7] на n-м фрейме сигнала:

$$E_n^{(0)} := R_n(0),$$

$$k_{ni} := \left[R_n(i) - \sum_{j=1}^{i-1} \alpha_{nj}^{(i-1)} R_n(i-j)\right] \Big/ E_n^{(i-1)}, 1 \le i \le p,$$

$$\alpha_{ni}^{(i)} := k_{ni},$$

$$\alpha_{nj}^{(i)} := \alpha_{nj}^{(i-1)} - k_{ni}\alpha_{n,i-j}^{(i-1)}, 1 \le j \le i-1,$$

$$E_n^{(i)} := (1-k_{ni}^2)E_n^{(i-1)},$$

$$a_{nj} := \alpha_{nj}^{(p)}, 1 \le j \le p,$$

где $R_n(i)$ – автокорреляционная функция, $\alpha_{nj}^{(i)}$ – j-й коэффициент линейного предсказателя порядка i, k_{ni} – i-й коэффициент отражения, $E_n^{(i)}$- среднеквадратичная погрешность предсказания для линейного предсказателя порядка i.

$$Q_{in}^2 = (a_{n1},...,a_{np}),\ i \in \overline{1,\eta(\mathbf{Q}^2)},\ n \in \overline{1,L}.$$

3. Коэффициенты отражения КЛП (или PARCOR-коэффициенты) [6–7], определяемые по алгоритму Дарбина на n-м фрейме сигнала:

$$Q_{in}^3 = (k_{n0},...,k_{np}),\ i \in \overline{1,\eta(\mathbf{Q}^3)},\ n \in \overline{1,L}.$$

4. Нормированная автокорреляция КЛП [6–7], получаемая по коэффициентам линейного предсказания:

$$r_n(0) = \sum_{j=0}^{p} a_{nj}^2,\ r_n(k) = 2\sum_{j=0}^{p-k} a_{nj}a_{n,j+k},$$

$$\left\|r_n(i)\right\| = \frac{r_n(i)}{r_n(0)},$$

$$Q_{in}^4 = (\left\|r_n(1)\right\|,...,\left\|r_n(p)\right\|),\ i \in \overline{1,\eta(\mathbf{Q}^4)},\ n \in \overline{1,L}.$$

5. Кепстр импульсной характеристики системы линейного предсказания [6–7], вычисляемый по коэффициентам линейного предсказания:

$$\hat{h}_n(0) = a_{n0} = 1,\ \hat{h}_n(j) = a_{nj} + \sum_{k=1}^{j-1} \frac{k}{j} \cdot \hat{h}_n(k) a_{n,j-k},\ n \in \overline{1,p},$$

$$Q_{in}^5 = (\hat{h}_n(1),..., \hat{h}_n(p)),\ i \in \overline{1,\eta(\mathbf{Q}^5)},\ n \in \overline{1,L}.$$

6. Площади поперечных сечений кусочно-постоянной акустической трубы [6–7], содержащей $(p+1)$ цилиндрическую секцию фиксированной длины, вычисляемые с помощью коэффициентов отражения:

$$A_{n1} = 1,\ A_{n,i+1} = \frac{1-k_{ni}}{1+k_{ni}} A_{ni},\ i \in \overline{2,p+1},$$

$$Q_{in}^6 = (A_{n2},..., A_{n,p+1}),\ i \in \overline{1,\eta(\mathbf{Q}^6)},\ n \in \overline{1,L}.$$

7. Нормированный энергетический спектр КЛП [1]:

$$W_n(k) = \frac{R_n(0) - \sum_{k=1}^{p} \alpha_{nk} R_n(k)}{r_n(0) - \sum_{s=1}^{p} r_n(s) \cos\left(\frac{2\pi}{N} sk\right)},\ k \in \overline{0, N/2-1},$$

$$\left\| W_n(k) \right\| = \frac{W_n(k)}{\sum_{i=0}^{N/2-1} W_n(i)},\ 0 \le k \le N/2-1,$$

$$Q_{in}^7 = (\left\| W_n(0) \right\|,..., \left\| W_n(N/2-1) \right\|),\ i \in \overline{1,\eta(\mathbf{Q}^7)},\ n \in \overline{1,L}.$$

