ПРОБЛЕМЫ СОЗДАНИЯ МНОГОУРОВНЕВОЙ СИСТЕМЫ РАСПОЗНОВАНИИ РЕЧИ

Курочкин С.Н
(Москва, МГГУ СТАНКИН)
Бродин А.Г.
(Москва МГТУ СТАНКИН)

В современных компьютерных системах все больше внимания уделяют построению интерфейса естественным вводом-выводом информации (распознавание рукописного текста, речевой диалог).

Наиболее перспективными на сегодняшний день являются системы речевого ввода. Задачу распознавания речевой информации можно разделить на две большие подзадачи:

1. Непосредственное распознавание отдельных слов.
2. Распознание смысловой нагрузки слов.

Непосредственное распознавание отдельных слов осложняется рядом факторов: различием языков, спецификой произношения, шумами, акцентами, ударениями и т. п.

В настоящее время можно выделить два основных направления при построении систем распознавания речи:

1. Эталонный - данный метод основан на сравнении некоторых характеристик речи (энергетических, спектральных и т.п.). В качестве эталонов в большинстве случаев используют целые слове. Данный метод удобен для использования в системах с ограниченным словарем (например, для ввода небольшого набора команд).
2. Фонемно-ориентированный метод. Основан на выделении фонем из потока речи. Фонема это единица речи представляющая собой единицу речи, Подобно тому, как слово состоит из букв, так и речь состоит из фонем. Для каждого языка имеется свой конечный набор фонем.

Сравнивая распознавание речевого потока методом распознавания целых слов и распознавание фонем можно сделать вывод: при небольшом количестве слов, используемых оператором более высокую надежность и скорость можно ожидать от распознавания целых слов, но при увеличении словаря скорость резко падает. Предположительно, размер словаря системы распознавания уже в сотню слов делает переход на уровень более низкий, чем распознавание слов в целом, актуальным.

Рассмотрим модель построения системы распознавания речи построенной на фонемно-ориектированном методе (Рис.1).

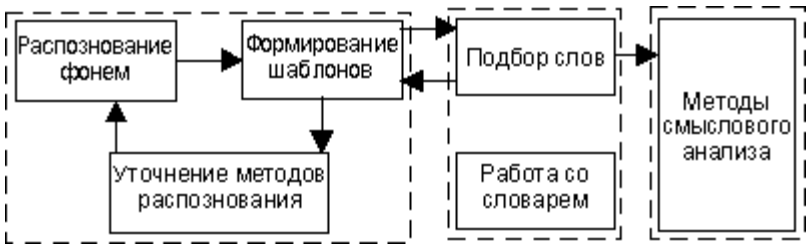


Рис. 1 Построение системы распознавания речи

Из списка фонем распознанных с определенной точностью, составляется шаблон, который передается на следующий уровень, где по нему происходит подбор наиболее подходящего слова, передача информации о выборе на более высокий уровень для дальнейшего анализа и на нижний, для подстройки системы на конкретного пользователя. Достоинством этой схемы является высокая адаптивность, дающая возможность динамической самоподстройки системы на оператора, и многоуровневая система проверок, повышающая точность работы.

Проанализируем возможные механизмы распознавания фонем. Звуки, участвующие в формировании речи, имеют две основных классификации: по артикуляционным признакам и по акустическим признакам.

Классификация звуков по артикуляционным признакам является крайне важным при

использовании методов генерации и распознавания речи с помощью моделирования носоглотки, но для решения задач деления на фонемы более интересно рассмотрение акустических различий звуков. По акустическим признакам звуки подразделяются:

Тональные звуки - образуются голосом при почти полном отсутствии шумов, что обеспечивает хорошую
слышимость звуков:

гласные а, э, и, о, у, ы.

Сонорные (звучные) - чье качество определяется характером звучания голоса, который играет главную роль в их образовании, а шум участвует в минимальной степени:

согласные м, м', н, н', л, л', р, р'.

Шумные - их качество определяется характером шума:

звонкие шумные длительные: в, в, з, з, ж;
звонкие шумные мгновенные: б, б’, д, д', г, г';
глухие шумные длительные: ф, ф', с, с', ш, х, х';
глухие шумные мгновенные: п, п', т, т', к, к'.

Заметим, что гласные и сонорные звуки состоят из участков затухания импульсов от основных (не обертонных) колебаний истинных голосовых связок. Для упрощения, будем называть эти участки доменами.

