

УДК 004.41/42(075.8)
ББК 32.973-018 я7

Т. И. Данильчук
*Байкальский государственный университет
экономики и права*
И. М. Трофимов
*Белгородский государственный технологический
университет им. В. Г. Шухова*

ОБРАБОТКА И АНАЛИЗ РЕЧЕВЫХ ФРАЗ С ИСПОЛЬЗОВАНИЕМ МАТЕМАТИЧЕСКИХ АЛГОРИТМОВ УРОВНЕЙ СИГНАЛОВ

Актуальность темы определяется необходимостью разработки практических систем автоматического распознавания и понимания речевых сигналов при изучении иностранных языков. Рассматриваются вопросы анализа распознавания речевой информации с помощью систем управления. В поисках решения данной проблемы исследуется спектральное представление речи. Приведены особенности устройства речевого аппарата, которые должны быть учтены при создании алгоритма распознавания речи и выдачи рекомендаций по улучшению произношения. Для преобразования речевого акустического сигнала в цепочку символов и слов предлагается алгоритм анализа слова и фразы. Анализируются принципы работы современных систем распознавания речи. Предложено описание структуры модели определения правильности произношения. Обработка речевых фраз с использованием предложенных математических алгоритмов уровней сигналов поможет определить степень влияния индивидуальных особенностей устройства речевого аппарата.

Ключевые слова: математические методы; программирование; распознавание речи; голосовой аппарат; анализ речи; анализ сигнала.

T. I. Danilchuk
Baikal State University of Economics and Law
I. M. Trofimov
Belgorod State Technical University named after V.G. Shukhov

USING MATHEMATICAL ALGORITHMS OF SIGNAL LEVELS FOR PROCESSING AND ANALYSIS OF SPEECH

Relevance of the topic is determined by the need to develop a practical system of automatic recognition and interpretation of speech signals in studying foreign languages. The paper addresses the problems of spoken language recognition analysis with the help of control systems that use a system of automatic recognition. To solve these problems, the authors study spectral representation of speech. They also describe peculiarities of the organ of speech that must be taken into consideration when creating a speech recognition algorithm and developing recommendations for improving pronunciation. To convert the voice acoustic signal into a chain of symbols and words, the authors propose an algorithm of word and phrase analysis. The article also analyzes working principles of contemporary speech recognition systems, and describes the structure of a model of determining correctness of pronunciation. Thus, speech processing by means of the proposed mathematical algorithms of signal levels will help to see the extent of the impact of individual peculiarities of the organ of speech.

Keywords: mathematical methods; programming; speech recognition; organ of speech; speech analysis; signal analysis.

В современных условиях информационного общества с каждым днем все более актуальным становится использование речевых технологий, таких как распознавание, голосовое управление сложными техническими системами, а

также автоматизированная постановка произношения. Существующие сегодня системы распознавания речи основываются на сборе всей доступной информации, необходимой для распознавания слов. Исследователи считают, что, таким образом, задача распознавания образца речи, основанная на качестве сигнала, подверженного изменениям, будет достаточной для распознавания. Однако, в настоящее время, даже при распознавании небольших сообщений нормальной речи, невозможно после получения разнообразных реальных сигналов осуществить прямую трансформацию в лингвистические символы. Хотя спектральное представление речи важно, необходимо помнить, что изучаемый сигнал уникален. Уникальность возникает по многим причинам:

- различие человеческих голосов;
- громкость речи;
- стилистические особенности произношения;
- нормальное варьирование движения артикуляторов (языка, губ, челюсти, неба).

Большинство программ по изучению иностранного языка с возможностью проверки произношения основывается на сравнении эталона с речью диктора. Однако, как было выяснено А. Нери, основной недостаток систем, построенных по принципу сравнения с эталоном, состоит в том, что, даже при очень хорошем произношении, речь обучаемого и эталон могут иметь совершенно разные спектры или формы во времени [3]. Это происходит из-за особенностей устройства речевого аппарата, который состоит из двух отделов: центрального и периферического. Центральный — это головной мозг с его корой, подкорковыми узлами, проводящими путями и ядрами соответствующих нервов. Центральный отдел не влияет на саму звуковую волну, получающуюся при самой речи, в отличие от периферического. Периферический отдел — это вся совокупность исполнительных органов речи.

Периферический речевой аппарат состоит из трех основных отделов, которые действуют совокупно:

1. Первый отдел — дыхательные органы. Все звуки речи образуются только при выдохе.
2. Второй отдел — пассивные органы речи (зубы, альвеолы, твердое небо, глотка, полость носа, гортань). Они оказывают наибольшее влияние на технику речи.
3. Третий отдел — активные органы речи (язык, губы, мягкое небо, маленький язычок, надгортанник, голосовые связки) [2].

Поэтому при создании алгоритма распознавания речи и выдачи рекомендации по улучшению произношения были учтены данные особенности. Разрабатываемая структура модели определения правильности произношения будет состоять из модулей: база слов, база советов, модуль анализа произношения, модуль генерации совета.

В базе слов будут содержаться эталонная запись слова или фразы и наборы характеристик слова, соответствующие правильному и неправильному произношению. Все слова будут разделены по группам. База советов, будет состоять из предложений, выдаваемых пользователю по улучшению произношения. Модуль анализа произношения будет состоять из двух частей: алгоритм анализа слова и алгоритм анализа фразы.

Так, при распознавании слова будет использован следующий алгоритм:

1. Определяется, к какой группе относится слово, для каждой группы характерны свои собственные корректирующие характеристики.
2. Вычисляются значения характеристик слова: общие и корректирующие.

