

**ОПТИМИЗАЦИЯ БИБЛИОТЕКИ ЭТАЛОНОВ ПРИ ИДЕНТИФИКАЦИИ ДИКТОРА ПО КРОССКОРРЕЛЯЦИОННЫМ ПОРТРЕТАМ**

© 2024 Е.Ю. Галицкая, В.Р. Крашенинников

Ульяновский государственный технический университет, Ульяновск, Россия

Статья поступила в редакцию 06.06.2024

В статье рассматривается применение модели речевого сигнала в виде кросскорреляционного портрета к задаче текстозависимой идентификации диктора. Эти портреты являются двумерными массивами, состоящими из выборочных значений локальных коэффициентов корреляции двух сигналов. На примере двух дикторов показано, что особенности голоса человека проявляются на портретах его речевых сигналов так, что каждый диктор имеет свой уникальный портрет его произнесения речевой команды. На этом свойстве основан метод идентификации диктора по портретам его произнесений команд. Основу метода составляет сравнение портретов команд «неизвестного» диктора (диктора, которого необходимо идентифицировать и для которого в базе данных хранятся его эталонные портреты) с заранее заданными эталонными портретами для каждого класса дикторов. Свойство сохранения индивидуальности портретов на фоне довольно сильных шумов позволяет применять метод в обстановке акустических шумов. От выбора эталонных портретов существенно зависит частота правильной идентификации дикторов. Отсюда возникает задача выбора таких произнесений для каждого класса команд, при которых портрет команды диктора будет наиболее близок ко всевозможным портретам «своего» диктора и наиболее удален от портретов команд «чужого» диктора. В работе предложен метод направленного перебора, который позволяет из имеющегося множества произнесений выбрать наиболее удачные для использования в качестве эталонных. На реальном речевом материале проведен эксперимент, который доказал эффективность предложенного в работе метода оптимизации библиотеки эталонов.

*Ключевые слова:* речевая команда, идентификация диктора, кросскорреляционный портрет.

DOI: 10.37313/1990-5378-2024-26-4(3)-363-369

EDN: LHVLIO

**ВВЕДЕНИЕ**

Активное внедрение голосового управления во все сферы жизни человека вызвало новую проблему – возможность попыток несанкционированного доступа, когда посторонний человек с помощью голоса может получить доступ к «чужим» объектам, сведениям и т. п. Возникает задача идентификации личности по голосу. Использование уникальности голоса человека является единственным способом идентифицировать человека по телефону [1], что в современном мире сотовой связи является очень перспективным. На данный момент не определены характеристики голоса человека, которые со 100% вероятностью идентифицируют его. Задачей голосовой идентификации занимаются как отечественные фирмы, например, «Центр речевых технологий», SPIRIT Corp., ООО «Одитек» и другие, так и зарубежные, например, IBM T.J. Watson Research Center, Oregon Graduate Institute и другие [2]. В основе всех разработанных си-

стем можно выделить два этапа [3]. Первый – регистрация диктора в системе, выделение из его голоса определенных характеристик, то есть создание эталона. Второй этап – процесс идентификации. Из вновь поступившего голоса, выделяют аналогичные характеристики и сравнивают их с хранящимися в базе для определения наиболее близкого класса дикторов. Среди основных проблем при построении систем голосовой идентификации личности следует отметить изменчивость голоса человека и окружающие шумы [4]. В работе [5] приведен пример системы идентификации диктора, показывающей частоту правильной идентификации 100% (на определенной выборке команд) при отношении сигнал/шум, равном 2, но уже только около 60% при отношении 1.

Существуют системы текстозависимой и текстонезависимой идентификации [6]. В работах [7–10] описаны разработанные системы идентификации диктора. Вероятность правильной распознавания в этих системах составляет от 0,81 до 0,989. В настоящей работе будут рассматриваться текстозависимые системы с ограниченным словарем. Один из методов идентификации диктора в таких системах является использование модели речевого сигнала в виде кросскорреляционного портрета, то есть изо-

*Галицкая Екатерина Юрьевна, соискатель степени кандидата технических наук на кафедре ПМИ.*

*E-mail: katrisa@yandex.ru*

*Крашенинников Виктор Ростиславович, доктор технических наук, профессор, профессор кафедры «Прикладная математика и информатика». E-mail: kvrulstu@mail.ru*

бражения, состоящего из выборочных значений локальных коэффициентов кросскорреляции сигнала [11]. Преимуществом этого метода является возможность его применения в условиях различных акустических шумов. Недостатком является зависимость частоты правильной идентификации от выбора библиотеки эталонов, которая при различном составе библиотеки может быть в очень широком диапазоне. В настоящей работе предлагается способ оптимизации состава библиотеки эталонов для повышения точности идентификации личности.

