УДК 007.52

ПРОБЛЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ И ПРЕДЛАГАЕМЫЙ ПУТЬ ИХ РЕШЕНИЯ

Антипин А. Ф., Шишкина А. Ф.
Стерлитамакская государственная педагогическая академия им. Зайнаб Биишевой

Работа с текстом зачастую предполагает ручной ввод с клавиатуры, при этом средняя скорость набора обычного человека составляет 100-120 знаков в минуту, что соответствует 1-2 предложениям. В связи с этим, трудно переоценить труд стенографистов и наборщиков текстов, чья скорость работы выше в 3-4 раза. Однако в эпоху стремительного развития средств вычислительной техники, Интернет- и мультимедиа-технологий возникает острая необходимость в технологиях для обработки аудио- и видеоинформации, распознавания речи, мимики и жестов, позволяющих сократить до минимума долю ручного труда, возложив всю рутинную работу на электронно-вычислительные машины. Кроме того, подобные технологии существенно повысят скорость обмена информацией между человеком и машиной, а, следовательно, и между людьми, находящимися в разных точках земного шара.

Наиболее известной российской компанией, специализирующейся на разработке и создании инновационных систем в области обработки и анализа аудиоинформации, распознавания речи и смежных с ними областей, является ООО "Центр речевых технологий" – компания с более чем 20-летней историей. Несмотря на то, что данная компания является весомым игроком международного рынка речевых технологий, она все же не гарантирует полноценного распознавания слитной речи для своих программных продуктов.

Наиболее известные российские системы и программы распознавания речи, такие как «Цезарь», «Нестор», «Горыныч» и другие, далеки от совершенства, особенно при распознавании слитной русской речи в условиях повышенного уровня шума (на улице, в цехах и т.д.). Поэтому в большинстве случаев подобные программы используются в процессе подготовки протоколов и стенограмм совещаний, переговоров и т.д., хотя потенциальная сфера применения данной технологии очень высока. Здесь необходимо отметить, что в настоящее время удачными, но обеспечивающими далеко не 100% результат, являются разработки в области распознавания английской речи, что связано, во-первых, с их наибольшей применимостью, во-вторых, с относительной простотой английского языка по сравнению с русским языком.

Для распознавания звуков речи и перевода их в буквенные обозначения необходимо использовать классификацию звуков. Одна из наиболее стройных и экономных классификаций звуков построена с учетом акустической природы речи. Акустическая классификация построена в 1955 году Р. О. Якобсоном (США), Г. Фантом (Швеция) и М. Халле (США) на основе анализа спектров звуков.

Ротовая полость представляет собой систему акустических резонаторов, настроенных на несколько различных гармоник. При изменении положения речевых органов меняются и частоты этих гармоник [1]. Когда в такую систему резонаторов попадает сложный звук, в нем усиливаются частоты, совпадающие с резонансными, и ослабляются другие частоты.

Области усиления энергии в спектре звука называются формантами. Для различных звуков русской речи положение формант в спектре различное. Кроме того, отличаются и амплитуды частот спектра. По совокупности этих параметров есть возможность распознать отдельные звуки речи.

Считается, что для характеристики звуков речи достаточно выделения четырех формант, которые нумеруются в порядке возрастания их частоты. Однако в большинстве случаев для различения звуков достаточно первых двух формант. Среднее расстояние между формантами для мужских голосов составляет приблизительно 1000 Гц, для женских и детских – несколько больше.

На рис. 1 в верхнем ряду показаны типичные формы звуковой волны трех гласных русского языка - "а", "и", "о". Графики приведены для мужского голоса в случае, когда гласные произносятся отдельно друг от друга. В нижнем ряду рисунка представлены в том же порядке спектры гласных звуков. По горизонтальной оси откладывается частота, измеряемая в Гц, по вертикальной оси – амплитуда в дБ. В спектрах хорошо различимы пики гармоник основного тона и форманты речи. Если провести плавную огибающую, охватывающую гармоники в областях спектральных максимумов, можно выделить частоту, уровень и ширину формант. Например, на первом графике для звука «а» хорошо заметны три форманты, для которых можно определить диапазон частот.

Сравнение спектров различных звуков позволяет понять, чем в спектральном отношении одни звуки отличаются от других. Например, если сравнивать гласный звук «и» с другими гласными звуками, то можно отметить, что для этого звука относительно большую роль играют высшие форманты.

