УДК 004.56.5

СПЕКТРАЛЬНЫЙ И КЕПСТРАЛЬНЫЙ АНАЛИЗ ЗВУКА ДЛЯ ИДЕНТИФИКАЦИИ ГОЛОСА

Я. В. Грушко[1, а], В. Я. Данилов[1, б]

[1] *Учебно-научный комплекс «ИПСА» НТУУ «КПИ», кафедра ММСА*

Аннотация

В статье представлены алгоритмы спектрального и кепстрального анализа для идентификации мужского и женского голосов. Создан прототип системы распознавания, основанной на использовании мел-кепстрального анализа и энергетического распределения спектра голосового сигнала. Приведены вероятности дикторозависимого распознавания с использованием обоих подходов.

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

Введение

Голос – такая же неотъемлемая черта каждого человека, как и его лицо или отпечатки пальцев. Широкое распространение средств связи (стационарные и мобильные телефонные сети, интернет-телефония и т.д.) открывают большие возможности для применения данного идентификатора; кроме того, распознавание по голосу весьма удобно для пользователей и требует от них минимум усилий.

Технологии и средства идентификации по голосу применяются в ряде областей, непосредственно связанных с обработкой обращений пользователей по телефону (колл-центры и т.п.), что позволяет ускорить обслуживание абонентов и разгрузить операторов. В более значимых проектах (особенно связанных с необходимостью защиты конфиденциальной информации) идентификация по голосу играет немаловажную роль при разработке комплексных систем безопасности, в борьбе с терроризмом и др.

Необходимо учитывать, что голос (наряду с почерком, походкой и т.п.) относится к т.н. «поведенческим» идентификаторам. Он подвержен существенным изменениям под воздействием эмоциональных факторов (настроение человека) и состояния здоровья (ангина, насморк, бронхит и т.д.). На качестве идентификации могут сказываться также внешние условия (например, посторонние шумы от дорожного движения, разговоров других людей). Если для передачи голосовой информации используются линии связи, помехи в них также способны затруднить распознавание пользователя. Поэтому достичь высокой точности и надежности идентификации является чрезвычайно сложной задачей.

Методы распознавания акустического сигнала разделяют на дикторозависимые и недикторозависимые[1]. Представляемый в работе метод относится к классу дикторозависимых методов распознавания, который учитывает голосовые признаки говорящего. Для анализа голосового сигнала в дикторозависимых методах обычно применяют кепстральный анализ [2, 3] который, по-сути, представляет собой анализ спектра спектра анализируемого сигнала, называемый «кепстром» [4].

Популярным справочным руководством по идентификации человеческого голоса является работа И. Клемента, С. Мохова, Д. Николакопулуса, С. Синклара и др. «Modular Audio Recognition Framework v.0.3.0.6 (0.3.0 final) and its Applications» [5], где изложены основные подходы записи и анализа звуковых данных. В НТУУ «КПИ» проблемой анализа и параметризации речевых сигналов занимались Данилов В.Я. и Добрушкин Г.А. [6, 7, 8], которые рассмотрели дикторозависимую модель распознавания речевой информации на основе искусственных иммунных систем [8].

1. Метод

Кепстральный анализ голоса, выполненный в настоящей работе, основывается на методе идентификации, предложенном Р. Хасаном и др. [9].

Блок-схема метода приведена на(рис. 1)

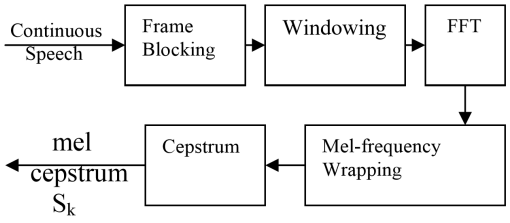

Рис. 1. Блок-схема метода мел-кепстрального анализа звука для идентификации голоса.

На начальном этапе, голосовой сигнал, записанный в виде WAV-файла в анализирующую систему,

[а] gornahur@ukr.net
[б] danilov1950@rambler.ru

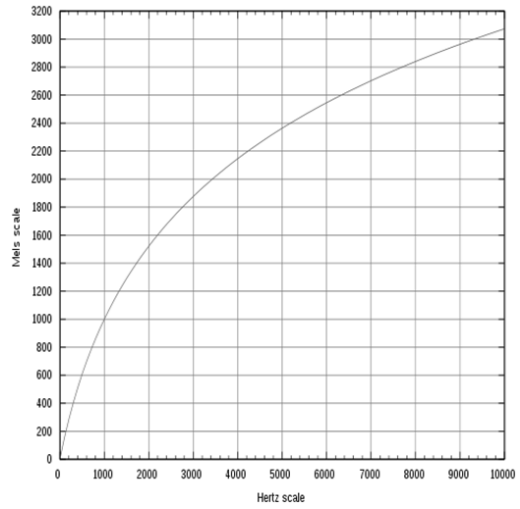

Рис. 4. Функция перехода от частоты в Гц к частоте в мелах

разбивается на фреймы с перекрытием $N/2$, где N – количество точек, составляющее период дискретного сигнала. Голосовой сигнал представленый на (рис. 2)

