

УДК 81'322

## ЛИНГВИСТИЧЕСКИЕ ПОДХОДЫ К АВТОМАТИЧЕСКОМУ РАСПОЗНАВАНИЮ РЕЧИ

© 2013 г.

*А.М. Грачев*

Нижегородский госуниверситет им Н.И. Лобачевского

grachol@yandex.ru

*Поступила в редакцию 13.12.2013*

Анализируется возможность использования автоматического распознавания речи. Приведена разработка лингвистически ориентированных распознающих систем, которые позволили выработать ряд рекомендаций для лучшего распознавания речи.

*Ключевые слова:* распознавание речи, математические модели, спектрограмма, фонема, идентификация, декодирование.

Для современного этапа развития лингвистической науки характерна ориентация на междисциплинарность и на расширение сфер практического применения лингвистической теории.

В этой связи особую актуальность приобретают исследования, посвященные проблеме автоматического распознавания речи. Эта проблема имеет широкий спектр возможностей прикладного применения в разных областях – от криминалистики до лингводидактики.

Предметом исследования в настоящей работе являются лингвистически ориентированные системы автоматического распознавания речи у нас в стране и за рубежом.

Рассматривая анализ речевых процессов у человека, В.И. Галунов подчеркивает, что близость систем автоматической обработки речи к организации обработки речи у человека представляется критической. Математические модели обработки речевой информации показали свою ограниченность и не позволяют решить сложные, но решаемые легко человеком проблемы, связанные с распознаванием слитной речи без подстройки под диктора [1, с. 129].

По мнению американского исследователя В. Зу [2], одного из самых активных сторонников этого подхода, неудачи акустико-фонетического распознавания в различных системах АРР обусловлены двумя причинами:

1) использованием слишком упрощенных представлений о соотношении речевого сигнала и его фонемного (аллофонического) отражения в языке;

2) применением таких способов первичной акустической обработки речевого сигнала, которые слишком далеки от того, как это осуществляется в слуховой системе человека.

Для того чтобы лучше понять, как человек распознает звучащую речь, в конце 70-х – начале 80-х годов были активизированы эксперименты по чтению (дешифровке) неизвестных («слепых») динамических спектрограмм речи на материале разных языков, в том числе на русском. Эксперименты показали, что фонетист, интерпретирующий спектрограммы, может выделить из акустического сигнала значительный объем фонетической информации, при этом точность и надежность распознавания оказываются выше, чем соответствующие показатели у систем распознавания речи. Было обнаружено также, что чтение спектрограмм и их фонемная (фонетическая) интерпретация основаны на выделении и использовании большого количества разнообразных акустических ключей (полезных признаков). Наконец, подробный анализ дешифровки спектрограмм человеком привел к заключению, что фонетическое декодирование базируется на вполне определенных акустико-фонетических правилах, которые поддаются формализации и могут быть переданы машине.

Эти результаты вселяли надежду на то, что построение надежной системы распознавания речи на основе использования фонетических знаний вполне возможно. Однако, подводя итоги экспериментов по чтению спектрограмм, В. Зу в 1985 г. отметил, что создание высококачественного фонетического распознавателя все-таки остается исключительно трудной задачей из-за неполноты знаний о лингвистически существенных акустических признаках, отсутствия надежных способов их выделения в сигнале и сложности правил фонетической интерпретации. Его прогноз состоял в том, что могут

понадобиться десятилетия для того, чтобы овладеть этими знаниями и научиться ими пользоваться. Этот прогноз, по-видимому, оправдывается, так как в рамках лингвистического подхода, основанного на знаниях, пока не удалось построить надежных систем распознавания речи.

В.И. Галунов в своем обзоре современного состояния речевых технологий в России и за рубежом отмечает, что многие варианты использования речевых технологий (автоматическое распознавание и синтез речи) предполагают автоматическую идентификацию пользователя. Прежде всего это касается использования речевых технологий в телефонных сетях, где приходится решать все задачи обработки в их единстве: распознавание и синтез речи, распознавание личности говорящего и компрессия речи [1].

Задачи определения личности говорящего традиционно принято подразделять на верификацию и идентификацию. При этом под верификацией подразумевается, что диктор предъявляет эталонный образец своего голоса, произнося парольную фразу (называет имя, фамилию, PIN-код, пароль или имя доступа — login). Система автоматического распознавания индивидуальных характеристик голоса и речи (или эксперт) должна подтвердить или отвергнуть индивидуальность данного конкретного лица. Предъявить пароль (а следственно, и право индивидуального доступа к системе или средству информации) может как истинный носитель данной индивидуальности, так и злоумышленник. Исходя из риска потерь в случае возможного несанкционированного доступа, можно для данной системы определить допустимую вероятность пропуска «чужого».