Использование домен при распознавании речи вполне очевидно. По сути, домен (вспомним, что пока домен рассматривается в приложении только к сонорным и гласным звукам) содержит в себе информацию достаточную для распознавания звука. Если взглянуть на образ протяженно произнесенной гласной (или сонорного звука), то за исключением небольших по длине участков в начале и конце образа звук состоит из домен с высокой степенью идентичностью, даже для различных людей многие характеристики, а соответственно и общий вид домен во многом схожи, что придает особую универсальность методам распознавания при выделении и распознавании фонем через домены. Еще одним достоинством домен является относительная простота их выделения. По определению, домен начинается с максимального значения в определенном диапазоне, после которого идет затухающий по некоторому закону колебательный процесс. Как дополнительные условия, которые можно использовать при расчленении речи на домены, можно перечислить:

- стабильную (в диапазоне) длину домен;
- постоянную, с некоторой точностью, величину максимумов, по которым происходило вычленение домен.

Дополнительно будем рассматривать шумные длительные звуки как один домен. Это позволит легко выделять корень этих звуков из общего потока и облегчит их анализ.

Анализ образов шумных мгновенных (взрывных) звуков показывает наличие участков по структуре схожих с определенным для гласных и сонорных звуков понятием домена. Но наряду с совокупностью общих признаков прослеживается различие: для вышесказанных участков в шумных мгновенных звуках отсутствует та строгая идентичность домен между собой. Во всех мгновенных звуках присутствует момент, сильно облегчающий их выделение из речи - перед произнесением таких звуков наблюдается непродолжительная по меркам восприятия, но весьма значительная, в масштабах длительностей домен, пауза. Эго помогает выделению домен. Поэтому в зависимости от различных алгоритмов выделения может быть удобно, разбивать такого рода звуки на несколько домен, или же воспринимать их целиком как один.

При разбиении потока речи на домены мы получаем еще один уровень в распознавании. В общей иерархии он находится еще ниже, чем уровень распознавания фонем . Рассмотрим функционирование такой системы (Рис.2).

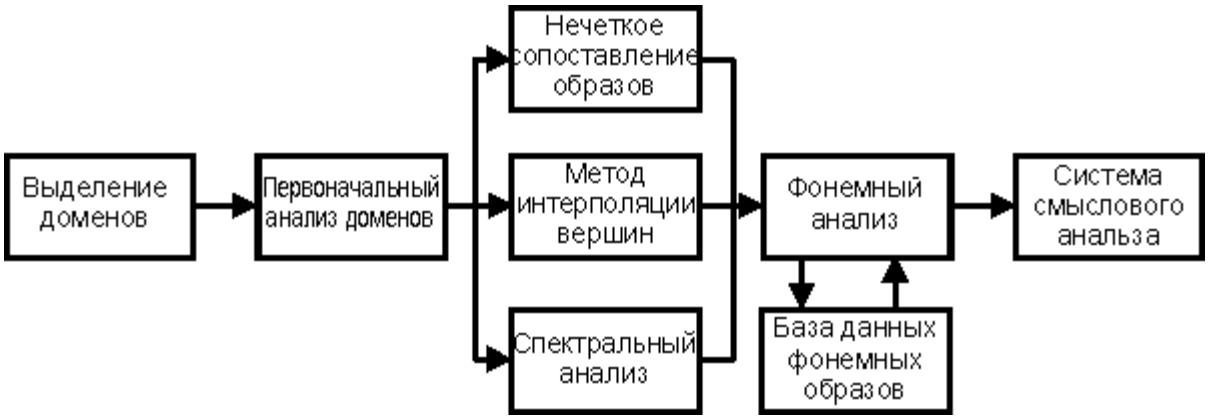


Рис. 2 Использование доменов в системе распознавания речи

Первоначально производится деление потока речи на домены, используя такие свойства доменов как, стабильная длинна на протяжении одной фонемы и большую амплитуду первого колебания в домене.

В дальнейшем происходит первичный анализ домена для определения методов его дальнейшей обработки. Эти методы различны для тональных, сонорных и шумных звуков. На втором этапе также производится выделение отдельных слов слитной речи.