### МОДЕЛЬ РЕЧЕВОГО СИГНАЛА В ВИДЕ КРОССКОРРЕЛЯЦИОННОГО ПОРТРЕТА

В работе [12] описана модель речевого сигнала (РС) в виде кросскорреляционного портрета (ККП), который представляет собой двумерный массив из выборочных коэффициентов кросскорреляции отрезков двух сигналов. Пусть имеется два РС:  $S1$  и  $S2$ , состоящие из  $N_1$  и  $N_2$  отсчетов, соответственно. Разобьем каждый из сигналов на  $M$  отрезков одинаковой длины (возможно кроме последнего):  $s1_i, s2_i$ , где  $s1_i, s2_i$  – отрезки сигналов  $S1$  и  $S2$ , соответственно,  $i = 1 \dots M$ . Длина (количество отсчетов) отрезка  $s1_i$  равна  $N1$ , длина  $s2_i$  –  $N2$ . Вычислим выборочный коэффициент корреляции отрезка  $s1_i$  и отрезка длиной  $N1$ , отложенного от начала отрезка  $s2_i$ . Сдвинем начало отрезка  $s2_i$  на  $k = 1 \dots K$  отсчетов вправо, каждый раз вычисляя коэффициент корреляции отрезка  $s1_i$  и полученных сдвигом отрезков. Выполнив аналогичную процедуру для отрезков  $s1_i, s2_i$  ( $i = 2 \dots M - 2$ ), получим двумерный массив, то есть ККП:

$$W(S1, S2) = r(i, k) = \frac{\frac{1}{N_2} \sum_{j=0}^{N_2-1} s1_{i+N_1+j} * s2_{i+N_2+j+k} - \mu1_i * \mu2_{i+k}}{\sqrt{\sigma1_i^2 * \sigma2_{i+k}^2}}, \quad (1)$$

где  $i = 0 \dots M - 3, k = 0 \dots K, \mu1_i = \frac{1}{N_1} \sum_{j=0}^{N_1-1} s1_{i+N_1+j}, \mu2_i = \frac{1}{N_2} \sum_{j=0}^{N_2-1} s2_{i+N_2+j+k}$  – выборочные средние,  $\sigma1_i^2 = \frac{1}{N_1} \sum_{j=0}^{N_1-1} s1_{i+N_1+j}^2 - \mu1_i^2, \sigma2_i^2 = \frac{1}{N_2} \sum_{j=0}^{N_2-1} s2_{i+N_2+j}^2 - \mu2_{i+k}^2$  – выборочные дисперсии. Если  $S1 = S2$ , то ККП превращается в автокорреляционный портрет (АКП) команды  $S1$ :  $W(S1, S1) = W(S1)$ , так как он состоит из коэффициентов автокорреляции сигнала  $S1$ .

Исследования этой модели выявили ее основные свойства [13]. Это индивидуальность (отличия между РС влекут за собой разницу их ККП), сохранение индивидуальности на фоне

довольно сильных шумов, инвариантность к линейным преобразованиям РС и некоторые другие. Кроме того, оказалось, что ККП не только обладают свойством индивидуальности для каждого класса речевых команд (РК), но и свойством индивидуальности для каждого диктора. На основе данного свойства в работе [11] было предложено применять модель РС в виде ККП к задаче идентификации диктора. Оказалось, что расстояния между ККП одного класса РК, произнесенных разными дикторами, больше, чем расстояния между ККП одного класса РК, произнесенных одним диктором. Под расстоянием между ККП  $W1$  и  $W2$ , понимается величина

$$\rho = \rho(W1, W2) = \sum_{i=0}^{M-3} d(w1_i, w2_i), \quad (2)$$

где  $w1_i, w2_i$  –  $i$ -е строки портретов  $W1$  и  $W2$ ,  $d$  – некоторая функция, которая позволяет численно сравнить два вектора между собой. Мы будем в качестве функции  $d$  использовать метрику квадрата евклидова расстояния.

