

СТАТЬЯ

УДК 004.42

КЕПСТРАЛЬНЫЙ АНАЛИЗ ЗВУКА ДЛЯ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ ПО ГОЛОСУ**Музафаров Р.Р., Соловьев Н.А.***ФГБОУ ВО «Оренбургский государственный университет», Оренбург, e-mail: post@mail.osu.ru*

Более 40 лет тому назад уже была поставлена задача распознавания человека по голосу, и ее изучением все еще занимаются во многих научно-исследовательских институтах, центрах разработки программного обеспечения и т.д. Конечно, за последние годы качество систем голосовой идентификации выросло, но оно все еще не близко к идеалу. Исходя из вышесказанного, можно прийти к мнению, что помимо исследований уже созданных алгоритмов нужно искать и новые решения в данной области. Задачей голосового распознавания является выделение, классифицирование и реагирование на речь человека из входного голосового сигнала. Благодаря развитию информационных технологий в голосовой биометрии заинтересованы государственные учреждения, бизнес-структуры и многочисленные категории пользователей информационных услуг. Существуют два типа систем: текстозависимые и текстонезависимые. Первые используют заранее определенные фразы или предложения. Вторые же лишены привязки к заранее заданным фразам, тем самым они могут работать со словами и предложениями любой длины. В данной работе изучается задача автоматической идентификации пользователя по его голосовым характеристикам и реализуется алгоритм, который решает задачу текстонезависимой идентификации.

Ключевые слова: идентификация голоса, спектральный анализ, кепстральный анализ, мел-кепстральные коэффициенты

CEPSTRAL SOUND ANALYSIS FOR VOICE IDENTIFICATION**Muzafarov R.R., Solovlev N.A.***Orenburg State University, Orenburg, e-mail: post@mail.osu.ru*

More than 40 years ago, the task of recognizing a person by voice was already set and it is still being studied in many research institutes, software development centers, etc. Of course, in recent years, the quality of voice recognition systems has increased, but it is still not close to ideal. Based on the foregoing, one can come to the conclusion that, in addition to researching already created algorithms, it is necessary to look for new solutions in this area. The task of voice recognition is to isolate, classify, and respond to human speech from the input voice signal. Thanks to the development of information technologies, state institutions, business structures and numerous categories of users of information services are interested in voice biometrics. There are two types of systems: text-dependent and text-independent. The former use predefined phrases or sentences. The latter are devoid of binding to predefined phrases, thus they can work with words and sentences of any length. In this work, the problem of automatic user identification by his voice characteristics is studied and an algorithm is implemented that solves the problem of text-independent identification.

Keywords: voice identification, spectral analysis, cepstral analysis, mel-cepstral coefficients

Речь – это неотъемлемая черта каждого человека, которая позволяет ему изучать окружающий мир, делиться с другими людьми своими знаниями и опытом. Благодаря голосовому аппарату человек может осуществлять речевые высказывания в звуковой форме, которые называются «устной речью». Индивидуальное строение голосовых органов характеризуется неповторимыми голосовыми характеристиками для каждого человека.

В связи с развитием информационных технологий и широким распространением средств связи открываются огромные возможности для использования голоса в качестве признака, с помощью которого можно будет устанавливать личность пользователя; помимо этого, голосовая идентификация очень удобна для конечных пользователей и не требует от них наличия сложного технического оборудования.

Средства и технологии голосовой идентификации прекрасно себя чувствуют во многих сферах: электронная коммерция, банковские технологии, криминалистика, безопасность, борьба с терроризмом и др.

В России в 2020 г. было установлено, что с использованием информационно-телекоммуникационных технологий было совершено более 500 тыс. преступлений. Из данных Министерства внутренних дел Российской Федерации следует, что это число на 73,4% больше, чем было в предыдущем году. Около 410 тыс. преступлений из них были совершены путем кражи или мошенничества [1].

