УДК 004.522

О РАСПОЗНАВАНИИ РЕЧИ

С.Л. БАБАРИНОВ
М.А. БУДНИКОВА

Белгородский государственный национальный исследовательский университет

e-mail:
babarinov@bsu.edu.ru

В данной статье рассмотрены факторы, по которым можно произвести классификацию систем распознавания речи. Затронуты основные проблемы распознавания речи как таковой, так и современные тенденции в области.

Ключевые слова: распознавание речи, классификация, системы распознавания, речь, систематизация

В ходе развития компьютерных систем становится очевидным, что эффективность использования этих систем может быть повышена в случае использования естественного и распространённого для человека инструмента общения – речи. В частности это позволит ускорить ввод информации и управление компьютерными, и особенно, мобильными системами.

В настоящее время во всем мире ведутся работы по созданию более естественных, чем существующие, для человека средств общения с компьютером, среди которых присутствует и речевой ввод информации. Многими разработчиками отмечены успехи в области распознавания речи, но массового использования таких технологий на российском рынке не наблюдается, что может быть следствием зависимости от диктора, недостаточной точностью распознавания для непрерывной речи и высокой чувствительностью к наличию различного вида помех. Неплохих успехов добилась корпорация Google Inc., предлагающая речевой ввод при осуществлении поиска в сети Интернет. В русском сегменте лидирует компания Яндекс с новой разработкой Yandex.SpeechKit которая представляет собой мультиплатформенную библиотеку, предоставляющая разработчикам мобильных приложений доступ к технологии распознавания речи Яндекс [8]. Тем не менее, проблема речевого ввода информации осложняется рядом факторов: различием структуры языков, спецификой произношения, шумами и помехами, акцентами, ударениями и т.п.

Существующие на сегодняшний день системы распознавания речи основываются на сборе всей доступной и даже избыточной информации, необходимой для распознавания лексических элементов. Системы подобные распознавания речи крупных компаний, таких как Google Inc. Используют широкую базу сэмплов речевых паттернов сотен и даже тысяч дикторов, что позволяет им добиваться уверенного распознавания многих слов неизвестных дикторов. Некоторые исследователи считают [4,5], 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, пока невозможно после получения разнообразных реальных сигналов осуществить прямую трансформацию в лингвистические символы, что является желаемым результатом.

Распознавание речи представляет собой сложную, поэтапную задачу распознавания образов. В ходе решения этой задачи, речевые данные анализируется, и классифицируются согласно заданной иерархии. Классифицированные образы могут представлять собой различные структурные элементы, отрезки речевых данных определённой длительности (фонема, слоги, слова). Чем больше мы предполагаем априорной информации о входном сигнале, тем качественнее мы можем его обработать и распознать.

В общем случае, каждая отдельная задача идентификации речи сводится к тому, чтобы выделить, классифицировать и соответствующим образом отреагировать на акустические колебания, представляющие собой человеческую речь из входного сигнала.

Это может быть и выполнение определенного действия на команду человека, и выделение определенного слова-маркера из большого массива телефонных переговоров, и системы для голосового ввода текста.

Голосовое управление основано на технологии распознавания речи: система получает информацию о колебаниях воздуха через микрофон, сравнивает полученные данные с командами, которые записаны в системе и, в случае совпадения, выполняет предписанное действие. Чем больше лексических единиц поддается распознаванию с высокой точностью, тем больше шанс, что система распознает команду без ошибок.

Для распознавания речи акустический сигнал при помощи детектирующих и оцифровывающих устройств и машинной обработки фиксируется и преобразуется в цифровую форму. В результате дискретизации непрерывный (аналоговый) сигнал переводится в последовательность чисел. Наиболее популярные методы цифровой обработки речевых сигналов: частотный анализ в базисе Фурье, вейвлет анализ, кепстральный анализ, субполосный анализ.

Далее приводятся основные факторы, по которым возможно классифицировать различные системы распознавания.

Размер словаря. Частота ошибок системы распознавания напрямую зависит от количества слов в словаре системы распознавания. Так словарь из нескольких десятков слов может быть распознан с достаточно высокой точностью. В то время как частота ошибок при увеличении количества слов до сотен и тысяч влечет собой серьезные потери в точности распознавания, причем они тем больше, чем больше слова имеют сходств друг с другом (различия в приставках, окончаниях).

Дикторозависимость. Дикторозависимая система предназначена для использования одним пользователем, т.е. настраивается под индивидуальные характеристики речи, в то время как дикторонезависимая система предназначена для работы с любым диктором и не учитывает индивидуальных особенностей произношения. Дикторонезависимость – труднодостижимая цель, так как при обучении системы, она настраивается на параметры того диктора, на примере которого обучается. Частота ошибок в дикторонезависимых системах серьезно превышает дикторозависимые системы.

