УДК 621.391
DOI: 10.15587/2313-8416.2019.172408

АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ПОЛА ДИКТОРА НА ОСНОВЕ РАСПРЕДЕЛЕНИЯ КОШИ В ОКТАВНОЙ ПОЛОСЕ ЧАСТОТ

© С. В. Омельченко

В работе получены алгоритмы определения пола диктора на основе использования распределения Коши в октавной полосе частот со среднегеометрической частотой 125 Гц. Построены классификаторы на основе максимума логарифма функции правдоподобия. Рассмотрен алгоритм определения пола диктора, где учитывается не только логарифм распределения Коши в октавной полосе частот, но и оценки среднего значения частот формант и частот антиформант. Проведены исследования вероятности правильного распознавания алгоритмов определения пола диктора
Ключевые слова: *распределение Коши, частоты формант, частоты антиформант, моментные функции, гендерное распознавание*

1. Введение

Развитие методов искусственного интеллекта и диалоговых систем взаимодействия человека с машиной ставит новые задачи дальнейшего поиска новых эффективных методов определения гендерной принадлежности людей по их голосу. При использовании для настройки параметров системы распознавания речи результатов определения гендерной принадлежности позволит повысить качество работы систем распознавания речи.

Поэтому актуальным является исследование методов определения гендерной принадлежности людей по их голосу.

2. Литературный обзор

Для решения задачи распознавания пола по речи человека известно множество подходов к выбору признаков и правил принятия решений.

В работах [1, 2] в качестве признаков при определении пола дикторов применяются совместное использование оценок частоты основного тона и кепстральных признаков.

В работе [3] предложено использовать классификатор на основе обобщённого метода моментов (англ. GMM — Generalized Method of Moments). Точность определения пола для тестовых произнесенных слов на чешских и словацких языках была около 90 %.

В работе [4] для гендерного распознавания предложено использовать средние значения, коэффициенты асимметрии и эксцесса для частот основного тона и частот формант.

Известен подход [5], где использован многослойный перцептрон глубокой модели обучения с использованием акустических свойств голоса и речи для определения пола диктора. Для своих экспериментов они использовали набор данных из 3168 образцов человеческого голоса. При использовании такой модели классификации удалось достичь точности 96,74 %.

Алгоритмы гендерного распознавания использующие в качестве признаков частоту основного тона и Мел-кепстральные коэффициенты (MFCC) рассмотрены в [6]. Правило принятия решений построено на основе линейной и логистической регрессии. Для такого алгоритма точность правильного распознавания составила 95 %.

Исследования по совместному использованию в качестве признаков оценок частот формант и MFCC приведены в работе [7]. При этом точность правильного распознавания равнялась 94 %.

В роботе [8] использованы в качестве признаков MFCC, а решающие правила построены на основе полигауссовских смесей распределений. При этом система обеспечивает точность 92 % правильного распознавания.

Алгоритмы гендерного распознавания, использующие в качестве признаков коэффициенты линейного предсказания, коэффициенты отражения, представлены в литературе [9, 10]. Здесь используют для построения решающих правил скрытые марковские процессы и полигауссовские распределения.

Поэтому всем алгоритмам гендерного распознавания по звуковым сигналам присущи такие недостатки, как сложность реализации алгоритмов и не достаточное качество распознавания.

Из проведенного литературного анализа можно сделать вывод, что перспективным для решения поставленной задачи является поиск новых сочетаний признаков, алгоритмов принятия решений о поле человека.

3. Цель и задачи исследования

Целью работы является синтез новых эффективных методов определения пола людей по их речи.

Для достижения цели были поставлены следующие задачи:

1. Произвести выбор новых комбинаций информативных признаков, позволяющих осуществить разделение полов людей по их речи.

2. Выполнить синтез новых правил принятия решений о поле диктора.

3. Выполнить экспериментальную проверку предложенного метода определения пола на реальных звуковых сигналах и оценить вероятность их правильного распознавания.

4. Математическая постановка задачи гендерного распознавания

На вход системы определения пола поступает последовательность цифровых отсчетов речевого сигнала, введенного с микрофона через звуковую карту в персональный компьютер. Дискретизация речевого сигнала может быть выбрана 8 кГц, что соответствует полосе 4 кГц, как правило, используемой для передачи речи по телефонной связи.

Нужно выполнить синтез методов распознавания пола, которые по предъявленным реализациям в виде отсчетов речевых сигналов, выносили бы решения с максимальной средней вероятностью правильного распознавания.

