

УДК 519.2

В. Р. Масалимова¹, И. П. Насонов¹, Г. Д. Гефан¹

¹*Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

ГАУССОВЫ МОДЕЛИ ГОЛОСА ЧЕЛОВЕКА В ЗАДАЧАХ ВЕРИФИКАЦИИ И ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО РЕЧЕВЫМ СИГНАЛАМ

Аннотация: *Одна из актуальных проблем в области защиты информации - внедрение систем распознавания личности по голосу. Эти системы позволяют идентифицировать личность человека по совокупности уникальных характеристик голоса.*

В данной работе был создан и опробован алгоритм, основывающийся на Гауссовой модели голоса человека. Модель реализована на языке программирования Python. Этот алгоритм в дальнейшем может быть использован в качестве инструмента для верификации и идентификации личности по речевым сигналам говорящего.

Тестовые условные векторы из различных наборов сравнивались с эталонной моделью. Желаемый результат был подтвержден: чем существеннее различия между тестовым вектором и тренинговой (эталонной) моделью, тем меньшее значение принимает коэффициент подобия. Данный результат говорит о том, что вычислительный алгоритм успешно прошёл проверку и может использоваться для дальнейших испытаний с реальными речевыми векторами.

Ключевые слова: *верификация, идентификация, математическая модель голоса, речевые векторы, тренинговый и тестовый наборы.*

V. R. Masalimova¹, I. P. Nasonov¹, G. D. Gefan¹

¹*Irkutsk State Transport University, Irkutsk, Russian Federation*

GAUSSIAN MODELS OF THE HUMAN VOICE IN THE PROBLEMS OF VERIFICATION AND IDENTIFICATION OF THE PERSON FROM SPEECH SIGNALS

Abstract: *At this stage of development of data protection technologies, additional protection is being implemented through the analysis of human biometric data. One of the urgent problems in the field of information protection is the introduction of voice recognition systems. These systems make it possible to identify a person's personality by a set of unique characteristics of the voice.*

In this paper, an algorithm based on a Gaussian model of the human voice was created and tested. The model is implemented in the Python programming language. This algorithm can later be used as a tool for verification and identification of a person by the speaker's speech signals.

Test conditional vectors from different sets were compared with the reference model. The desired result was confirmed: the more significant the differences between the test vector and the training (reference) model, the lower the similarity coefficient takes. This result suggests that the computational algorithm has been successfully tested and can be used for further tests with real speech vectors.

Key words: *verification, identification, mathematical model of voice, speech vectors, training and test kits.*

Введение

С развитием информационных технологий появляется необходимость в дополнительной защите информации с помощью биометрических данных, таких как сетчатка глаза, отпечаток пальца, симметрия лица, особенности голоса. Внедрение систем распознавания личности по голосу – одна из актуальных проблем, связанных с безопасностью автоматизированных систем. Они позволяют идентифицировать личность человека по совокупности уникальных характеристик голоса. Данные системы осуществляют проверку посредством верификации и идентификации личности, которая запрашивает право доступа к определенным элементам системы.

Верификация и идентификация

Верификация личности – это процедура, позволяющая определить, имеет ли данный человек право доступа к запрашиваемой информации. Верификация необходима для получения доступа к конфиденциальной информации или другим ограниченными ресурсам. Для этого используются различные методы, например, запрос пароля или проверка биометрических данных. Одним из таких методов является система автоматического распознавания голоса, которая может подтвердить или отвергнуть гипотезу о личности говорящего.

Идентификацию можно разделить на закрытую и открытую. В первом случае известно, что диктор является представителем замкнутой группы, и системе необходимо только определить, какому диктору принадлежит голос. Диктор, речевая модель которого имеет минимальную величину отличия от исследуемого высказывания, признается идентифицированным [1].

Задачу открытой идентификации обычно сводят к многократной верификации, то есть не просто вычисляют степень отличия некоторого высказывания от имеющихся эталонов, но и проверяют, не превосходит ли даже минимальное из отличий некоторого порога. Если дело обстоит именно так, то делается вывод, что опознать говорящего невозможно, поскольку его нет в базе.

В задачах верификации (идентификации) личности говорящего существует две гипотезы: выдвинутая (основная, нулевая) гипотеза H_0 о том, что говорящий имеет право доступа, и конкурирующая гипотеза H_1 о том, что права доступа он не имеет. Система может совершить две ошибки. Первая ошибка заключается в том, что система не сможет найти зарегистрированного пользователя в базе данных и откажет ему в доступе. Другая ошибка совершается тогда, когда система принимает постороннее лицо за зарегистрированного пользователя, и, как следствие, предоставляет ему право доступа. Свести вероятности этих ошибок к нулю не получится, так как с уменьшением вероятности одной ошибки возрастает вероятность другой ошибки, и наоборот [1].