Тип речи. Обычно для ввода используются либо отдельные слова и словосочетания, либо требуется найти слова маркеры в слитной речи. Распознавание слитной речи намного труднее в связи с тем, что границы отдельных слов не четко определены и их произношение сильно искажено смазыванием и проглатыванием некоторых произносимых звуков.

Область применения. Назначение системы определяет требуемый уровень абстракции, на котором будет происходить распознавание. Существует несколько видов систем, для которых применяется распознавание речи: системы поиска слов маркеров, системы речь-в-текст и системы речь-в-речь. В системе поиска слова или фразы маркера распознавание маркера происходит как распознавание единого образа. Системы речь-в-текст и речь-в-речь (системы синхронного перевода) требуют повышенной точности распознавания, учитывающий не только распознаваемый в данный момент паттерн, но и предыдущие, уже распознанные паттерны. Таким образом, система должна анализировать не только некоторые отдельные слова и словосочетания, но также и учитывать контекст, в котором они были произнесены.

Тип лексической структурной единицы. При анализе речи, в качестве базовой единицы анализа могут быть выбраны отдельные слова и словосочетания, слоги, а также такие элементы как фонемы, аллофоны, дифоны и, реже, трифоны. От типа лексической структурной единицы зависит как сложность системы в целом, так и качество распознавания, и размер используемого словаря.

Механизм работы. В современных системах широко используются различные подходы к механизму функционирования распознающих систем. Вероятностно-сетевой подход состоит в том, что речевой сигнал разбивается на определенные части (кадры, либо по фонетическому признаку), после чего происходит вероятностная оценка того, к какому именно элементу распознаваемого словаря имеет отношение данная часть и (или)

весь входной сигнал. Подход, основанный на решении обратной задачи синтеза звука, состоит в том, что по входному сигналу определяется характер движения артикуляторов речевого тракта и, по специальному словарю происходит определение произнесенных фонем.

Использование дополнительной неречевой информации. В последнее время набирают популярность системы распознавания речи, использующие неакустические параметры такие как: движения губ, языка, мышц лица (фиксируемые камерой), ультразвук, колебания в костях черепа, а также электромиографию фиксирующую активность голосовых связок и гортани [4,5,10]. Основной причиной появление таких методов является желание повысить количество информации, пригодной для распознавания. Подобного рода ларингофоны впервые использовались в системах связи танковых войск, где уровень акустических шумов внутри машин был довольно высок и требовались дополнительные источники информации, кроме непосредственной фиксации звуковых коллебаний. Системы распознавания речи, использующие такие подходы позволяют добиваться более высокой точности, а также снимать многие ограничения, накладываемые на акустический тракт в виду воздействия на него помех и шумов.

Согласно представленным факторам составлена классификация, которая, по мнению авторов, наиболее полно описывает существующие системы распознавания речи (рисунок).