Рассмотрим распределение вероятностей расстояния для одного класса РК у разных дикторов на следующем примере. Пусть имеется два диктора: женщина ( $D_1$ ) и мужчина ( $D_2$ ). Каждый из них произносит команду  $P = \text{"Кабина"}$  по 10 раз в бесшумной обстановке:  $p_1^1, \dots, p_{10}^1, p_1^2, \dots, p_{10}^2$ , где  $p_j^i$  –  $j$ -е произнесение команды  $i$ -ым диктором. Из этих команд можно сформировать по 100 разнообразных ККП ( $M = 100, K = 15$ ) для каждого диктора:  $\Omega_1, \Omega_2$ , где  $\Omega_1, \Omega_2$  – множество ККП для диктора  $D_1$  и  $D_2$ , соответственно. Построим гистограмму внутриклассовых расстояний для множества ККП  $\Omega_1$  и межклассовых расстояний между множествами ККП  $\Omega_1$  и  $\Omega_2$ . На рисунке 1 слева показана гистограмма внутриклассовых расстояний между ККП  $\Omega_1$ . Среднее значение и СКО соответственно равны 38.42 и 8.73. На рисунке 1 справа показана гистограмма межклассовых расстояний между ККП  $\Omega_1$  и  $\Omega_2$ . Среднее значение и СКО соответственно равны 245.03 и 28.68.

Заметна существенная разница между средними значениями внутри- и межклассовых расстояний между командами одного класса для разных дикторов. Это различие позволяет строить на основе ККП текстозависимый метод идентификации дикторов. Аналогичные результаты были получены для других классов команд и других дикторов.

Для исследования влияния шума добавим аддитивно к командам из рассмотренного выше примера шум (например, шум дождя [14]) с различным отношением сигнал/шум (ОСШ) в дисперсиях. Это отношение будем вычислять следующим образом. Найдем среднее значение  $\mu$  всех ОСШ для 10 произнесений РК диктором  $D_1$ . Подберем коэффициент к шуму так, чтобы среднее значение ОСШ для произнесений диктора  $D_2$  было равно  $\mu$ . Сформируем ККП из зашум-

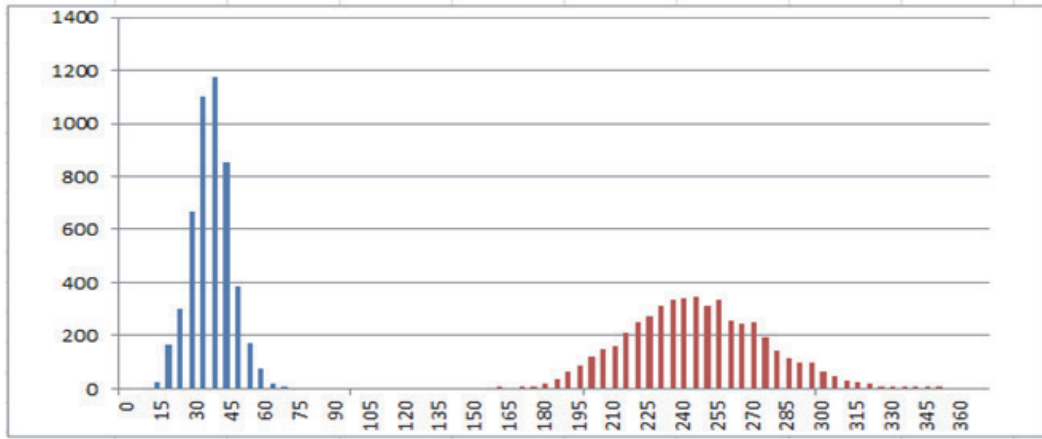


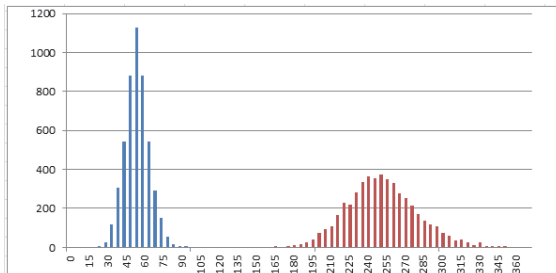
Рис. 1. Гистограмма внутриклассовых (левая) расстояний класса  $\Omega_1$  и межклассовых (правая) расстояний между портретами классов  $\Omega_1$  и  $\Omega_2$

ленных РК и построим гистограммы распределения вероятностей расстояний. На рисунке 2 показаны гистограммы расстояний для ОСШ = 5; 2,5; 1; 0,5; 0,33; 0,1. Левая гистограмма соответствует внутриклассовым расстояниям, правая – межклассовым расстояниям.