Параметризация речевых сигналов

На первом этапе решения задачи верификации (идентификации) личности говорящего производится параметризация речевых сигналов. Параметрическая модель речевого сигнала описывается конечным числом переменных, называемых параметрами или признаками модели, совокупность которых образует пространство признаков.

Таким образом, после параметризации получается последовательность m -мерных речевых векторов вида:

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) \\ \dots \\ y_m(n) \end{bmatrix}, \quad (n = \overline{1, N}), \quad (1)$$

где n относится к данному моменту времени. Если эти векторы характеризуют говорящего, личность которого известна, то набор называется тренинговым или эталонным (для данного говорящего). Если же векторы относятся к говорящему, личность которого верифицируется или идентифицируется, то набор называется тестовым. Тренинговые наборы служат для построения эмпирических эталонных моделей речи известных говорящих, и по степени сходства тестового набора с той или иной эталонной моделью делается вывод о верификации (идентификации) говорящего [2].

Модели речевых сигналов. Гауссовы модели

Модели речевых сигналов строятся на статистической обработке эмпирических данных, относящихся к речевым векторам разных дикторов. Решение такой задачи различается по типу

работы с исходными данными.

К наиболее известным методам построения моделей речевых сигналов относятся: векторное квантование [3], гауссова модель [4], метод опорных векторов [5], авторегрессионная векторная модель [6] и марковская модель [7]. Последние два из перечисленных методов рассматривают набор речевых векторов в динамике.

В данной работе был рассмотрен подход, который не рассматривает динамику речевого вектора. В таком случае может использоваться любой набор речевых векторов независимо от порядка. Исходя из выбранного подхода, была использована Гауссова модель [2]. Целью работы является создание и опробование алгоритма, который в дальнейшем может быть использован как инструмент верификации и идентификации личности по голосу говорящего.

Гауссовы модели рассматривают речевой вектор \mathbf{y} как m -мерную случайную величину с нормальным распределением плотности вероятности

$$b(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} |Y|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^T Y^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right]. \quad (2)$$

В данной формуле:

$\boldsymbol{\mu}$ – средний вектор, вычисленный по всему набору эталонных векторов;

Y – ковариационная матрица

$$Y = \frac{1}{N} \begin{pmatrix} \Sigma(y_1 - \mu_1)^2 & \Sigma(y_1 - \mu_1)(y_2 - \mu_2) & \dots & \Sigma(y_1 - \mu_1)(y_m - \mu_m) \\ \Sigma(y_2 - \mu_2)(y_1 - \mu_1) & \Sigma(y_2 - \mu_2)^2 & \dots & \Sigma(y_2 - \mu_2)(y_m - \mu_m) \\ \dots & \dots & \dots & \dots \\ \Sigma(y_m - \mu_m)(y_1 - \mu_1) & \Sigma(y_m - \mu_m)(y_2 - \mu_2) & \dots & \Sigma(y_m - \mu_m)^2 \end{pmatrix} \quad (3)$$

В одномерном случае ($m = 1$) ковариационная матрица становится обычной выборочной дисперсией, а (2) превращается в формулу плотности распределения вероятностей одномерной нормальной случайной величины и описывает кривую Гаусса.

Таким образом, для идентификации диктора будем использовать его эталонную модель с двумя параметрами: средний вектор $\boldsymbol{\mu}$ и ковариационная матрица Y . Система получает на вход тестовый набор речевых векторов неизвестного диктора. Каждый такой вектор «пропускается» через эталонную модель. Это означает, что все тестовые векторы последовательно подставляются вместо \mathbf{y} в формулу (2), что позволяет получить вероятность того, что эти векторы были сгенерированы эталонным диктором. В задаче верификации, если эта вероятность превосходит некоторый порог, то говорящий верифицируется. В задаче идентификации тот диктор из базы данных, для которого эта вероятность принимает максимальное значение, превосходя при этом некоторый порог, считается идентифицированным [2].

Если компоненты вектора принять случайными величинами, независимыми друг от друга, то все недиагональные элементы матрицы ковариации (3) равны нулю. В этом случае матрица содержит только главную диагональ (состоящую из дисперсий компонент), а все остальные элементы – нули

$$Y = \frac{1}{N} \begin{pmatrix} \Sigma(y_1 - \mu_1)^2 & 0 & \dots & 0 \\ 0 & \Sigma(y_2 - \mu_2)^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma(y_m - \mu_m)^2 \end{pmatrix} \quad (4)$$

Так как в данном случае в математической модели будет рассматриваться одна и та же ковариационная матрица, коэффициент при экспоненте не будет изменяться, что позволяет нам не учитывать его при подсчёте плотности.