Для решения задачи распознавания пола необходимо выполнить обнаружение речевой информации и найти совокупности временных границ начала и конца слова [11]. Это позволяет исключить из гендерного распознавания совокупности отсчетов, соответствующих шумам или помехам.

5. Алгоритмы гендерного распознавания по речевым сигналам

Выделенные, после сегментации речи, совокупности отсчетов, соответствующие речи диктора, разбиваются на одинаковые блоки в диапазоне 256–512 отсчетов.

Цифровая фильтрация может быть выполнена в виде:

$$\hat{x}(t)=\mathrm{Re}\left((N)^{-\frac{1}{2}}\sum_{m=0}^{N-1}C(m)H_{кор}(m)\exp\left(i\left(\frac{2\pi t}{N}\right)m\right)\right),$$
$$C(m)=(2N)^{-\frac{1}{2}}\sum_{\tau=0}^{2N-1}y_{\tau}^{j}\exp\left(-i\left(\frac{2\pi\tau m}{2N}\right)\right), \quad (1)$$

где $y_i^j=\begin{cases}s_i^j, & i=0,1,...,(N-1),\\ 0, & i=N,(N+1),...,(2N-1)\end{cases}$ – отсчеты речевого сигнала; $H_{кор}(m)$ – передаточная характеристика цифрового фильтра; *Re* – оператор взятия вещественной части сигнала; C(m) – спектр входного сигнала; s_i^j - i -й отсчет j -го блока входного сигнала; N – количество отсчетов в блоке.

На рис. 1 видно наличие существенных различий оценок плотностей вероятностей речевых сигналов в октавной полосе частот со среднегеометрической частотой 125 Гц для мужчин и женщин.

Для плотностей вероятностей для мужских и женских речевых сигналов используются распределения Коши, которые различаются параметрами масштаба.

Распределение Коши занимает особое место среди многочисленных функций распределения непрерывных случайных величин, известных в теории вероятностей.

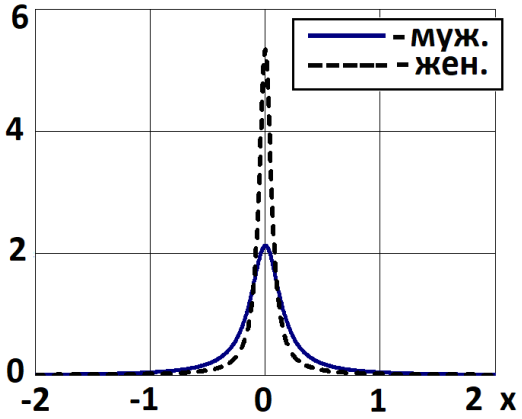

Рис. 1. Плотности вероятностей для мужских и женских речевых сигналов

Плотность распределения Коши имеет вид:

$$f(x)=\frac{\lambda}{\pi\cdot\left(\lambda^2+(x-\mu)^2\right)}, \quad (2)$$

где параметры λ и μ называют параметрами масштаба и положения, соответственно. Параметр положения μ совпадает с модой и медианой распределения, параметр масштаба λ совпадает со срединным отклонением.

Распределение данного вида называют двухпараметрическим, приняв $\mu=0$ получаем однопараметрическое распределение.

Логарифм функции правдоподобия примет вид:

$$\ln N_i=\sum_{j=1}^{n}\ln\left(\frac{\lambda_i}{\pi\cdot\left(\lambda_i^2+(x_j-\mu_i)^2\right)}\right). \quad (3)$$

где λ_i – параметр масштаба для i -го отсчета; μ_i – параметр положения для i -го отсчета; n – количество отсчетов.

Зависимость логарифма функции правдоподобия мужского голоса от параметра масштаба приведена на рис. 2, а женского голоса – на рис. 3.

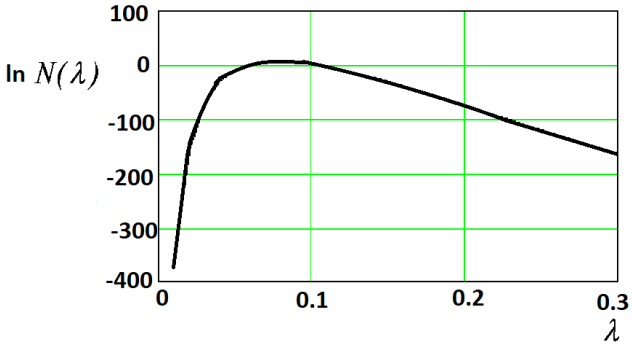

Рис. 2. Зависимость логарифма функции правдоподобия мужского голоса от параметра масштаба

