**УДК 004.934.1’1**

***Саввина Г.В.***
Донецкий государственный институт искусственного интеллекта, Украина

# Система распознавания изолированных слов с предварительной сегментацией

Статья посвящена описанию системы пофонемного распознавания изолированных слов. В процессе выполнения работы был предложен способ повышения надёжности системы пофонемного распознавания изолированных слов. Основные преимущества использованного подхода – сокращение числа возможных границ фонем и определение длины участка речи, относимого к текущей или следующей за ней фонеме, исходя из рассматриваемого речевого сигнала.

## Введение

Создание средств человеко-машинного общения является важной практической задачей, на решение которой направлены усилия множества специалистов по распознаванию речи, моделированию процессов речеобразования и восприятия речи человеком.

Задача дикторозависимого распознавания изолированных слов малого словаря к настоящему времени достаточно хорошо разработана во всём мире. Основной распознаваемой единицей является слово или фраза целиком. При этом по некоторой реализации слова строится вектор (или фиксированный по объему набор векторов) признаков, характеризующих это слово, который принимается за эталон. Набор эталонов, соответствующих словам, которые должны распознаваться системой, составляет словарь эталонов. В процессе распознавания исследуемому речевому сигналу (РС) ставится в соответствие набор векторов признаков, а затем в словаре эталонов методом динамической временной деформации (ДВД) [1] определяется эталон, который наиболее близок к этому набору.

Распознавание слов среднего и большого словаря, при котором слова рассматриваются целиком (а не как составные единицы), проблематично ввиду трудоёмкости создания словаря эталонов и больших затрат времени на распознавание. При пофонемном распознавании слово рассматривается как последовательность фонем. Преимущество такого подхода заключается в том, что словарь можно задавать в текстовом виде, не задавая при этом голосовые эталоны всех слов словаря. Обучение системы состоит в обучении соответствующего набора фонем, распознавание слов можно производить не целиком, а по частям, соответствующим фонемам, причем для всех одинаково начинающихся слов нужно производить только одну обработку их общей части.

Одна из последних разработок отдела распознавания речевых образов Института проблем искусственного интеллекта (РРО ИПИИ) – создание лабораторной системы пофонемного распознавания изолированных слов большого словаря

для русского языка [2]. Эта система позволяет с минимальными затратами времени и труда производить обучение системы распознавания, выполнять распознавание в режиме реального времени, динамически менять словарь распознавания. Упомянутая система в процессе распознавания определяет границы фонем, составляющих слово, и сопоставляет с их эталонами соответствующие участки речи. Недостатком указанной системы является недостаточно высокая вероятность правильного распознавания. Случаи неверного распознавания связаны с неверным определением границ фонем.

Настоящая работа ставит своей целью повышение точности распознавания системы пофонемного распознавания изолированных слов за счёт предварительной сегментации речевого сигнала и использования признаков, присущих отдельным сегментам.

## Система пофонемного распознавания

Схематично процесс распознавания РС системой можно описать следующим образом. Оцифрованный с частотой дискретизации 22050 Гц и разрядностью 8 бит РС подвергается предварительной обработке, результатом которой является последовательность векторов признаков (реализация). В качестве вектора признаков используется вектор кумулятивного отношения [3]. Реализация пофонемно сопоставляется со всеми словами словаря, по минимуму меры близости между реализацией и словом словаря принимается решение о произнесённом слове. Опишем основные принципы системы, благодаря которым осуществляется пофонемное распознавание.

Слова словаря представляют последовательностью фонем. Считается, что задан алфавит фонем. Каждой из фонем ставится в соответствие один или несколько векторов признаков, именуемых эталонами фонемы или кодовыми векторами. Совокупность эталонов всех фонем образует кодовую книгу. Определение границ фонем слова в реализации сводится к последовательному определению границ между парами фонем слова (первой и второй, второй и третьей фонемами и т.д.).

Моментом *перехода* от одной фонемы к другой считают момент, когда два подряд идущих вектора признаков окажутся ближе к эталону следующей фонемы, чем к эталону данной. Минимальная длина фонемы считается равной длине трех векторов.

Мера близости между словом и реализацией определяется как сумма мер близости между фонемами слова и соответствующими им участками реализации. Мера близости между фонемой и соответствующим ей участком реализации определяется как сумма евклидовых расстояний между векторами признаков, входящими в этот участок, и ближайшими к ним эталонами рассматриваемой фонемы.

Слово с минимальной мерой близости до реализации, для которого найдены границы всех фонем, считается результатом распознавания.

Попытаемся определить причины ошибочной классификации РС алгоритмом пофонемного распознавания. Одна из них обусловлена большим разнообразием и изменчивостью РС. Поэтому отрезок РС, соответствующий двум последовательно идущим векторам признаков и принадлежащий рассматриваемой фонеме, может быть ошибочно отнесён к следующей, и наоборот отрезок РС,

соответствующий двум последовательно идущим векторам признаков и принадлежащий следующей фонеме, может быть отнесён к рассматриваемой. В этом случае ошибочное срабатывание (несрабатывание) правила условной сегментации приведёт к ошибочной классификации РС.