8. Нормированный энергетический спектр, вычисленный на основе энергетического спектра ДПФ [1]:

$$\hat{s}_n(m) = s_n(m) w(m),$$

где $w(n) = 0{,}54 + 0{,}46 \cos\frac{2\pi n}{\Delta N}$ – окно Хемминга.

$$S(k) = \sum_{n=0}^{\Delta N-1} \hat{s}(n)\ e^{-j\frac{2\pi nk}{\Delta N}},\ 0 \le k \le \Delta N/2-1,$$

$$\left\| S_n(k) \right\| = \frac{S_n(k)}{\sum_{i=0}^{N/2-1} S_n(i)},\ 0 \le k \le N/2-1,$$

$$Q_{in}^8 = (\left\| S_n(0) \right\|,..., \left\| S_n(N/2-1) \right\|),\ i \in \overline{1,\eta(\mathbf{Q}^8)},\ n \in \overline{1,L}.$$

9. Нормированное количество импульсов равной длины.

Для каждого n-го фрейма вычисляется d_{nz} – количество импульсов длины z [8], $z \in \overline{1,len}$, которое далее нормируется

$$\| d_{nz} \| = \frac{d_{nz}}{\sum_{s=1}^{len} d_{ns}},$$

$$Q_{in}^{9} = (\|d_{n1}\|,...,\|d_{n,len}\|),\ i \in \overline{1,\eta(\mathbf{Q}^{9})},\ n \in \overline{1,L}.$$

10. Мел-частотные кепстральные коэффициенты (MFCC) [9], вычисляемые с использованием обратного дискретного косинусного преобразования на n-м фрейме сигнала:

$$MFCC_{nk} = \sum_{l=1}^{P} E_{nl} \cos\left(k(l-0.5)\pi / P\right),\ k \in \overline{1,P},$$

где $E_{nl} = \lg\left(\sum_{k=k1_i}^{k2_i} \left(\widehat{S}_n(k)\right)^2 w(k-(k1_l + \Delta K_l/2)) \right)$ – логарифмированная энергия l-той мел-частотной полосы. Для ее вычисления использовалась треугольная оконная функция Бартлета:

$$w(k) = \begin{cases} 0, & k < -\Delta K_l/2 \\ 1+\dfrac{2k}{\Delta K_l}, & -\Delta K_l/2 \le k \le 0 \\ 1-\dfrac{2k}{\Delta K_l}, & 0 \le k \le \Delta K_l/2 \\ 0, & k > \Delta K_l/2 \end{cases},$$

где $\Delta K_l = k2_l - k1_l$ – четное число, $k1_l, k2_l$ – границы частотных диапазонов l-той полосы.

$$Q_{in}^{10} = (MFCC_{n1},..., MFCC_{nP}),\ i \in \overline{1,\eta(\mathbf{Q}^{10})},\ n \in \overline{1,L}.$$

На основе каждой s-й системы признаков формируются эталоны шипящих звуков.

$$Q_i^s = ((Q_{i11}^s,..., Q_{i1k}^s,..., Q_{i1K}^s),..., (Q_{ij1}^s,..., Q_{ijk}^s,..., Q_{ijK}^s),..., (Q_{iL1}^s,..., Q_{iLk}^s,..., Q_{iLK}^s)),\ s \in \overline{1,10},\ i \in \overline{1,\eta(\mathbf{Q}^s)},$$

где $\eta(\mathbf{Q}^s)$ – количество эталонов шипящих фонем для s-й системы признаков;

L – количество участков равной длины (фреймов), на которые разбивается сигнал;

K – количество признаков, описывающих один фрейм.

На третьем этапе верификации диктора, используя полученные векторы признаков шипящих звуков, производится собственно верификация. При этом используется алгоритм DTW [10].

В основе алгоритма DTW лежит рекуррентная формула

$$C_{i11}^s = D_{i11}^s,\ C_{imj}^s = D_{imj}^s + \min(C_{i,m-1,j}, C_{i,m,j-1}, C_{i,m-1,j-1}),\ m \in \overline{1,L},\ j \in \overline{1,L},$$

где C_{imj}^s – расстояние между левыми частями фонемы (фреймы от 1 до m) и i-м эталоном (фреймы от 1 до j);

D_{imj}^s – расстояние между m-м фреймом фонемы и j-м фреймом i-го эталона.