Нельзя забывать, что голос (так же, как и походка, почерк и т.п.) относится к так называемым «поведенческим» или «динамическим» идентификаторам, т.е. голос, под влиянием эмоциональных факторов (настроение человека) или состояния здо-

ровья человека (насморк, бронхит, ангина и т.д.), склонен к серьезным изменениям, что в свою очередь может сильно повлиять на результат идентификации.

Также можно столкнуться и с другими проблемами, если распознавание проводится не в лабораторных условиях. Во-первых, аппаратура, которая записывает и обрабатывает голосовой сигнал, может вызывать искажения. Во-вторых, неизбежно наложение внешних акустических шумов на исходный сигнал, которые могут сильно повлиять на необходимые для идентификации характеристики. Поэтому достичь высокой точности и надежности идентификации – очень сложная задача.

Методы распознавания акустического сигнала бывают дикторозависимые и недикторозависимые [2]. Описываемый в работе метод относится к классу дикторозависимых методов распознавания, который учитывает голосовые признаки диктора. Для исследования голосового сигнала в дикторозависимых методах обычно используют кепстральный анализ [3, 4], который, по сути, является анализом спектра исследуемого сигнала, называемый «кепстром» [5].

Множество работ известных авторов посвящены теории обработки сигнала и методам идентификации диктора. Среди них работы таких авторов, как Матвеев [6], Dehak [7], Раев [8], Пеховский [9] и Симончик [10].

На основе сказанного ранее можно прийти к мнению, что нет сомнений в актуальности автоматизации процесса идентификации пользователя по голосу.

Цель исследования: повышение надежности голосовой идентификации личности в условиях внешнего механического шума.

Материалы и методы исследования

Преобразование Фурье для сигнала используется для получения информации, которая недоступна в исходном виде [11].

Большинство сигналов представляются во временной области, т.е. сигнал представляет зависимость амплитуды от времени. Такое представление не является лучшим, потому что наиболее значимая информация скрыта в частотной области. Преобразование Фурье позволяет перейти от временного представления сигнала к частотному. Временное и частотное измерение представлено на рис. 1.

Интегральное преобразование Фурье задается следующей формулой:

$$f(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ix\omega} dx. \quad (1)$$

Цифровая обработка сигналов работает сигналами с дискретными преобразованиями. Поэтому будет удобно представить сигнал в дискретном виде, воспользовавшись преобразованием Фурье:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-2\pi i}{N} kn}. \quad (2)$$

Необходимо отметить, что существуют недостатки разложения сигналов в ряды Фурье. Поэтому предпочтительнее использовать оконное преобразование Фурье с движущейся по сигналу оконной функцией. Преобразование Фурье выполняется для каждого окна (фрагмента) исходного сигнала, который предварительно был разделен на равные по длительности фрагменты (рис. 2). Тем самым выполняется переход к частотно-временному представлению сигналов, при этом в пределах каждого фрагмента сигнал считается стационарным. Этот способ помогает получать, анализировать и строить в виде спектрограмм динамические спектры и исследовать их поведение во времени. Спектрограмма строится в трех координатах – частота, время и амплитуда.