Разработка математической модели

Математическая модель разработана для ознакомления с работой системы идентификации и верификации личности по его условным речевым векторам с помощью Гауссовых моделей. Модель реализована на языке программирования Python. Для написания программы использован дополнительный модуль NumPy для работы с матрицами.

Будем считать, что речевой вектор имеет $m = 10$ компонент. Прежде всего, необходимо задать эталонную модель голоса набором $N = 50$ случайных векторов. Назовём этот набор «своим». В нём каждая координата – случайная величина с математическими ожиданиями, представленными в таблице 1.

Таблица 1

Формирование набора «Свой»

i	1	2	3	4	5	6	7	8	9	10
$M(x)$	1	2	3	4	5	5	4	3	2	1

Каждая координата задаётся как случайная величина с равномерным распределением на интервале $(i - 0.5; i + 0.5)$. Например, координата 4 из таблицы 1 генерируется, как случайное число, равномерно распределенное на интервале $[3.5; 4.5]$. Нужно сгенерировать 50 векторов и по ним сформировать средний вектор μ и ковариационную матрицу. Средний вектор μ будет иметь координаты, схожие со значениями $M(x)$ из таблицы 1.

Далее формируем следующие тренировочные наборы: «Чужой», «Свой, но размытый», «Смещенный». Набор «Чужой» будет формироваться по тому же принципу: в каждом интервале сигнал задается, как случайная величина с равномерным распределением на интервале $(i - 0.5; i + 0.5)$. Но математические ожидания – совершенно иные в сравнении с набором «Свой» (таблица 2).

Таблица 2

Формирование набора «Чужой»

i	1	2	3	4	5	6	7	8	9	10
$M(x)$	5	4	3	2	1	1	2	3	4	5

«Свой, но размытый» – этот набор векторов будет иметь те же математические ожидания, что и «Свой» набор (таблица 1). Но в каждом интервале сигнал задается, как случайная величина с равномерным распределением на интервале $(i - 1; i + 1)$, поэтому в данном случае дисперсия увеличится в 4 раза.

Далее формируем набор «Смещенный», что можно интерпретировать как условный переход от мужского голоса к женскому голосу или, иначе говоря, сдвиг интенсивности сигнала к более высоким частотам. В каждом интервале сигнал задается, как случайная величина с равномерным распределением на интервале $(i - 0.5; i + 0.5)$. Однако распределение математических ожиданий как бы смещено на один интервал вправо относительно значений «Своего» набора (таблица 3).

Таблица 3

Формирование набора «Смещенный»

i	1	2	3	4	5	6	7	8	9	10
$M(x)$	0	1	2	3	4	5	5	4	3	2

На рисунке 1 показан средний вектор эталонного набора и по одному случайному вектору из остальных (тестовых) наборов.