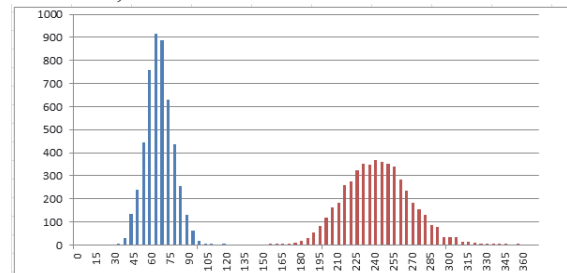
Как видно из рисунка, средние значения гистограмм по-прежнему находятся далеко друг

от друга при ОСШ = 1 и выше. При ОСШ ниже 0,5 гистограммы начинают сближаться, так как портреты при сильных шумах становятся фактически портретами шума, и точная идентификация дикторов становится затруднительной. Данное свойство различия ККП одного класса РК для разных дикторов на фоне шума позволяет применять ККП к задаче идентификации

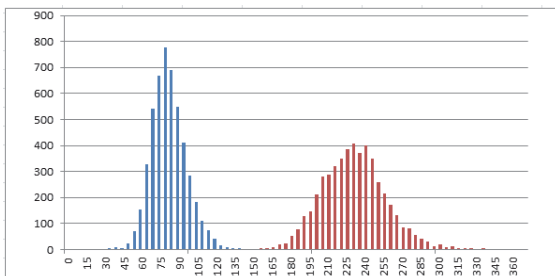
ОСШ = 5



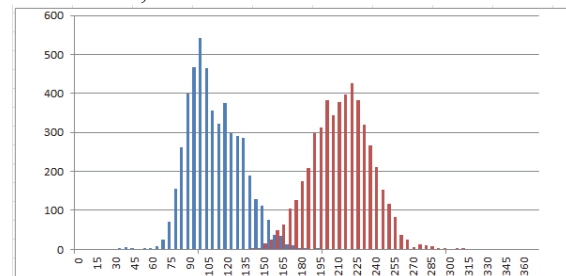
ОСШ = 2,5



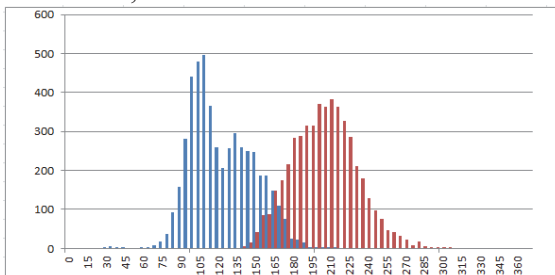
ОСШ = 1



ОСШ = 0,5



ОСШ = 0,33



ОСШ = 0,1

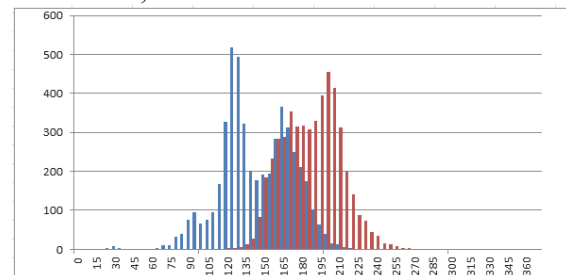


Рис. 2. Гистограмма внутриклассовых (левая) расстояний класса  $\Omega_1$  и межклассовых (правая) расстояний между портретами классов  $\Omega_1$  и  $\Omega_2$  при различном ОСШ

диктора на фоне различных акустических шумов. Однако при сильных шумах для получения достаточно уверенной идентификации дикторов требуется предпринять меры по ослаблению влияния шумов.

### МЕТОД ИДЕНТИФИКАЦИИ ДИКТОРОВ ПО КРОССКОРРЕЛЯЦИОННЫМ ПОРТРЕТАМ. ОПТИМИЗАЦИЯ БИБЛИОТЕКИ ЭТАЛОНОВ

Метод идентификации диктора на основе ККП состоит в следующем. Пусть имеется  $M$  дикторов:  $D_j, j = 1 \dots M$ , и словарь из  $N$  команд:  $P_i, i = 1 \dots N$ . Для каждого диктора  $D_j$  в базе данных хранятся эталонные ККП всех классов команд  $P_i: W_i^{stj} = (p_{1i}^{stj}, p_{2i}^{stj})$ , где  $p_{1i}^j, p_{2i}^j$  – первое и второе, соответственно, эталонные произнесение команды класса  $P_i$  диктором  $D_j$ . Требуется идентифицировать диктора  $D$ , которым может быть любой из  $M$  имеющихся дикторов.

Диктор  $D$  произносит по одному разу каждую из  $N$  команд:  $r_1, \dots, r_N$ . Из этих произнесений строятся ККП для каждого класса дикторов:  $W_i^j = W(p_{1i}^{stj}, r_i)$ . Определяется суммарное расстояние для каждого класса дикторов:  $s^j = \sum_{i=1}^N \rho(W_i^j, W_i^{stj})$ . Диктор  $D$  относится к классу с минимальным расстоянием.