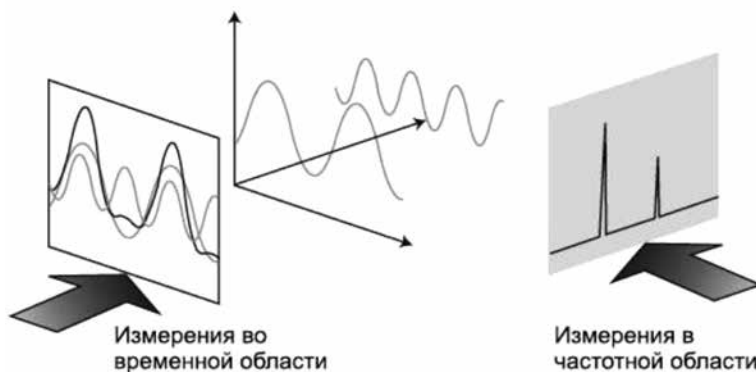


Рис. 1. Измерения сигнала во временной и частотной области

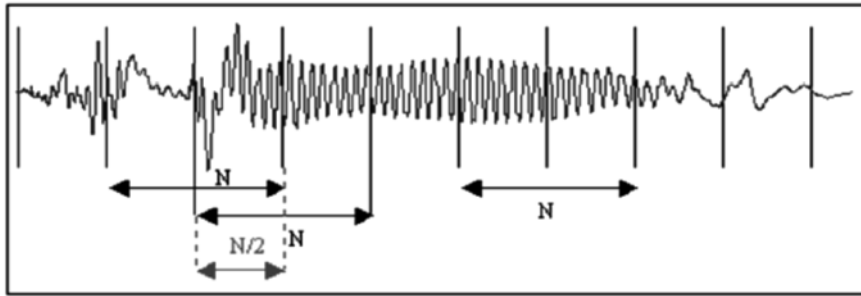


Рис. 2. Разбиение голосового сигнала на фрагменты

Поскольку анализируемый сигнал не является периодическим, приходится на следующем этапе умножать каждый фрагмент на оконную функцию, устраняющую разрывы на границах периодов. В качестве оконной функции мы выбрали функцию Хэмминга:

$$W(\tau - t) = 0,54 - 0,46 \cos \frac{2\pi i}{N}. \quad (3)$$

График, на котором изображена функция Хэмминга, представлен на рис. 3.

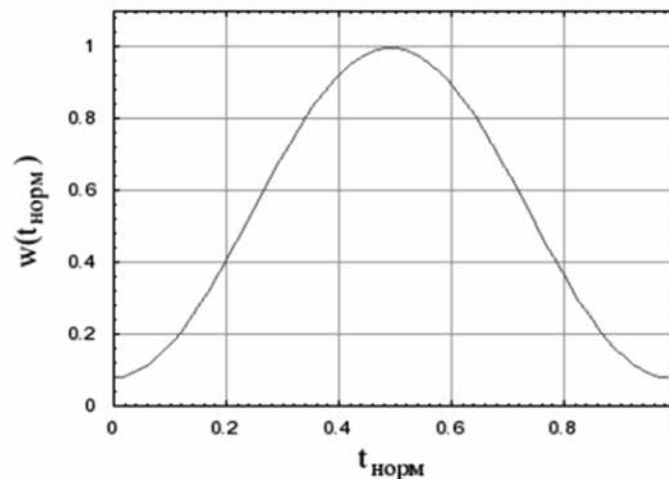


Рис. 3. Функция Хэмминга

Оконное преобразование Фурье (ОПФ) в интегральном виде задается следующей формулой:

$$F(t, \omega) = \int_{-\infty}^{\infty} f(\tau) W(\tau - t) e^{-i\omega\tau} d\tau. \quad (4)$$

Формула ОПФ в дискретном виде:

$$F(m, \omega) = \sum_{n=-\infty}^{\infty} f[n] \omega[n - m] e^{-j\omega n}. \quad (5)$$

Высота звука, которую человек воспринимает, не связана линейно с его частотой, она связана с тембром и уровнем громкости. Поэтому для измерения высоты воспринимаемого звука была придумана величина «мел».