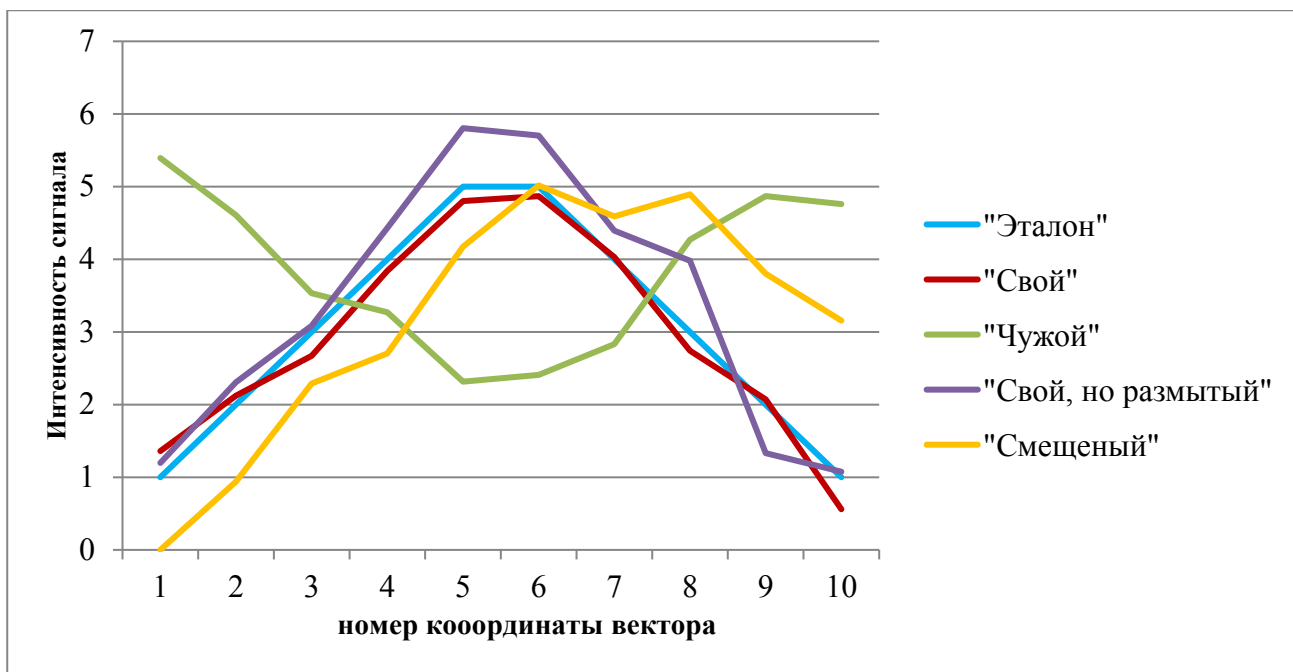


Рис. 1. Входные данные Гауссовой модели

Теперь пропускаем через эталонную модель случайные векторы, генерируемые различным образом. После этого на выходе мы получаем 50 коэффициентов подобия для каждого, сгенерированного ранее тестового набора. (Коэффициентом подобия мы называем показатель экспоненты в формуле (1)).

Коэффициенты подобия для векторов из набора «Свой» колеблются в значениях примерно от -3 до -9 , тогда как для векторов из набора «Чужой» коэффициенты подобия принимают значения от -400 до -550 . (Рис. 2). Это впечатляющий результат, говорящий о том, что если «чужой» вектор очень сильно отличается от эталонного, то принять его за «свой» нет никаких шансов.

Если в тестовой аудиозаписи будут помехи, приводящие к размытию спектра, то коэффициенты подобия уменьшаются, их значения будут в интервале от -10 до -25 (Рис. 2, «Свой, но размытый»). При смещении условного спектра коэффициенты подобия будут колебаться в диапазоне от -30 до -85 (Рис.2 «Смещенный»).

Заключение

В работе создан и опробован алгоритм, основанный на Гауссовой модели человеческого голоса, который в дальнейшем может быть использован как инструмент верификации и идентификации личности по речевым сигналам говорящего.

Условные тестовые векторы из различных наборов сравнивались с эталонной (тренинговой) моделью. Ожидаемый результат подтвердился: чем существеннее различия между тестовым вектором и тренинговой моделью, тем меньшее значение принимает коэффициент подобия. Это говорит о том, что вычислительный алгоритм успешно прошёл проверку и может использоваться для дальнейших испытаний с реальными речевыми векторами.