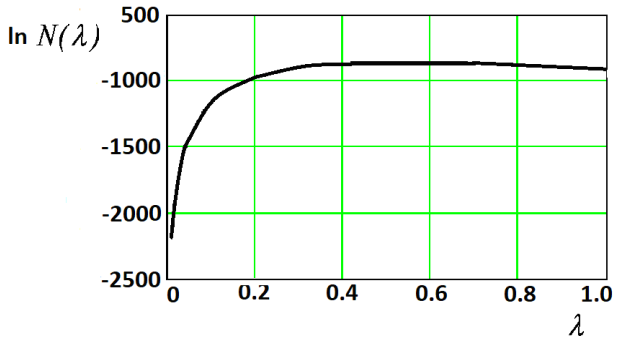

Рис. 3. Зависимость логарифма функции правдоподобия женского голоса от параметра масштаба

Плотности распределения Коши для указанного параметра мужского голоса для масштаба $\lambda = 0,08$ приведены на рис. 4, а женского голоса с параметром масштаба $\lambda = 0,4$ – на рис. 5.

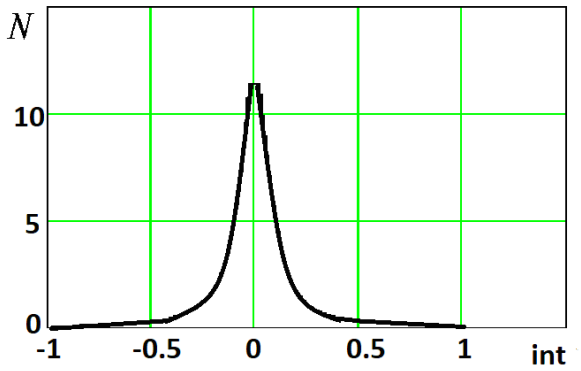

Рис. 4. Плотность распределения Коши для указанного параметра масштаба 0,08

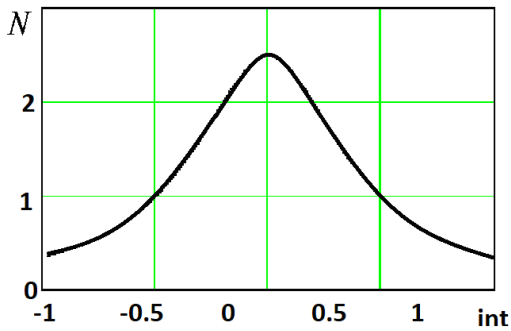

Рис. 5. Плотность распределения Коши для указанного параметра масштаба 0,4

Номер типа сигнала для распределения Коши в октавной полосе частот со среднегеометрической частотой 125 Гц находят в виде:

$$i = \arg\max\left(\ln N_u,\ u = \overline{1,2}\right), \qquad (4)$$

где $\arg\min\left(f(j),\ j = \overline{0,L}\right)$ – функция вычисления номера j, при котором функция $f(j)$ максимальна на множестве $j = \overline{0,L}$.

Для оценивания формантных и антиформантных частот необходимо оценить параметры авторегресии скользящего среднего (АРСС) речевого сигнала. Для оценивания параметров АРСС, как правило, применяются процедуры раздельного оценивания параметров авторегрессии (АР) и параметров скользящего-среднего (СС).

Сначала оцениваются коэффициенты АР $\hat{a}_u$, например, методом Левинсона, а затем их оценки используют для построения обратного фильтра, который будет применен к исходным данным.

Алгоритм оценивания ошибки предсказания описывается выражением:

$$y_t = x_t - \sum_{u=1}^{p} \hat{a}_u x_{t-u}, \qquad (5)$$

где $\hat{a}_u$ – оценки коэффициентов АР; p – порядок модели АР.

Последовательность остаточных ошибок на выходе этого фильтра должна характеризовать процесс скользящего среднего, к которому будет применена процедура оценивания СС-параметров.

Оценка нормированной корреляционной функции ошибки предсказания сигнала:

$$K_{yj} = \frac{1}{(T+1-j)\cdot(L2+1-L1)} \times \times \sum_{v=L1}^{L2} \sum_{i=0}^{T-j} \left(y_{i+j}^{(v)} \cdot y_i^{(v)}\right), \qquad (6)$$

где v – номер выборки; T – количество отсчетов в периоде наблюдения.