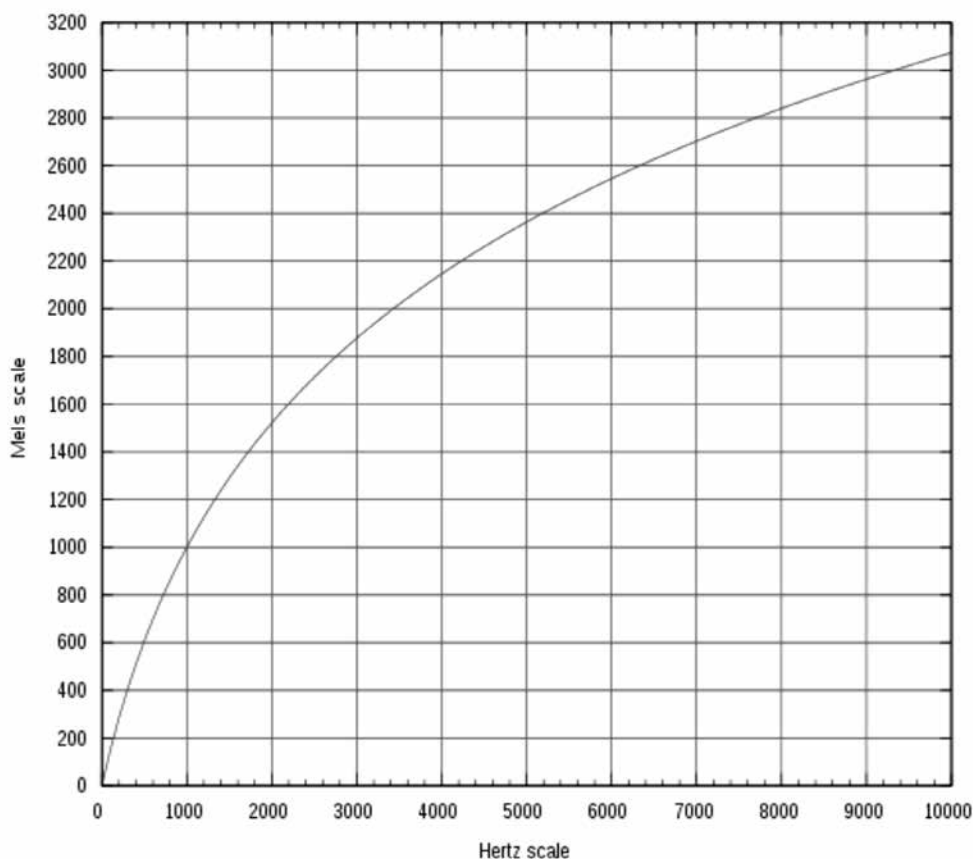


Рис. 4. График зависимости высоты звука в мелах от частоты колебаний

Переход к новой шкале описывается зависимостью

$$M(f) = 1125 \ln \left(1 + \frac{f}{700} \right), \quad (6)$$

График зависимости высоты звука в мелах от частоты колебаний изображен на рис. 4.

Вектор признаков будет состоять из мел-кепстральных коэффициентов. Вычисляются они по формуле

$$c_n = \sum_{k=1}^K (\log S_k) \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], \quad (7)$$

Задача классификации решалась методом городских кварталов «манхэттенское расстояние», т.е. вычислением расстояния от вектора признаков анализируемого сигнала $\mathbf{p}$ до вектора признаков $\mathbf{q}$ вектора шаблона:

$$d_1(p, q) = p - q_1 = \sum_{i=1}^n |p_i - q_i|. \quad (8)$$

Результаты исследования и их обсуждение

В ходе работы на языке C# был создан прототип системы распознавания, интерфейс которой приведен на рис. 5.