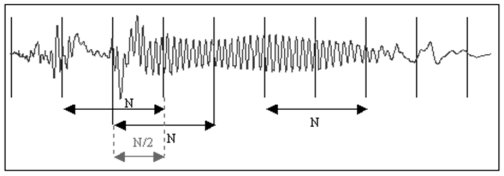

Рис. 2. Голосовой сигнал

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

$$W_{Hamm}[i] = 0.54 - 0.46cos\frac{2\pi i}{N} \tag{1}$$

График на котором изображена функция Хэмминга представлен на (рис. 3).

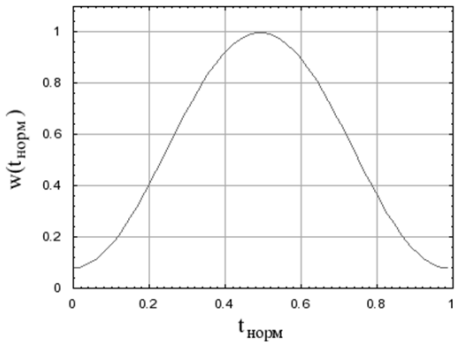

Рис. 3. Функция Хэмминга

Эта функция плавно сводит на нет сигнал вблизи краев анализируемого участка. Далее, выполняется быстрое преобразование Фурье (FFT) сигнала, в основу которого положен алгоритм Кули-Тьюки с основанием 2. На следующем этапе мы конвертируем частоты полученные быстрым преобразованием Фурье в мел частоты. Переход от обычной частоты (Гц) к мел-частоте выполняется по следующей формуле:

$$m = 1127 \ln\left(1 + \frac{f}{700}\right) \tag{2}$$

где m – частота в мелах, f – частота в герцах.

График функции перехода от частоты в герцах к частоте в мелах изображен на (рис. 4)

Затем мы выполняем расчёт mel-фильтров (от др.-греч. $\mu\varepsilon\lambda o\varsigma$ — звук) т.е. переходим к психофизической единице высоты звука, основанной на субъективном восприятии среднестатистическими людьми. Здесь неявно делается предположение, что голосовой аппарат человека приспособлен к его слуховому аппарату, т.е. наиболее важные идентифицирующие голос признаки следует искать по мел-частотной шкале, воспринимаемой ухом.

Это предположение, вообще говоря, не очевидно и требует экспериментальной проверки, которая была проведена в рамках представляемой работы. Т.е. была выполнена идентификация с целью различения мужского и женского голосов по мел-кепстральным коэффициентам (вектору признаков сигнала **C**), которые находятся на последнем этапе метода [9] по формуле:

$$C_n = \sum_{k=1}^{K} (\log S_k)\, [n\left(k - \frac{1}{2}\right)\frac{\pi}{K}] \tag{3}$$

здесь: C_n – мел-кепстральный коэффициент под номером n, S_k – мел-коэффициент под номером κ, K – наперед заданное количество мел-кепстральных коэффициентов,$n \in [1, K]$, а также по вектору признаков **E**, полученному из анализа энергетического распределения спектра сигнала по обычной (Гц) шкале по формуле:

$$E_i \sim \sum_{n_i}^{n_{i+1}} X_k^2 \tag{4}$$

где

$$i \in (1, K)$$

Для определения эффективности идентификации пола говорящего указанными способами статистический анализ и дальнейшая обработка данных были сведены к минимуму. Эффективность определялась методом городских кварталов «манхэттенское расстояние» [10], т.е. вычислением расстояния от вектора признаков пробного сигнала **p** до вектора признаков **q** вектора шаблона:

$$d_1(p, q) = \|p - q\|_1 = \sum_{i=1}^{n} |p_i - q_i| \tag{5}$$

а также непосредственным сравнением компонент этих векторов после каждого эксперимента.

$$Similarity = \sum_{i=1}^{M} 1, i \in \{i : \frac{min(p_i, q_i)}{max(p_i, q_i)} < 0,5\} \tag{6}$$

Таблица 1. Таблица вероятности правильной идентификации мужского и женского глоса.