Вторая причина связана с принудительным заданием минимальной длины фонемы – 3 вектора признаков для всех типов фонем, тогда как известно, что длительность гласных может многократно превышать длительность согласных. Известно, что эта величина более чем в два раза превышает минимальную длительность фонемы «*р*». При быстром темпе произнесения длительность ряда других согласных и даже некоторых гласных в безударных положениях недостаточна для срабатывания правила условной сегментации. Попытки же заменить правило изменения минимальной длины фонемы в сторону уменьшения приводили к возрастанию числа ошибок распознавания, что связано с изменчивостью РС.

Таким образом, для решения задачи повышения надёжности распознавания необходимо найти и реализовать средства, препятствующие ошибочному срабатыванию правила условной сегментации, а также средства, определяющие минимальную длину сегмента РС, относимого к одной фонеме, исходя из распознаваемого РС.

## Предварительная сегментация речевого сигнала

В качестве средства, позволяющего решить выявленные проблемы, выбрана предварительная сегментация РС.

В смежной с распознаванием речи области – распознавании изображений – *сегментация* рассматривается как этап обработки, предшествующий непосредственно распознаванию. Под термином «*сегментация*» понимается разделение изображения на несколько областей или зон, которые отличаются друг от друга элементарными признаками и считаются однородными. При распознавании речи же под *сегментацией* подразумевается разбиение речевого сигнала на участки, соответствующие фонемам. А они отнюдь не всегда являются однородными участками. При описании предварительной сегментации под термином *сегментация* будем понимать разбиение речевого сигнала на однородные непересекающиеся участки, такие, что один или несколько соседних сегментов образуют фонему, один сегмент не может соответствовать более чем одной фонеме. Цель предварительной сегментации – разбиение РС на участки, относительно каждого из которых принимается решение о принадлежности к текущей или следующей фонеме.

В основу разрабатываемых методов сегментации положено предположение о том, что речь – это последовательность звуковых кодов. Кодирование производится за счёт изменения спектральных параметров и громкости звука, которые вызваны изменением положения артикуляторных органов. То есть коды не абсолютны, а относительны. Речь есть блочный код. Минимальной единицей информации (блоком), для которой возможна расшифровка, является слово.

Поскольку вектор кумулятивного отношения отражает относительные значения спектра РС, можно предположить, что в моменты, соответствующие смене положения артикуляторных органов, происходят максимальные изменения значений векторов признаков. Предварительная сегментация получается следующим

образом. Вычисляем расстояния между соседними векторами признаков. Находим все локальные максимумы полученного ряда чисел. Они и определяют границы искомых сегментов.

Обычно гласные разбиваются на 3 – 5 сегментов; сонорным согласным и паузе часто соответствует один сегмент. Предложенный способ предварительной сегментации был протестирован на банке фонетически размеченных слов, произнесённых 11 дикторами, с общим числом слов 1211. В 96,9 % случаев сегментация признана допустимой для дальнейшего распознавания.

# Заключение

В настоящей работе предложен способ повышения надёжности системы пофонемного распознавания изолированных слов [2] за счёт сокращения числа возможных границ фонем и определения длины участка речи, относимого к текущей или следующей за ней фонеме, исходя из рассматриваемого речевого сигнала.

Одна из особенностей работы по сравнению с ранее разрабатывавшимися в отделе РРО ИПИИ подходами к сегментации РС – стремление отказаться от наперёд заданных значений пороговых величин (длины сегмента). Вместо этого предполагается получать эти значения из РС, который распознаётся. Кроме того, данный подход позволяет оценить сверху число фонем в слове.

# Литература

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наук. думка, 1987. – 262 с.
2. Козлов А.В., Саввина Г.В., Шелепов В.Ю. Система пофонемного распознавания отдельно произносимых слов // Искусственный интеллект. – 2003. – № 1.
3. Старушко Д.Г. Вычисление кумулятивного отношения на основе быстрого преобразования Хартли и нейросетевое распознавание речевых единиц с его использованием. Быстрое вычисление кумулятивного отношения // Мат-лы Междунар. науч.-техн. конф. «Искусственный интеллект – 2002». – Т. 2. – Таганрог: Изд-во ТРТУ. – 2002. – С. 327-329.

The article is devoted to isolated word recognition system description. The system is based on phoneme recognition. There was the way of isolated word recognition system improving proposed in this work. The advantages of proposed method are: reduction of count of possible phomeme boundaries and determining portions of signal which belong to current phoneme or to the next one depending on speech signal.

Стаття присвячена опису системи пофонемного розпізнавання ізольованих слів. У процесі виконання роботи запропоновано спосіб підвищення надійності системи пофонемного розпізнавання ізольованих слів. Основні переваги використаного підходу – скорочення кількості можливих границь фонем і визначення довжини ділянки мовлення, що має бути віднесена до поточної або наступної фонеми, виходячи з мовленнєвого сигналу, що розглядається.

*Статья поступила в редакцию 28.07.03.*