3. Формируется набор характеристик слова.
 4. Полученный набор отправляется в модуль генерации совета.
- Алгоритм при распознавании фразы:
1. Определяется, к какой группе относится фраза.
 2. Вычисляются значения характеристик фразы.
 3. Фразы разбиваются на отдельные слова.
 4. Анализируется каждое слово.
 5. Формируется набор характеристик фразы.
 6. Полученный набор отправляется в модуль генерации совета.

Рассмотрим, как происходит разбиение фразы на отдельные слова. Современные методы выделения слов могут точно определить начальную и конечную точки произнесенного слова в звуковом потоке, основываясь на выделении пауз (относительный ноль). Данные методы оценивают энергию и среднюю величину в коротком отрезке времени, а также вычисляют относительный ноль.

Сначала происходит создание начальной и конечной точек (если аудиозапись сделана в идеальных условиях). В этом случае отношение сигнал — шум велик, таким образом определить слово не составляет особого труда. В реальных условиях при записи накладывается фоновый шум, который имеет большую интенсивность и может нарушить процесс выделения слов в потоке речи. В таких случаях для выделения слова используются алгоритмы уровня сигнала.

Наиболее оптимальный алгоритм отделения слов — алгоритм Рабинера-Ламеля [1]. В нем используется понятие строб-импульсов — сигналов, подаваемых на логическую схему, чтобы управлять выходом логической схемы таким образом, что выход может быть либо включен, либо выключен. Составляется набор $\{s_1, s_2, \dots, s_n\}$, где n — количество строб-импульсов, s_1 — числовое значение, общая энергия находится по формуле:

$$E(n) = \frac{1}{n} \sum_{i=1}^n s_i^2.$$

Средний уровень пересечения нулевого уровня:

$$ZCR(n) = \sum_{i=1}^{n-1} \text{sign}(s_i) \text{sign}(s_{i+1}),$$

где

$$\text{sign}(s_i) = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{if } s_i < 0 \end{cases}$$

Метод использует три числовых уровня: два для энергии (верхний, нижний) и один для среднего пересечения нулевого уровня¹. Точка, начиная с которой энергия перекрывает верхний уровень и уровень положительных и отрицательных значений, не отменяет установленный уровень, который считается отправной точкой голосового звучания. Поиск первой такой точки производится путем скрещивания импульсов от начала и до конца, что определит первую область с речью. Проход в обратном направлении позволит определить конечную точку последней области с речью. Для дальнейшего разбиения фразы на слова данный алгоритм используется еще раз, но не на всей записи, а на выбранном фрагменте, полученном при первом применении алгоритма. Начало области (без речи) начинается в точке, в которой энергия становится меньше значения нижнего уровня.

¹ URL: <http://revistaie.ase.ro/content/46/s%20-%20furtuna.pdf>.

Таким образом, современные методы позволяют точно определить начальную и конечную точки произнесенного слова. Вместе с тем, существуют проблемы при распознавании коротких сообщений нормальной речи, поскольку очень велика степень влияния индивидуальных особенностей устройства речевого аппарата. Поэтому, основной задачей распознавания речи является абстрагирование от индивидуальных особенностей. В то же время, для определения правильности произношения, необходимо учитывать и анализировать индивидуальные особенности строения речевого аппарата, чтобы понять, как они влияют на правильность произношения.

Список использованной литературы

1. Рабинер Р. Л. Цифровая обработка речевых сигналов / Р. Л. Рабинер, Р. В. Шафер ; пер. с англ. под ред. М. В. Назарова, Ю. Н. Прохорова. — М. : Радио и связь, 1981. — 496 с.
2. Филичева Т. Б. Основы логопедии : учеб. пособие / Т. Б. Филичева, Н. А. Чевелева, Г. В. Чиркина. — М. : Просвещение, 1989. — 223 с.
3. Фролов А. В. Синтез и распознавание речи. Современные решения / А. В. Фролов, Г. В. Фролов. — М. : Связь, 2003. — 216 с.

References

1. Rabiner L. R., Schafer R. W. *Digital processing of speech signals*. Englewood Cliffs, New Jersey, Prentice-Hall, 1978. 512 p. (Russ. ed.: Rabiner R. L., Shafer R. V. *Tsifrovaya obrabotka rechevykh signalov*. Moscow, Radio and Svyaz Publ., 1981. 496 p.)
2. Filicheva T. B., Cheveleva N. A., Chirкина G. V. *Osnovy logopedii* [The basics of speech-language therapy]. Moscow, Prosveshchenie Publ., 1989. 223 p.
3. Frolov A. V., Frolov G. V. *Sintez i raspoznavanie rechi. Sovremennye resheniya* [Synthesis and speech recognition. Current solutions]. Moscow, Svyaz Publ., 2003. 216 p.

Информация об авторах

Данильчук Татьяна Ивановна — старший преподаватель, кафедра экономики труда и управления персоналом, Байкальский государственный университет экономики и права, 664003, г. Иркутск, ул. Ленина, 11, e-mail: danilchuk-ti@isea.ru.

Трофимов Иван Михайлович — аспирант, кафедра программного обеспечения вычислительной техники и автоматизированных систем, Белгородский государственный технологический университет им. В. Г. Шухова, 308012, г. Белгород, ул. Костюкова, 46, e-mail: trofimov.ivan.m@gmail.com.

Authors

Danilchuk Tatiana Ivanovna — Senior Instructor, Dep-t of Labour Economics and HR Management, Baikal State University of Economics and Law, 11 Lenin St., 664003, Irkutsk, Russia, e-mail: danilchuk-ti@isea.ru.

Trofimov Ivan Mihailovich — PhD student, Dep-t of Computer Software and Automated Systems, Belgorod State Technical University named after V. G. Shukhov, 46 Kostyukov st., 308012, Belgorod, Russia, e-mail: trofimov.ivan.m@gmail.com.