Вид сигнала, метод	Город. кварт	Similarity
звук, **E**	90%	90%
слово, **E**	70%	70%
звук, **C**	90%	70%
слово, **C**	85%	70%

2. Результаты

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

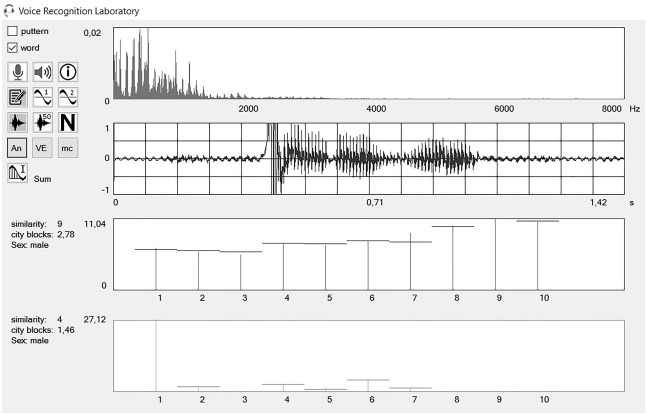

Рис. 5. Интерфейс программы распознавания голоса.

Графики n эелементов векторов энергетического распределения **E** спектра шаблонного и анализируемого голосовых сигналов представлены на (рис. 6)

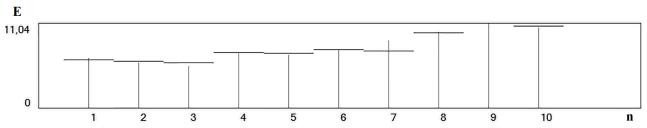

Рис. 6. Графики n эелементов векторов энергетического распределения **E** спектра голосовых сигналов. Горизонтальными линиями обозначены значения элементов вектора-шаблона, вертикальные линии отображают значения вектора анализируемого сигнала.

Графики векторов-признаков **C**, составленных из n мел-кепстральных коэффициентов шаблонного и анализируемого голосовых сигналов представлены на (рис. 7)

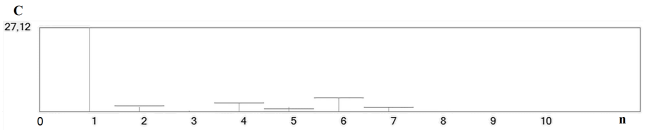

Рис. 7. Графики векторов-признаков **C**, составленных из n мел-кепстральных коэффициентов шаблонного и анализируемого голосовых сигналов. Горизонтальными линиями обозначены значения элементов вектора-шаблона, вертикальные линии отображают значения вектора анализируемого сигнала.

Результаты серии эксперементов по определению вероятности правильного распознавания мужского и женского голосов приведены в (табл. 1).

3. Выводы

В результате было установлено, что вероятность правильной идентификации, без использования каких бы то ни было тренировочных алгоритмов кластеризации, методом энергетического распределения спектра по обычной шкале не уступает методу мел-кепстральных коэффициентов по мел-шкале, в случае если анализируются отдельные гласные звуки, и составляет 90%. При анализе слов, вероятно, сказывается различная акустическая энергетика одних и тех же слов у разных людей одного пола, поэтому метод мел-кепстральных коэффициентов даёт лучшее значение вероятности идентификации и составляет величину 85% против 70% по энергетическому распределению.

Перечень использованых источников

1. Rabiner L., Juang B. Fundamental of Speech Recognition. — Englewood Cliffs : Prentice-Hall N.J., 1993.
2. 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 : M. Rosenblatt, Ed, 1963. — P. 209–243.
3. Оппенгейм А. В., Шафер Р. В. Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. — М. : Связь, 1979.
4. 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. 1. — P. 149 — 154.
5. Clement I. Modular Audio Recognition Framework v.0.3.0.6 (0.3.0 final) and its Applications. — Quebec, Canada : Montreal, 2007.
6. Добрушкин Г. О., Данилов В. Я. Применение вейвлет-преобразования для сегментации и удаления шума с речевых сигналов // Научные вести НТУУ КПИ. — 2009. — Т. 1. — С. 34–42.
7. Добрушкин Г. О., Данилов В. Я. Сопоставление качества Мел- и Барк- частотных кепстральных коэффициентов для параметризации речевого сигнала // Научные работы. — 2011. — Т. 160. — С. 167–171.
8. Добрушкин Г. О., Данилов В. Я. Основные подходы к распознаванию речевой информации // Вестник Винницкого политехнического института. — 2010. — С. 61–73.
9. Speaker identification using mel frequency cepstral coefficients / R. Hasan, M. Jamil, Rabbani G., Rahman S. // 3rd International Conference on Electrical and Computer Engineering. — 2004. — P. 28–30.
10. Krause E. F. Taxicab Geometry. — Dover, 1987.