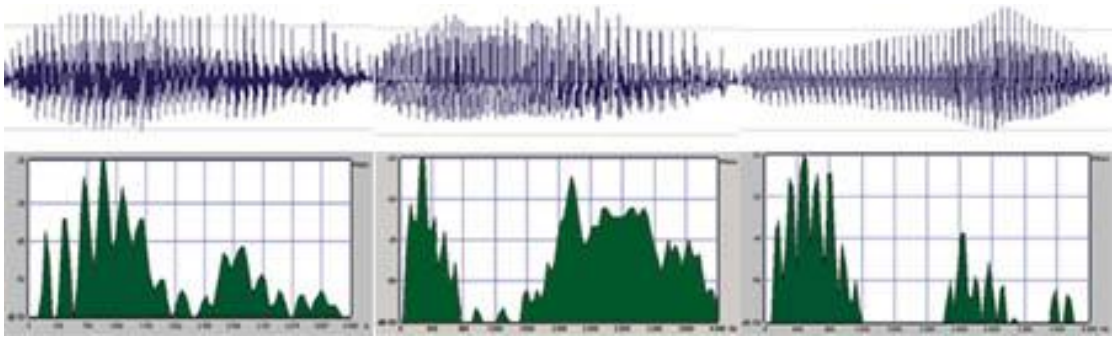

Рис. 1. Типичные формы и спектры звуковой волны трех гласных русского языка

Поскольку уже существуют абстрактные математические модели, описывающие речь, то к ее анализу можно подходить с физических позиций. То есть, можно исследовать звуковую волну, преобразованную математическими методами, не задумываясь о ее смысловой и эмоциональной нагрузке. Эта возможность уже реализована в автоматических или полуавтоматических системах идентификации и аутентификации голосов людей.

Задача идентификации голоса заключается в том, чтобы сравнить неизвестный речевой сигнал с имеющимися сигналами из базы. В процессе анализа характеристик выносится решение о том, чей голос звучит. Если при сравнении с голосом из базы данных система показывает достаточную близость, речевой сигнал приписывается диктору из базы.

Рассмотрим процедуру сравнения подробнее. Анализ речевого сигнала начинается с того, что он переводится в цифровую форму и затем сегментируется. После этого акустический сигнал обрабатывается с помощью определенных алгоритмов - спектрального анализа, линейного предсказания, кепстральной обработки и других. В результате получается параметрическое описание сегментов речевого сигнала в виде ряда параметров. По этим параметрам и производится сравнение с голосами их базы.

Тот же алгоритм работы может быть положен в основу устройств, способных переводить звуки русской речи в текстовое выражение.

Как уже говорилось выше, автомата, который бы понимал речевой сигнал и переводил его в текстовую форму, пока не существует. Это связано со сложностью работы с речью. Прежде чем распознавать и сравнивать, необходимо выделить элементы речи и классифицировать их в соответствии с единицами языка (фонемами). Однако если для отдельно произнесенных звуков это не так сложно, работа со слитной речью вызывает существенные проблемы. В слитной речи нет четких стационарных процессов. Звуки и слова могут быть не разделены паузами, речь отличается высокой степенью вариантивности, различным темпом, не всегда четким произношением, поэтому алгоритм точной сегментации речевого потока на звуковые сегменты пока не разработан.

Таким образом, качество распознавания слитной русской речи определяет качество акустических и языковых моделей, используемых в системе распознавания, где не последнюю роль играет структура самой системы, которую в свою очередь определяют различные функциональные модули и блоки, входящие в ее состав.

Не секрет, что современные технологии распознавания основаны на применении различных подходов в достижении поставленной цели. Так, в "Центре речевых технологий" для описания акустических моделей используют комбинацию классической теории цифровой обработки сигналов и технологии искусственных нейронных сетей. Такие модели наиболее устойчивы к междикторской вариативности, а также к помехам и искажениям, вносимым окружением или каналом передачи [2].

Авторами предлагается следующая структура системы распознавания слитной речи, укрупненная блок-схема которой приведена на рис. 2.