Основой этого метода является схожесть портретов класса команд у одного диктора и различие этого класса команд у разных дикторов. Необходимо задать эталонные ККП для каждого диктора, с которыми и будет происходить дальнейшее сравнение. Естественно, что состав библиотеки эталонов оказывает существенное влияние на относительную частоту правильной идентификации (ЧПИ). Отсюда возникает задача выбора «наилучших» эталонов, то есть оптимизации состава библиотеки эталонов.

Сформулируем задачу оптимизации библиотеки при идентификации диктора по ККП. Пусть имеется  $M$  дикторов:  $D_j, j = 1 \dots M$  и словарь из  $N$  РК:  $P_i, i = 1 \dots N$ . Для каждого диктора имеется множество произнесений каждой команды из словаря:  $P_i^j = \{p_{i,1}^j, \dots, p_{i,l}^j\}$ , где  $i = 1 \dots N$  – номер класса РК,  $j = 1 \dots M$  – номера класса диктора,  $l = 1 \dots L$  – номер произнесения,  $L$  – количество произнесений. Требуется сформировать библиотеку эталонных команд для каждого диктора:  $E^j, j = 1 \dots M$ . Критерием качества сформированной библиотеки эталонов будем считать ЧПИ на множестве произнесений  $\Omega$ , на которых будет производиться идентификация дикторов.

Для решения данной задачи воспользуемся методом направленного перебора, который показал хорошие результаты для оптимизации библиотеки эталонов при распознавании РК по ККП [15] и модифицируем его для метода идентификации диктора.

Выберем произвольным образом для каждого класса команд у первого диктора набор эталонных РК:  $D_1: \{(p_{1rnd(L)}^1, p_{2rnd(L)}^1), \dots, (p_{1Nrnd(L)}^1, p_{2Nrnd(L)}^1)\}$ .

Аналогичную процедуру произведем для всех оставшихся дикторов  $D_2, \dots, D_M$ .

Заменим у первого диктора первую эталонную РК  $p_{1rnd(L)}^1$  на произнесение:  $p_{11}^1$ . Остальные эталонные РК для всех дикторов оставим без изменения. Определим ЧПИ на множестве  $\Omega$  с полученной библиотекой эталонов  $f_1$ . Далее заменим первую эталонную РК  $p_{11}^1$  на  $p_{12}^1$ . Определим ЧПИ на множестве  $\Omega$  с полученной библиотекой эталонов  $f_2$ . Аналогично произведем замены первой эталонной РК для диктора  $D_1$  на оставшиеся произнесения этой команды, каждый раз вычисляя ЧПИ. Определим произнесение, при котором ЧПИ максимальна и примем его за первую эталонную команду класса  $P_1$  диктора  $D_1$ .

Аналогично определим вторую эталонную РК класса  $P_1$  диктора  $D_1$ .

Повторим процедуры (2-3) для первой и второй эталонной РК  $P_1$  оставшихся дикторов.

Аналогичную процедуру направленного перебора проведем для всех остальных эталонных произнесений РК всех дикторов.

Полученную библиотеку РК примем за эталонную.

### РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

На реальном речевом материале был опробован описанный в данной работе метод оптимизации библиотеки эталонов в задаче идентификации дикторов. В эксперименте приняли участие 4 диктора: женщина и трое мужчин. Словарь РК состоял из названия цифр: «один», «два», «три», «четыре», «пять». Данные РК были произнесены каждым диктором по 30 раз в бесшумной обстановке. Аддитивно к каждой РК был добавлен шум дождя [14] с ОСШ = 0,9. ОСШ вычисляется как среднее ОСШ для всех РК, участвующих в эксперименте. Для выборки эталонных РК использовались первые 10 произнесений. Последующие 20 произнесений использовались для оценки качества сформированной библиотеки. Портреты строились с параметрами  $M = 60, K = 20$ . Случайным образом 20 раз был выбран набор эталонных РК для каждого диктора, ЧПИ распознавания по этим наборам были в диапазоне от 0,85 до 0,96. К наилучшему варианту с ЧПИ = 0,85 был применен описанный метод, в результате которого был получен вариант библиотеки с ЧПИ = 0,975. Таким образом, удалось существенно повысить ЧПИ. Исследования библиотеки эталонов также выявили, что наиболее высокие результаты показывает библиотека эталонов, где в качестве эталонного ККП, используется АКП.