Коэффициенты регрессии $b = (b_0, b_1, ..., b_p)$ вычисляются как нормированная корреляционная функция ошибки предсказания в виде:

$$b_u = \frac{K_{yu}}{K_{y0}}. \qquad (7)$$

Передаточная характеристика фильтра, описывающая голосовой тракт человека может представлена в виде:

$$H\left(\frac{n2\pi}{T}\right) = \frac{\left|\sum_{k=1}^{p}\left(b_k \cdot e^{-\frac{ikn2\pi}{T}}\right)\right|}{\left|1-\sum_{k=0}^{p}\left(a_k \cdot e^{-\frac{ikn2\pi}{T}}\right)\right|}. \qquad (8)$$

Формантные частоты оцениваются в соответствии с выражением:

$$f_v = \left(\frac{F_o}{N}\right) \arg loc\max\left(H\left(\frac{n2\pi}{T}\right),\ n = \overline{0,M}\right), \qquad (9)$$

где $M = Z\left(\frac{N}{2}-1\right)$, Z(u) – функция округления числа u к целому; $\arg loc\max(x)$ – векторная функция, ставящая в соответствие последовательности отсчетов $x_1, x_2, ..., x_N$ упорядоченное множество, которое состоит из индексов $f_1, f_2, \ldots, f_L$, удовлетворяющих условию локального максимума:

$$x_{f_i} > x_{f_{i-1}}, x_{f_i} \le x_{f_{i+1}}.$$

Затем определяется оценка антиформантных частот в соответствии с выражением:

$$fa_v = \left(\frac{F_\partial}{N}\right)\arg loc\min\left(H\left(\frac{n2\pi}{T}\right),\ n = \overline{0,M}\right), \quad (10)$$

где $M = Z\left(\frac{N}{2}-1\right)$, Z(u) – функция округления числа u к целому; $\arg loc\min(x)$ – векторная функция, ставящая в соответствие последовательности отсчетов $x_1, x_2, ..., x_N$ упорядоченное множество, которое состоит из индексов $f_1, f_2, \ldots, f_L$, удовлетворяющих условию локального минимума:

$$x_{f_i} < x_{f_{i-1}}, x_{f_i} \ge x_{f_{i+1}}.$$

Среднее значение частот формант и частот антиформант оценивают в соответствии с выражениями:

$$mf = \frac{1}{N}\sum_{k=1}^{N} f_k,$$

$$mf\text{a} = \frac{1}{N}\sum_{k=1}^{N} f_{a,k}.$$

где f_k – оценка форманты для k-ой выборки; $f_{a,k}$ – оценка антиформанты для k-ой выборки; N – количество выборок.

Используя центральные моменты μ_k порядка k вычисляем коэффициент асимметрии:

$$As = \frac{\mu_3}{\sigma^3},$$

где μ_3 – центральный момент третьего порядка, σ – среднеквадратическое отклонение.

Эксцесс вычисляется по формуле:

$$Es = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 – центральный момент четвертого порядка, σ – среднеквадратическое отклонение.

Решение о гендерной принадлежности выносится в соответствии с выражением:

$$i = \arg\max\left(\begin{array}{l} k_0 \cdot \ln N_u + \sum_{j=1}^{4} k_{f,j,u} \cdot \left(mf_j - fgr_j\right) + \sum_{j=1}^{4} kAs_j, u \cdot \left(Asf_j - Asfgr_j\right) + \\ \sum_{j=1}^{4} kEf_{j,u} \cdot \left(Esf_j - Esfgr_j\right) + \sum_{j=1}^{4} k_{a,j,u} \cdot \left(mfa_j - mfagr_j\right) + \\ \sum^{4} kaAs_j, u \cdot \left(Asfa_j - Asfagr_j\right) + \sum^{4} kaEs_{j,u} \cdot \left(Esfa_j - Esfagr_j\right),\ u = \overline{1,2} \end{array}\right|$$

при этом mf_k, mf_k – усредненные по реализациям значения частот формант и антиформант; Asf, Asfa – коэффициент асимметрии формант и атиформант; $Esf_j, Esfa_j$ – оценки коэффициентов эксцесса для частот j-ой форманты и j-ой антиформанты; $fgr_j, fagr_j$, $Asfgr_j, Asfagr_j$, $Esfgr_j, Esfagr_j$ – граничные коэффициенты для двух полов.

6. Экспериментальные исследования

Проведены экспериментальные исследования предложенных алгоритмов определения пола диктора на персональном компьютере с использованием цифровых отсчетов, введенных с микрофона.

Для распознавания были использованы эталоны в виде совокупности отсчетов, соответствующих произнесенным цифрам от одного до десяти. Для оценивания вероятности правильного распознавания последовательно были предъявлены всего 200 эталонов, в формировании которых участвовали 10 мужчин и 10 женщин.