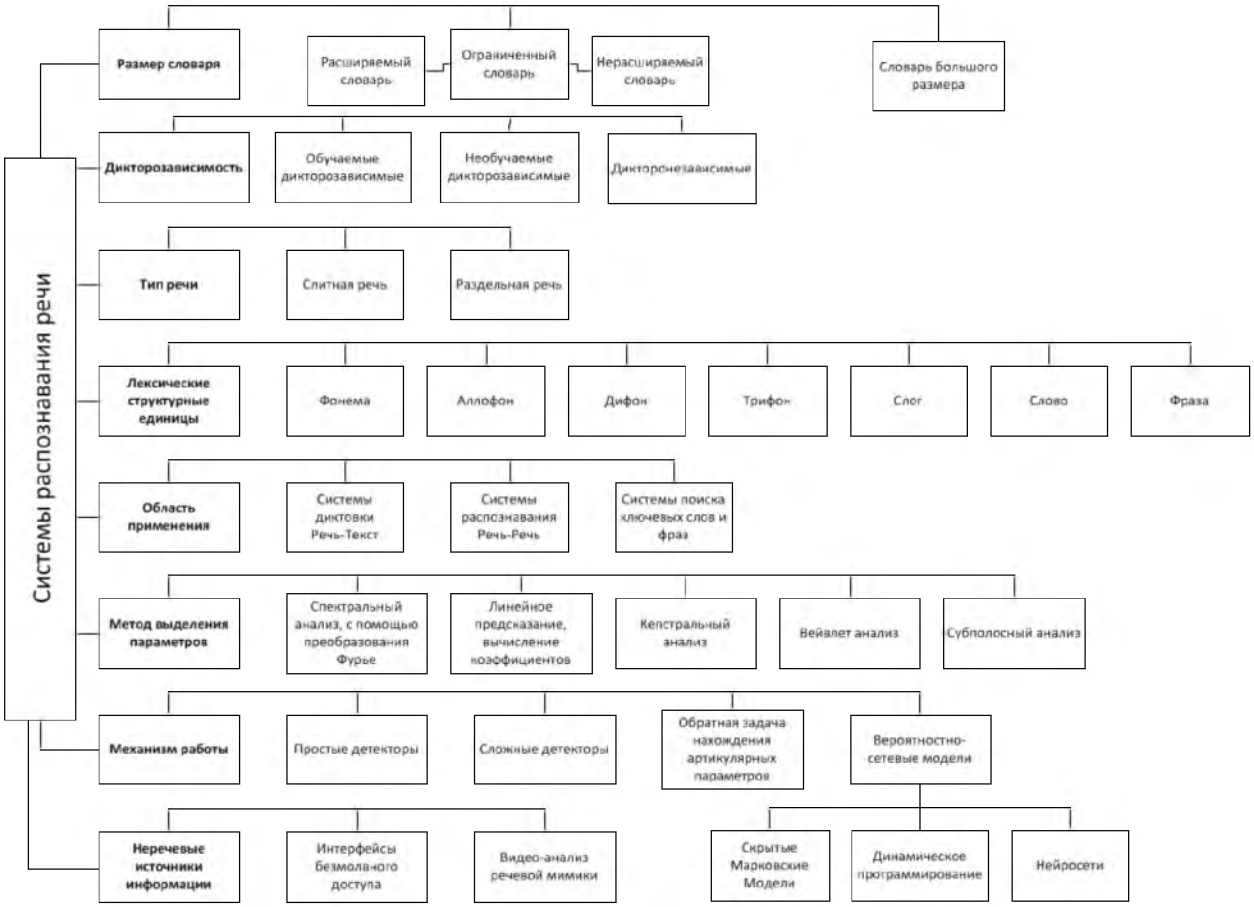

Рис. Классификация систем распознавания речи

Анализ основных факторов влияющих на структуру систем распознавания речи показал, что в настоящее время не существует такой системы, которая обладала универсальностью, надежностью и простотой. Представленная в данной работе классификация

систем распознавания речи позволит сузить область исследований в этом направлении при разработке новых алгоритмов, методов и систем распознавания речи.

Список литературы

1. Mohri, M., Pereira, F., & Riley, M. (2008). Speech recognition with weighted finite-state transducers. In Springer Handbook of Speech Processing (pp. 559-584). Springer Berlin Heidelberg.
2. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82-97.
3. Denby B, Schultz T, Honda K, Hueber T, Gilbert J.M., Brumberg J.S. (2010). Silent speech interfaces. Speech Communication 52: 270–287
4. Brumberg J.S., Nieto-Castanon A, Kennedy P.R., Guenther F.H. (2010). Brain–computer interfaces for speech communication. Speech Communication 52:367–379. 2010
5. Jorgensen C, Dusan S. (2010). Speech interfaces based upon surface electromyography. Speech Communication, 52: 354–366
6. IT.TUT.BY Интервью с директором ООО «Речевые Технологии» Виталием Киселевым. [Электронный ресурс] // IT.TUT.BY – Информационные технологии в Беларуси. 2008. Режим доступа: http://it.tut.by/news/97283.html
7. Речевые технологии. Программы распознавание речи [Электронный ресурс] – Дата обновления 20 янв 2010, Режим доступа: http://speech-soft.ru/index.php?a=inf&inf=rasp
8. Яндекс. SpeechKit API [Электронный ресурс] – Дата обновления 24 окт. 2013, Режим доступа: http://api.yandex.ru/speechkit/
9. Хабрахабр. Распознавание речи [Электронный ресурс] – Дата обновления 15 июля 2009, Режим доступа: http://habrahabr.ru/post/64572/
10. RealSpeaker. Аудио-видео распознаватель речи [Электронный ресурс] – Дата обновления 07 мая 2012, Режим доступа: http://realspeaker.net/ru/
11. Жиляков Е.Г., Бабаринов С.Л., Чадюк П.В. Исследование сервиса компании Google Inc. по распознаванию русской речи Научные ведомости Белгородского Государственного Университета, №15 (158) 2013 г., выпуск 27/1

ABOUT SPEECH RECOGNITION

S.L. BABARINOV
M.A. BUDNIKOVA

Belgorod State National Research University

e-mail: babarinov@bsu.edu.ru

This article discusses the factors, which can make a classification of speech recognition systems. Addressed the basic problems such as speech recognition, and modern trends.

Keywords: speech recognition, classification, recognition, speech, systematization