Для оценки влияния акустических шумов, был проведен следующий эксперимент на РК, рассмотренных выше. Для «чистых» РК случайным образом 5 раз была сформирована библиотека эталонов, каждый раз вычислялась ЧПИ. К библиотеке, показавшей наилучшие результаты, был применён метод направленного перебора. В результате была получена библиотека эталонов:  $E^{st}$ . Следует отметить, что библиотека эталонов состоит из АКП. На оставшихся РК производилась идентификация диктора (по 29 произнесений набора РК каждым диктором), в результате получилась ЧПИ=1.

Для исследования влияния шумов был проведен следующий эксперимент. Ко всем РК был аддитивно добавлены различные шумы с ОСШ, вычисляемым как среднее арифметическое ОСШ по всем РК. Процедура идентификации дикторов была выполнена с применением метода зашумления эталонов [16, 17]. Описанный эксперимент был проведен со следующими шумами:  $\theta_1$  – Шум дождя,  $\theta_2$  – Шум в метро,  $\theta_3$  – Шум грузовой машины в работе,  $\theta_4$  – Шум промышленного миксера,  $\theta_5$  – Шум пневматической дрели,  $\theta_6$  – Звук отбойного молотка,  $\theta_7$  – Шум пассажирского самолета (приземление),  $\theta_8$  – Производственный шум. Результаты распознавания представлены в таблице 1. Как видно из таблицы 1, при ОСШ>1 получалась ЧПИ больше 0,98.

Таблица 1 – ЧПИ и ОСШ для различных шумов

Шум	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$
ОСШ	1,34	1,62	1,26	1,34	1,55	1,67	1,56	1,46
ЧПР	0,982	0,982	0,991	0,991	0,991	0,991	0,991	0,982

## ЗАКЛЮЧЕНИЕ

В работе рассмотрен метод формирования библиотеки эталонов из заранее заданных произнесений для метода идентификации диктора по ККП. На реальном речевом материале было проведено исследование по оценке эффективности предложенного метода. В эксперименте приняло участие 4 диктора. В результате применения описанного в работе алгоритма удалось увеличить ЧПИ с 0,85 до 0,975. Таким образом, эксперимент подтверждает эффективность метода направленного перебора для формирования библиотеки эталонов при распознавании РК по ККП.

Для оценки эффективности применения ККП к идентификации диктора на фоне акустических шумов, были проведены 8 экспериментов при различных шумах. Результаты распознавания показали, что при ОСШ от 1,26 до 1,67 ЧПИ больше 0,982.

## СПИСОК ЛИТЕРАТУРЫ

1. *Матвеев, Ю. Н.* Технологии биометрической идентификации личности по голосу и другим модальностям / Ю.Н. Матвеев // Вестник Московского государственного технического университета им. Н.Э. Баумана. – 2012. – № 3(3). – С. 46-60.
2. *Ручай, А.Н.* Биометрика. Текстозависимая верификация диктора: учебное пособие. Челябинск: Челябинский государственный университет / А.Н. Ручай. – 2012. – 105 с.
3. *Частикова, В.А.* Нейросетевая система биометрической личности по голосу / В.А. Частикова, С.А. Жерлицын, Д.О. Войлова // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. – 2023. – №. 1 (316). – С. 70-79.
4. *Вишнякова, О.А.* Подходы к задаче идентификации диктора / О.А. Вишнякова, Д.Н. Лавров // Математические структуры и моделирование. – 2011. – № 1(22). – С. 48-54.
5. *Карпенко, А. П.* Идентификация личности по голосу на основе самоорганизующейся карты Кохонена [Электронный ресурс] / А.П. Карпенко, А.Н. Морозов // Наука и образование: научное издание МГТУ им. НЭ Баумана. – 2012. – № 11. – С. 373-394.
6. *Нифонтов, С.В.* Применение скрытых марковских моделей в текстонезависимых системах идентификации пользователей по голосу / С.В. Нифонтов, Ю.С. Белов // Электронный журнал: наука, техника и образование. – 2016. – № 2. – С. 116-124.
7. *Васильев, Р.А.* Анализ возможностей применения голосовой идентификации в системах разграничения доступа к информации / Р.А. Васильев, Д.Б. Николаев // Научный результат. Информационные технологии. – 2016. – Т. 1. – № 1. – С. 48-57.
8. *Чернецова, Е.А.* Алгоритм идентификации личности по голосу для санкционирования доступа к информации / Е.А. Чернецова, А.Д. Шишкин // Международный научно-исследовательский журнал. – 2019. – № 2 (80). – С. 59-64.
9. *Брюхомицкий, Ю.А.* Метод текстонезависимой идентификации личности по голосу / Ю.А. Брюхомицкий, В.М. Федоров // Известия Южного федерального университета. Технические науки. – 2018. – № 8 (202). – С. 173-181.
10. *Бучнева, Т.И.* Нейронные сети в задаче идентификации личности по голосу / Т.И. Бучнева, М.Ю. Кудряшов // Вестник Тверского государственного университета. Серия: Прикладная математика. – 2015. – №2. – С. 119-126.
11. *Галицкая, Е.Ю.* Применение кросскорреляционных портретов речевых команд для голосовой идентификации диктора / Е.Ю. Галицкая, В.Р. Крашенинников // Автоматизация процессов управления. – 2022. – № 4(70). – С. 58-66.
12. *Krashennnikov V. R., Armer A.I., Kuznetsov V.V., Lebedeva E.Yu.* Cross-correlation portraits of voice signals in the problem of recognizing voice commands according to patterns // Pattern Recognition and