В качестве D_{imj}^s выбрана евклидова метрика

$$D_{imj}^s = \sqrt{\sum_{k=1}^{K} (\widehat{Q}_{mk}^s - Q_{ijk}^s)^2},$$

где $\widehat{Q}^s_{mk}$ – k-й признак m-го фрейма фонемы;

Q^s_{ijk} – k-й признак j-го фрейма i-го эталона;

s – тип системы признаков.

Количественная оценка сопоставления шипящего звука верифицируемого диктора $\widehat{Q}^s$ с эталонами Q^s_i для s-й системы признаков вычислялась в соответствии с функционалом ошибки

$$\Im = \min_i C^s_{iLL} \to \min .$$

Результат верификации диктора определялся в соответствии с правилом

$$result = \begin{cases} "свой", & n \in N_1 \\ "чужой", & n \in N_2 \end{cases}, \; n = \arg\min_i C^s_{iLL} .$$

Для выбора системы признаков было проведено численное исследование, в котором участвовало 100 дикторов (50 «своих», 50 «чужих»): мужчины и женщины с разными голосовыми данными. В табл.1 приведены результаты верификации диктора по шипящему звуку («ш»). Численное исследование позволяет сделать вывод, что наиболее перспективными являются признаки MFCC и коэффициенты отражения КЛП.

Таблица 1. Результаты численного исследования систем признаков

Система признаков	**Вероятность верификации**
Нормированная автокорреляция	0,44
Коэффициенты КЛП	0,78
Коэффициенты отражения КЛП	0,96
Нормированная автокорреляция КЛП	0,78
Кепстр КЛП	0,68
Площади поперечных сечений акустической трубы КЛП	0,42
Нормированный энергетический спектр КЛП	0,4
Нормированный энергетический спектр ДПФ	0,32
Нормированное количество импульсов равной длины	0,56
MFCC	0,96

3. Выводы

Новизна. В статье предложены методика верификации диктора на основе шипящих звуков, авторские методы ФЕОК-ДВП и ФЕОК-НВП, осуществляющие обобщенную классификацию звуков в дискретном сигнале с целью выделения шипящих. Было проведено численное исследование систем признаков шипящих звуков, при этом в качестве метода распознавания был выбран алгоритм DTW. Преимуществом предлагаемого подхода является то, что для верификации диктора вместо всего слова достаточно использовать входящий в него шипящий звук, что расширяет область применения системы верификации.

Практическое значение. Основные положения работы были использованы при разработке системы верификации диктора, которая может использоваться в криминалистике и охранных системах.

СПИСОК ЛИТЕРАТУРЫ

1. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь,1981. – 496 с.
2. Атал Б.С. Автоматическое опознавание дикторов по голосам // ТИИЭР. – 1976. – Т. 64, № 4. – С. 48–66.
3. Чуи К. Введение в вэйвлеты. – М.: Мир, 2001. – 412 с.
4. Малла С. Вэйвлеты в обработке сигналов. – М.: Мир, 2005. – 671 с.
5. Добеши И. Десять лекций по вейвлетам. – М.: РХД, 2004. – 464 с.
6. Rabiner L.R., Jang B.H. Fundamentals of speech recognition // New Jersey: Prentice Hall PTR, Englewood Cliffs, 1993. – P. 507.
7. Маркел Д.Д., Грэй А.Х. Линейное предсказание речи. – М.: Связь,1980. – 308 с.
8. Молдокулова Н.В., Трунин-Донской В.Н. Лингво-акустические проблемы создания системы распознавания слитной речи на ЭВМ / Под ред. Ю.Й. Журавлева; АН КиргССР, Вычислительный центр АН СССР. – Фрунзе: Илим, 1989. – 136 с.
9. Davis S.B., Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences // IEEE Trans. on Acoustic, Speech and Signal Proc. – 1980. – Vol. 28, N 4. – P. 357–366.
10. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наукова думка, 1987. – 261 с.

Стаття надійшла до редакції 30.08.2007