В целом система верификации может быть охарактеризована ошибками 1-го рода (захват ложной цели или, в нашей интерпретации, это принятие злоумышленника за зарегистрированного пользователя) и ошибками 2-го рода (пропуск цели или отказ признать зарегистрированного пользователя). Каждая данная система может перестраиваться таким образом, что ошибки одного рода могут быть уменьшены за счет увеличения ошибок другого рода (даже при сохранении всех других факторов, влияющих на вероятность ошибки: длительности и характера речевого сообщения, помехи и т.п.). Изменение соотношения ошибок 1-го и 2-го рода достигается путем изменения порога принятия решения.

Классическая процедура идентификации диктора на закрытом множестве подразумевает,

что имеется ограниченная и строго контролируемая группа пользователей системы. При поступлении речевого сигнала на вход системы определения индивидуальных речевых характеристик эта система должна определить, кто из пользователей в настоящий момент вступает в речевой контакт с системой ограниченного доступа. В определенном смысле так формулируемая задача идентификации говорящего проще задачи верификации, сформулированной выше. В этом случае исключена ситуация допуска возможного злоумышленника. Разделение же допущенных пользователей (особенно при их небольшом количестве) может оказаться более простой задачей, чем задача верификации.

Пропускная биометрическая система может быть охарактеризована средней вероятностью правильной идентификации.

По мнению Е.И. Галяшиной, все перечисленные выше виды задач верификации и идентификации могут быть разбиты на два больших класса в зависимости от используемого речевого материала: текстозависимые и текстонезависимые. Разница между этими двумя классами заключается в том, является ли исследуемый речевой отрезок лингвистически подобным сравнительному образцу или нет. По способу решения задачи идентификации могут быть разделены на автоматические системы и субъективные (экспертные) [3, с. 134].

В целом на сегодняшний день разработка лингвистически ориентированных распознающих систем привела к формулировке следующих важных рекомендаций:

1) первичная акустическая обработка речевого сигнала должна основываться на сведениях о свойствах периферического слухового анализатора; должна быть реализована возможность выделения акустических событий и признаков, которые обнаруживаются слуховой системой человека, по крайней мере, в известном на сегодня объеме;

2) на ранних этапах применения фонетических знаний не следует добиваться полного фонетического распознавания неизвестной фразы. Вместо этого нужно выделять надежные акустические признаки звуковых единиц (признаки главных классов и способов образования) и использовать их для первичной сегментации речевого сигнала, а также для отбора из машинного словаря системы наиболее подходящих слов-гипотез;

3) необходимо учитывать просодическую информацию, прежде всего для определения

ударных слогов, которые могут быть подвергнуты детальному фонетическому декодированию и использоваться далее в целях ограничения возможного набора слов-гипотез;

4) каждое слово-гипотеза, прошедшее «мягкий» отбор по грубым классифицирующим признакам и фонемному составу ударного слога, должно далее проверяться (верифицироваться) на наличие в распознаваемом акустическом сигнале тех акустических признаков, которые задаются его полным звуковым обликом (транскрипцией) [4].

#### Список литературы

1. Галунов В.И. Современные проблемы распознавания речи // Информационные технологии и вычислительные системы. М., 2004. Вып. 2. С. 41–45.
2. Зу В.В. Лингвистический подход к автоматическому распознаванию речевых сигналов // Труды института инженеров по электротехнике и радиоэлектронике (ТИИЭР). Речевая связь с машинами. 1985. № 73. С. 51–54.
3. Галяшина Е.И. Основы судебного речеведения: Монография. М.: СТЭНСИ, 2001. 236 с.
4. Кодзасов С.В., Кривнова О.Ф. Общая фонетика: Учебное пособие. М.: РГГУ, 2001. 592 с.

## LINGUISTIC APPROACHES TO AUTOMATED SPEECH RECOGNITION

*A.M. Grachev*

The possibility of using automatic speech recognition is analysed. The development of linguistics-oriented identification systems, which resulted in the working-out of a number of recommendations for better speech recognition, is provided.

*Keywords:* speech recognition, mathematical models, spectrogram, phoneme identification, decoding.