- Image Analysis. – 2011. – V. 21. – № 2. – P. 192-194.
13. Крашенинников, В.Р. Математические модели речевых сигналов в виде корреляционных портретов в задачах распознавания речевых команд и идентификации дикторов / Крашенинников В.Р., Галицкая Е.Ю. // Автоматизация процессов управления. – 2024. – № 1(75). – С. 40-48.
  14. База звуков [Электронный ресурс] – URL: <http://www.zvukogram.ru> (дата обращения 14.05.2024).
  15. Krasheninnikov V.R., Krasheninnikova N.A., Kuznetsov V.V., Lebedeva E.Yu. Optimization of dictionary and model library for recognition of speech commands // Pattern Recognition and Image Analysis. – 2011. – V. 21. – № 3. – P. 505-507.
  16. Тампель, И.Б. Автоматическое распознавание речи: учебное пособие / Тампель И.Б., Карпов А.А. – СПб.: Университет ИТМО, 2016. – 138 с.
  17. Галицкая, Е.Ю. Применение зашумления эталонов при распознавании речевых команд по их кросс-корреляционным портретам / Е.Ю. Галицкая, В.Р. Крашенинников // Автоматизация процессов управления. – 2023. – № 1(71). – С.88-99.

## OPTIMIZATION OF A LIBRARY OF STANDARDS FOR SPEAKER IDENTIFICATION BY CROSS-CORRELATION PORTRAITS

© 2024 E.Yu. Galitskaya, V.R. Krasheninnikov

Ulyanovsk State Technical University, Ulyanovsk, Russia

The article discusses the application of a speech signal model in the form of a cross-correlation portrait to the problem of text-dependent speaker identification. These portraits are two-dimensional arrays consisting of sample values of the local correlation coefficients of two signals. Using the example of two speakers, it is shown that the features of a person's voice appear in the portraits of his speech signals so that each speaker has his own unique portrait of his pronunciation of a speech command. The method of identifying a speaker based on portraits of his command utterances is based on this property. The method is based on comparing the team portraits of an "unknown" speaker (the speaker who needs to be identified and for whom his reference portraits are stored in the database) with predefined reference portraits for each class of speakers. The property of preserving the individuality of portraits against a background of fairly strong noise allows the method to be used in an environment of acoustic noise. The frequency of correct identification of speakers significantly depends on the choice of reference portraits. This raises the problem of choosing such utterances for each class of commands in which the portrait of the announcer's command will be closest to all possible portraits of "your" announcer and most distant from the portraits of the commands of the "foreign" announcer. The paper proposes a method of directed enumeration, which allows one to select the most successful ones from the available set of utterances for use as reference ones. An experiment was conducted on real speech material, which proved the effectiveness of the method proposed in the work for optimizing the standard library.

*Key words:* speech command, speaker identification, cross-correlation portrait.