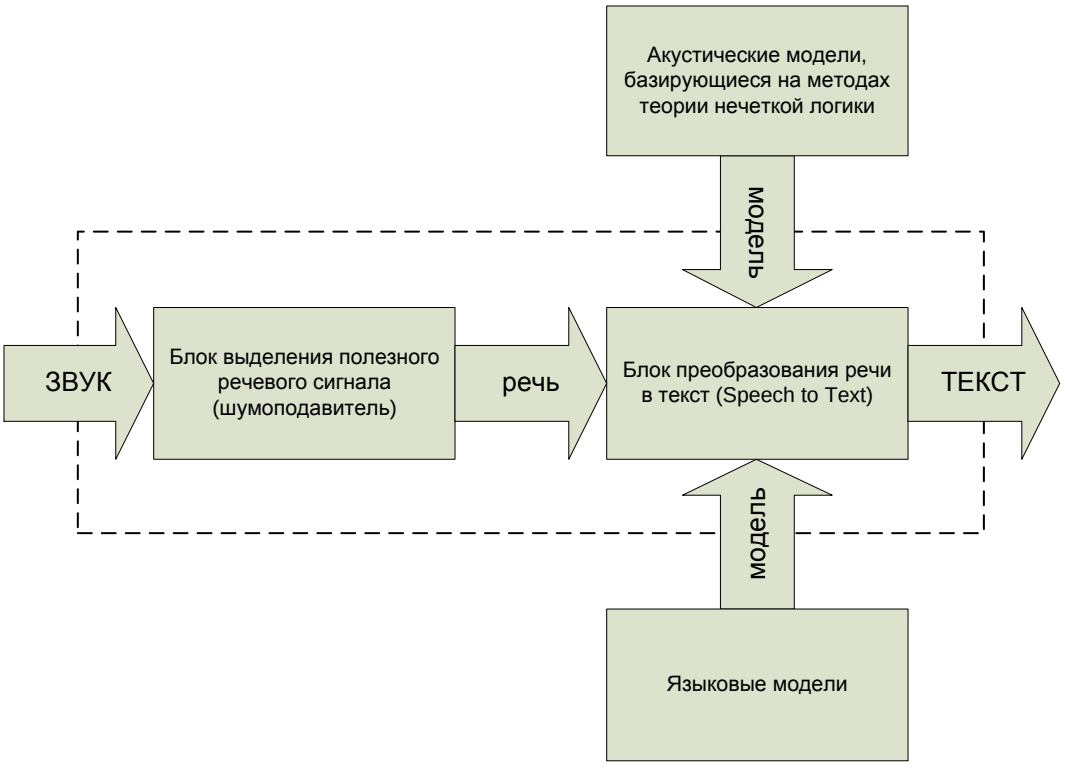

Рис. 2. Блок-схема системы распознавания слитной речи

Ядром системы распознавания слитной речи служит блок преобразования речи в текст, на вход которого поступает звуковой сигнал, очищенный от помех и искажений в блоке выделения полезного речевого сигнала (шумоподавителе [3]). Процесс распозна-

вания речи включает использование информации, содержащейся в акустических и языковых моделях, подключаемых в систему извне (выбор модели зависит от языка).

Использование для описания акустических моделей элементов теории нечеткой логики, в частности, многомерных логических регуляторов с переменными в виде совокупности аргументов двузначной логики [4], позволит повысить эффективность распознавания речи. Многомерные логические регуляторы являются альтернативной ступенью развития нечётких регуляторов, где основной упор делается на повышение быстродействия системы автоматического регулирования, компенсацию взаимного влияния контуров регулирования, их использование в интеллектуальных системах управления, что значительно расширяет область применения регуляторов данного типа.

Качество распознавания речи непосредственно зависит от качества и объёмов словарей, используемых в языковых моделях. В современных программах распознавания речи зачастую используются гибридные языковые модели, основанные на синтаксисе того или иного языка и классических статистических моделях, что обусловлено крайне высокой сложностью построения русской языковой модели. В связи с этим, авторами предполагается составление собственных тематических словарей и включение их в гибридную русскую языковую модель.

В заключение необходимо отметить, что предложенная авторами структура системы распознавания слитной речи в перспективе позволит сделать систему независимой от конкретного языка.

Список литературы:

1. Нейман Л.В., Богомильский М.Р. Анатомия, физиология и патология органов слуха и речи: Учеб. для студ. высш. пед. учеб. заведений / Под ред. В.И. Селиверстова. – М.: ВЛАДОС, 2001. – 224 с.

2. Центр речевых технологий | ЦРТ. URL: http://www.speechpro.ru/ (дата обращения: 28.03.2012).

3. Шишкина А.Ф. Шумоподавитель в ипульсных усилителях с линейной дельта-модуляцией // Радиоэлектроника, электротехника и энергетика: труды Международной конференции студентов, аспирантов и молодых ученых. В 2 т. – Томск, 6-8 октября 2011 г.: Томский политехнический университет. Т.1. – Радиоэлектроника, электротехника и энергетика. – 340 с. – С. 56-60.

4. Антипин А.Ф. Система автоматизированной разработки многомерных логических регуляторов с переменными в виде совокупности аргументов двузначной логики // Автоматизация в промышленности. 2011. № 3.