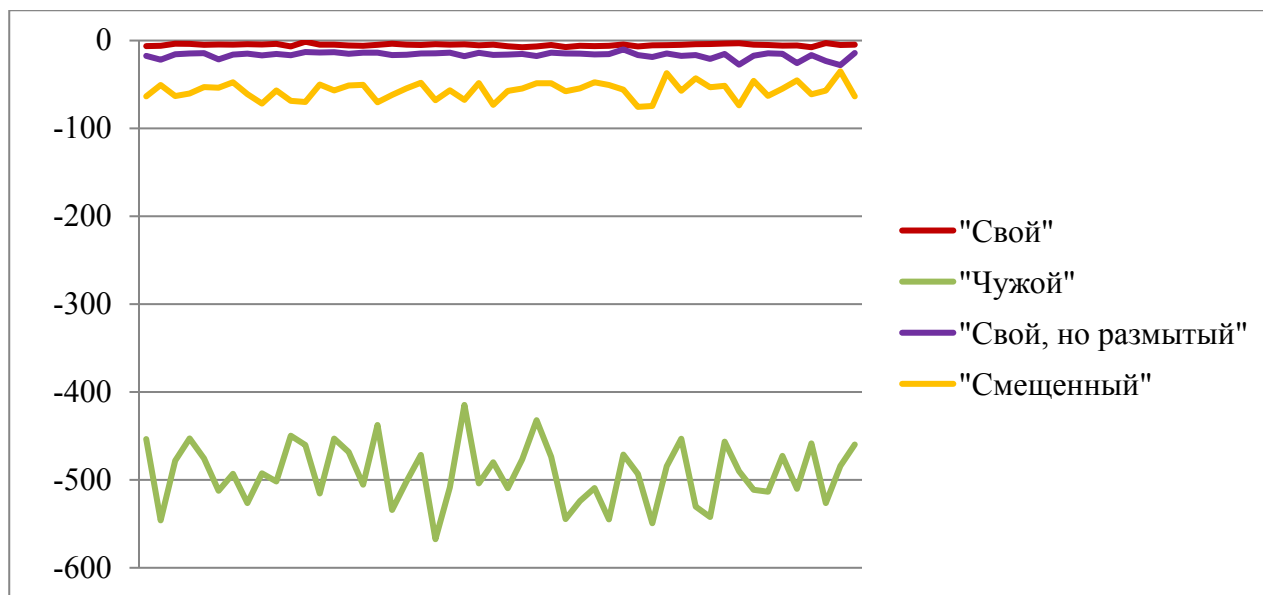


Рис. 2. Выходные данные: показатели подобия тестовых наборов по сравнению с эталонной моделью голоса

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Г.Д. Гефан, В.Б. Иванов, Авторегрессионные векторные модели и их применение в задачах идентификации личности по речевым сигналам // Информационные технологии и проблемы математического моделирования сложных систем. Иркутск: ИИТМ ИрГУПС, 2010. – Вып. 8, с. 74-81
2. Г.Д. Гефан, В.Б. Иванов, Математическое моделирование систем верификации и идентификации личности по речевым сигналам // Информационные системы контроля и управления в промышленности и на транспорте. Иркутск: ИрГУПС, 2010. – Вып. 17, с. 24-31
3. Picone J. Signal modeling techniques in speech recognition – 1993 [Электронный ресурс]. – Режим доступа: <http://aprodeus.narod.ru>
4. Bimbot F., Magrin-Chagnolleau I., Mathan L. Second Order Statistical Measures For Text Independent Speaker Identification // Speech Communication. – 1995. – 17(1-2). – P. 177-192.
5. Burges C.J.C. A tutorial on Support Vector Machines for pattern recognition // Data Mining and Knowledge Discovery. – 1998. – V. 2. – No 2. – P. 1-47.
6. Montacie C., Le Floch J.-L. AR-vector models for free-text speaker recognition // Proceeding of ICSLP. – 1992. – V.1. – P. 611-614.
7. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // Proc. IEEE. – 1989. – 77(2). – P. 257-286.

REFERENCES

1. G.D. Gefan, V.B. Ivanov, Autoregressive vector models and their application in the problem of personality identification by speech signals // Information technologies and problems of mathematical modeling of complex systems. Irkutsk: IITM IrGUPS, 2010. - Issue. 8, p. 74-81
2. G.D. Gefan, V.B. Ivanov, Mathematical modeling of systems for checking and identifying a person by speech signals // Information systems for control and management in industry and transport. Irkutsk: IrGUPS, 2010. - Issue. 17, p. 24-31
3. Picone J. Signal modeling techniques in speech recognition – 1993 [Электронный ресурс]. – Режим доступа: <http://aprodeus.narod.ru>
4. Bimbot F., Magrin-Chagnolleau I., Mathan L. Second Order Statistical Measures For Text Independent Speaker Identification // Speech Communication. – 1995. – 17(1-2). – P. 177-192.
5. Burges C.J.C. A tutorial on Support Vector Machines for pattern recognition // Data Mining and Knowledge Discovery. – 1998. – V. 2. – No 2. – P. 1-47.

6. Montacie C., Le Floch J.-L. AR-vector models for free-text speaker recognition // Proceeding of ICSLP. – 1992. – V.1. – P. 611-614.

7. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // Proc. IEEE. – 1989. – 77(2). – P. 257-286.

Информация об авторах

Масалимова Валерия Ринатовна – студентка, специальность «Безопасность автоматизированных систем», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: lera.masalimova@yandex.ru

Насонов Игорь Павлович – студент, специальность «Безопасность автоматизированных систем», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: tj97.dom@mail.ru

Гефан Григорий Давыдович - доцент кафедры «Математика», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: grigef@rambler.ru

Information about the authors

Masalimova Valeria Rinatovna - student, specialty "Safety of automated systems", Irkutsk State Transport University, Irkutsk, e-mail: lera.masalimova@yandex.ru

Nasonov Igor Pavlovich - student, specialty "Safety of automated systems", Irkutsk State University of Railways, Irkutsk, e-mail: tj97.dom@mail.ru

Gefan Grigory Davydovich - Associate Professor of the Department of Mathematics, Irkutsk State University of Railway Transport, Irkutsk, e-mail: grigef@rambler.ru