DOI: 10.37313/1990-5378-2024-26-4(3)-363-369

EDN: LHVLIIO

## REFERENCES

1. Matveev, Yu.N. Tekhnologii biometricheskoy identifikatsii lichnosti po golosu i drugim modal'nostyam / Yu.N. Matveev // Vestnik Moskovskogo gosudarstvennogo tekhnicheskogo universiteta im. N.E. Baumana. – 2012. – № 3(3). – S. 46-60.
2. Ruchaj, A.N. Biometrika. Tekstozavisimaya verifikatsiya diktora: uchebnoe posobie. Chelyabinsk: Chelyabinskij gosudarstvennyj universitet / A.N. Ruchaj. – 2012. – 105 s.
3. Chastikova, V.A. Nejrosetevaya sistema biometricheskoy lichnosti po golosu / V.A. Chastikova, S.A. Zherlicyn, D.O. Vojlova // Vestnik Adygejskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskie i tekhnicheskie nauki. – 2023. – № 1 (316). – S. 70-79.
4. Vishnyakova, O.A. Podhody k zadache identifikatsii diktora / O.A. Vishnyakova, D.N. Lavrov // Matematicheskie struktury i modelirovanie. – 2011. – № 1(22). – S. 48-54.
5. Karpenko, A. P. Identifikatsiya lichnosti po golosu na osnove samoorganizuyushchetsya karty Kohonena [Elektronnyj resurs] / A.P. Karpenko, A.N. Morozov // Nauka i obrazovanie: nauchnoe izdanie MGTU im. NE Baumana. – 2012. – № 11. – S. 373-394.
6. Nifontov, S.V. Primenenie skrytyh markovskih modelej v tekstonezavisimyh sistemah identifikatsii pol'zovatelej po golosu / S.V. Nifontov, Yu.S. Belov // Elektronnyj zhurnal: nauka, tekhnika i obrazovanie. – 2016. – № 2. – S. 116-124.
7. Vasil'ev, R.A. Analiz vozmozhnostej primeneniya golosovoj identifikatsii v sistemah razgranicheniya dostupa k informatsii / R.A. Vasil'ev, D.B. Nikolaev // Nauchnyj rezul'tat. Informacionnye tekhnologii. – 2016. – T. 1. – № 1. – S. 48-57.
8. Chernecova, E.A. Algoritm identifikatsii lichnosti po golosu dlya sankcionirovaniya dostupa k informatsii / E.A. Chernecova, A.D. Shishkin // Mezhdunarodnyj nauchno-issledovatel'skij zhurnal. – 2019. – № 2 (80). – S. 59-64.
9. Bryuhomickij, Yu.A. Metod tekstonezavisimoj

- identifikacii lichnosti po golosu / YU.A. Bryuhomickij, V.M. Fedorov // Izvestiya YUzhnogo federal'nogo universiteta. Tekhnicheskie nauki. – 2018. – № 8 (202). – S. 173-181.
10. *Buchneva, T.I.* Neironnye seti v zadache identifikacii lichnosti po golosu / T.I. Buchneva, M.Yu. Kudryashov // Vestnik Tverskogo gosudarstvennogo universiteta. Seriya: Prikladnaya matematika. – 2015. – №2. – S. 119-126.
  11. *Galickaya, E.Yu.* Primenenie krosskorrelyacionnyh portretov rechevyh komand dlya golosovoj identifikacii diktora / E.Yu. Galickaya, V.R. Krasheninnikov // Avtomatizaciya processov upravleniya. – 2022. – № 4(70). – S. 58–66.
  12. *Krasheninnikov V. R., Armer A.I., Kuznetsov V.V., Lebedeva E.Yu.* Cross-correlation portraits of voice signals in the problem of recognizing voice commands according to patterns // Pattern Recognition and Image Analysis. – 2011. – V. 21. – № 2. – P. 192-194.
  13. *Krasheninnikov, V.R.* Matematicheskie modeli rechevyh signalov v vide korrelyacionnyh portretov v zadachah raspoznavaniya rechevyh komand i identifikacii diktora / Krasheninnikov V.R., Galickaya E.YU. // Avtomatizaciya processov upravleniya. – 2024. – № 1(75). – S. 40-48.
  14. Baza zvukov [Elektronnyj resurs] – URL: <http://www.zvukogram.ru> (data obrashcheniya 14.05.2024 ).
  15. *Krasheninnikov V.R., Krasheninnikova N.A., Kuznetsov V.V., Lebedeva E.Yu.* Optimization of dictionary and model library for recognition of speech commands // Pattern Recognition and Image Analysis. – 2011. – V. 21. – № 3. – P. 505-507.
  16. *Tampel', I.B.* Avtomaticheskoe raspoznavanie rechi: uchebnoe posobie / Tampel' I.B., Karpov A.A. – SPb.: Universitet ITMO, 2016. – 138 s.
  17. *Galickaya, E.Yu.* Primenenie zashumleniya etalonov pri raspoznavanii rechevyh komand po ih krosskorrelyacionnym portretam / E.Yu. Galickaya, V.R. Krasheninnikov // Avtomatizaciya processov upravleniya. – 2023. – № 1(71). – S. 88-99.