Вероятность правильного распознавания при равновероятном использовании мужских и женских эталонов имеет вид:

$$P_{pr} = \frac{m_m}{n_m}\cdot 0,5 + \frac{m_w}{n_w}\cdot 0,5,$$

где m_m, m_w – количество принятий правильных решений о мужчина и женщинах дикторах; n_m, n_w – общее количество испытаний при предъявлении эталонов для дикторов мужчин и женщин, соответственно.

При принятия решений в соответствии с формулой (4) при использовании логарифма функции правдоподобия и распределения Коши средняя вероятность правильного распознавания составила $P_{pr} = 0,91$, а для алгоритма совместного принятия решений в соответствии с формулой (11) – $P_{pr} = 0,96$. Проведенные экспериментальные исследований алгоритмов подтверждают возможности определения пола диктора с использованием совокупности выбранных информативных признаков и решающих правил.

7. Выводы

1. Выбраны новые классификационные признаки, включающие совместное использование различий параметров распределения Коши в октавной полосе частот для различных полов и оценок среднего значения частот формант и частот антиформант, их коэффициентов асимметрии и эксцесса.

2. Выполнен синтез нового правила принятия решений о поле диктора на основе совместного использования логарифма функции правдоподобия распределения Коши в октавной полосе частот, учете оценок средних значений, коэффициентов асимметрии и эксцесса частот формант и антиформант.

3. Методом статистических испытаний проведена проверка предложенных алгоритмов гендерного распознавания. Получена оценка вероятности правильного распознавания пола диктора – 0,9.

Литература

1. Калюжный А. Я., Семенов В. Ю. Метод идентификации пола диктора на основе моделирования акустических параметров голоса гауссовыми смесями // Акустичний вісник. 2009. Т. 12, № 2. С. 31–38.

2. Practical Considerations for Real-Time Implementation of Speech-Based Gender Detection / Scheme E., Castillo-Guerra E., Englehart K., Kizhanatham A. // Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006. P. 426–436. doi: http://doi.org/10.1007/11892755_44

3. Pribil J., Pribilova A., Matousek J. GMM-based speaker gender and age classification after voice conversion // 2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE). IEEE, 2016. P. 1–5. doi: http://doi.org/10.1109/splim.2016.7528391

4. Omelchenko S. Development of the method of automatic determination of the speaker gender on the basis of joint evaluation of frequency moments of basic tons and formant frequencies // Technology audit and production reserves. 2018. Vol. 3, Issue 2 (41). P. 29–33. doi: http://doi.org/10.15587/2312-8372.2018.134977

5. Buyukyilmaz M., Cibikdiken A. O. Voice Gender Recognition Using Deep Learning // Proceedings of 2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA2016). Atlantis Press, 2016. P. 409–411. doi: http://doi.org/10.2991/msota-16.2016.90

6. Levitan S. I., Mishra T., Bangalore S. Automatic identification of gender from speech // Proceeding of Speech Prosody. 2016. P. 84–88. doi: http://doi.org/10.21437/speechprosody.2016-18

7. Faek F. Objective Gender and Age Recognition from Speech Sentences // Aro, The Scientific Journal of Koya University. 2015. Vol. 3, Issue 2. P. 24–29. doi: http://doi.org/10.14500/aro.10072

8. Harb H., Liming C. Gender identification using a general audio classifier // 2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No.03TH8698). IEEE, 2003. doi: http://doi.org/10.1109/icme.2003.1221721

9. Сорокин В. Н., Макаров И. С. Определение пола диктора по голосу // Акустический журнал. 2008. Т. 54, № 4. С. 659–668.

10. Robust GMM Based Gender Classification using Pitch and RASTA-PLP Parameters of Speech / Zeng Y., Wu Z., Falk T., Chan W. // 2006 International Conference on Machine Learning and Cybernetics. Dalian, 2006. P. 3376–3379. doi: http://doi.org/10.1109/icmlc.2006.258497

11. Пресняков И. Н., Омельченко С. В. Помехоустойчивые алгоритмы сегментации речи в системах обработки // Радиотехника. 2003. № 131. С. 165–177.

Рекомендовано до публікації д-р техн. наук Безрук В. М.
Дата надходження рукопису 28.05.2019

Омельченко Сергей Васильевич, кандидат технических наук, доцент, кафедра информационно-сетевой инженерии, Харьковский национальный университет радиоэлектроники, пр. Науки, 14, г. Харьков, Украина, 61166
E-mail: serhii.omelchenko@nure.ua