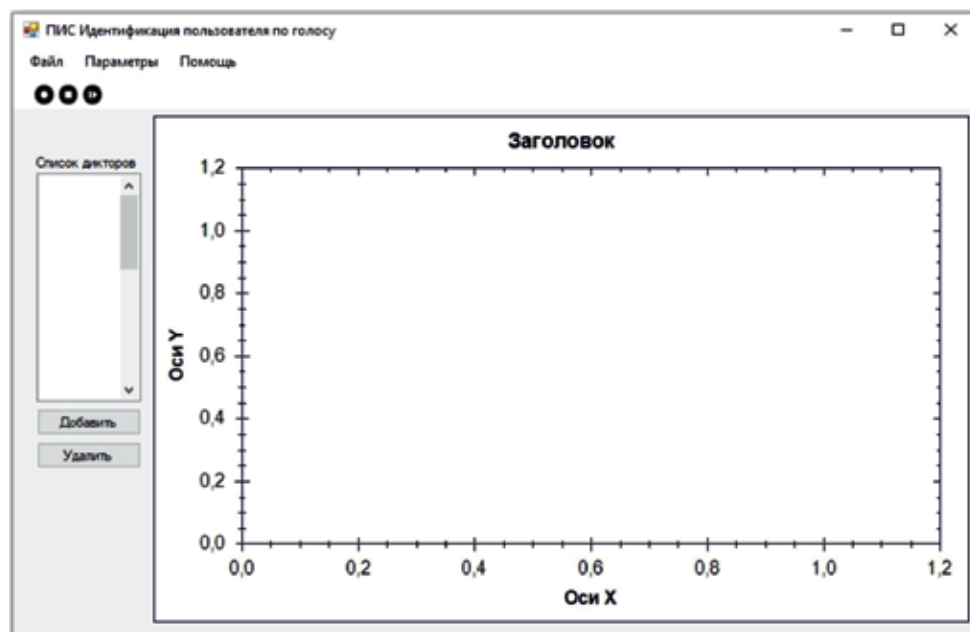


Рис. 5. Интерфейс программы идентификации диктора по голосу

Результаты серии экспериментов по определению вероятности правильного распознавания приведены в таблице.

Вероятности правильной идентификации дикторов

Количество дикторов	Количество тестов	Процент вероятности правильного распознавания, %
2	20	100
5	50	97
10	100	92

Заключение

В данной работе было проведено изучение использования систем голосовой биометрии в информационных системах. Были рассмотрены алгоритмы кепстрального анализа. Создан прототип системы распознавания, основанной на использовании мел-кепстрального анализа спектра голосового сигнала. Приведены результаты вероятности правильного распознавания.

Список литературы

1. В России число киберпреступлений за год увеличилось на 73%. [Электронный ресурс]. URL: <https://russian.rt.com/russia/news/828179-statistika-kiberprestupleniya-rossiya> (дата обращения: 12.05.2021).
2. Rabiner L., Juang B. Fundamental of Speech Recognition. Englewood Cliffs: Prentice-Hall N.J. 1993. P. 42–65.

3. Bogert B.P., Healy M.J.R., Tukey J.W. The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking, Proceedings of the Symposium on Time Series Analysis. New York: Wiley. 1963. Vol. 15. No. 8. P. 209–243.

4. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов: Пер. с англ. / Под ред. С.Ф. Боева. М.: Техносфера, 2012. С. 102–103.

5. Jeong J. Kepstrum Analysis and Real-Time Application to Noise Cancellation. Proceedings of the 8th WSEAS International Conference on SIGNAL PROCESSING, ROBOTICS and AUTOMATION. 2009. Vol. 15. No. 1. P. 82–91.

6. Матвеев Ю.Н. Технологии биометрической идентификации личности по голосу и другим модальностям // Вестник МГТУ. Приборостроение. Биометрические технологии. 2012. № 2. С. 46–59.

7. Dehak N., Kenny P., Dehak R., Dumouchel P., Ouellet P. Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech and Language Processing. 2011. Vol. 19. No. 8. P. 788–798.

8. Раев А.Н., Матвеев Ю.Н., Голощапова Т.И. Анализ влияния состояния наркотического опьянения на характеристики голосов дикторов // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 5. С. 87–93.

9. Пеховский Т.С., Лоханова А.И. Выбор UBM Модели с помощью Вариационного Байесовского Анализа для GMM-UBM Систем Распознавания Диктора // SPECOM-2011 Proceedings. 14th International conference on SPEECH and COMPUTER (Казань, 23–27 сентября 2011 г.). Казань: Издательство ПАЛАДИН, 2011. С. 327–332.

10. Симончик К.К. Метод и алгоритмы текстонезависимой верификации дикторов по голосу // LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. С. 188.

- Простыми словами о преобразовании Фурье. [Электронный ресурс]. URL: <https://habrahabr.ru/post/196374> (дата обращения: 24.05.2021).