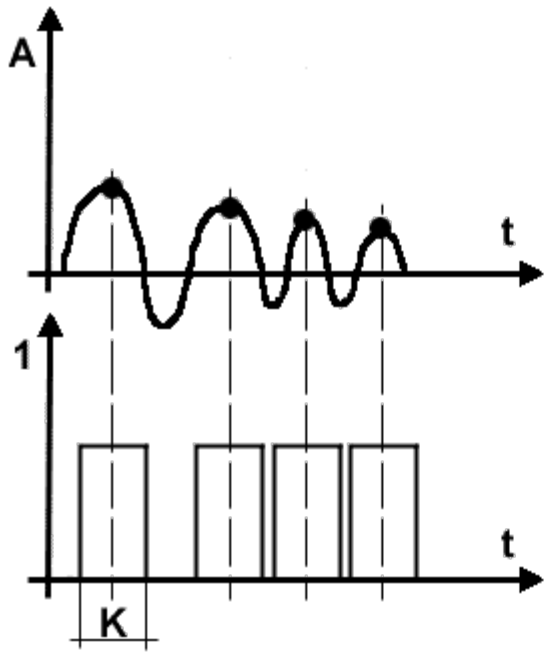
Подробнее остановимся на методах анализа домен. Целесообразно производить такой анализ в несколько этапов с постепенным уточнением результата:

1. Простейшими методами определяем диапазон возможных значений.
2. Более сложными методами анализа определяем вероятность принадлежности данного домена к различным фонемам из ранее определенного диапазона.

Для этой цели были разработаны несколько методов.

Метод нечеткого сопоставления образов при разработке данного метода была использована теория нечеткой логики. Суть метода состоит в следующем: на основе статистических данных составляется двоичный образ доменов для каждой фонемы [1].

Двоичный образ представляет сбой карту локальных выбросов в домене по амплитуде. При этом учитывается лишь местоположение выброса на временном диапазоне, величина амплитуды значения не имеет.



$$M = K\frac{(a1 + a2 + \ldots + an)}{n}$$

$$an = \begin{cases} 1 \\ 0 \end{cases}$$

M - вероятность совпадения
K - коэф. точности

Рис.3 Использование функции принадлежности

Используя функцию принадлежности можно получить вероятность идентичности анализируемого домена и двоичного образа.

Анализ доменов на основе интерполяции вершин. Вид кривой проведенной по вершинам доменов аналогичен для всех доменов данной фонемы и мало различается для различных людей, а также для разных условий произнесения [2]. Первый этап - построение интерполяционного многочлена Тейлора по вершинам домена включает в себя:

1. выборку вершин, т.е. положительных экстремумов домена;
2. расчет коэффициентов;
3. построение многочлена.

Порядок многочлена задается числом вершин данного домена. Получив функцию, записанную в виде многочлена Тейлора, приступаем к ее анализу (Рис.4).

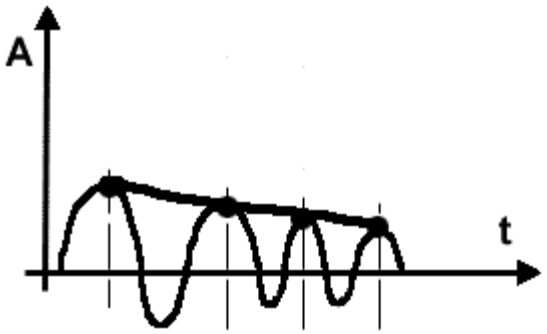


Рис.4 Интерполяция вершин.

Анализ по соотношениям значений функции относительно первого максимума данного домена совместно с анализом по знакам первых производных в наборе точек позволяет оценить общий вид функции и является универсальным, сочетая в себе надежность и гибкость.

Используя комбинацию данных методов можно с высокой точностью определить набор фонем для передачи на следующий уровень системы. С каждой фонемой на верхний уровень передается вероятность ее правильного определения.

Используя эти данные, формируется набор слов для последующей передачи на уровень смыслового анализа.

Предложенная система была частично реализована в опытном программном продукте для анализа свойств доменов и показала свою жизнеспособность. По нашему мнению, использование домен позволит создавать не ресурсоемкие универсальные системы распознавания речи.

ЛИТЕРАТУРА:

1. Киедзи Асаи, Дзюндзо Ватада, Сокуке Иваи и др. Прикладные нечеткие системы. Под редакцией Т.Тэрано, К. Асаи, М. Сугено. Издательсгво 'Мир' Москва 1993г.
2. Л. Рабинер, Б. Гоулд. Теория и применение цифровой обработки сигналов. Издательство 'Мир' Москва 1978г.