УДК 004.934.1'1(043.2)

Халаменда Н.Н.

РАСПОЗНАВАНИЕ ЧЕЛОВЕЧЕСКОЙ РЕЧИ В КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЯХ

Институт компьютерных технологий
Национального авиационного университета

Рассмотрено распознавание речи в компьютерных системах. Сформированы задачи для успешного распознавания речи. Рассмотрены недостатки существующих сегодня систем распознавания речи, а также предложены методы для устранения негативного эффекта

Введение

По мере развития компьютерных систем становится все более очевидным, что использование этих систем намного расширится, если станет возможным использование человеческой речи при работе непосредственно с компьютером, и в частности станет возможным управление машиной обычным голосом в реальном времени, а также ввод и вывод информации в виде обычной человеческой речи.

Существующие технологии распознавания речи не имеют пока достаточных возможностей для их широкого использования, но на данном этапе исследований проводится интенсивный поиск возможностей употребления коротких многозначных слов для облегчения понимания. Распознавание речи в настоящее время нашло реальное применение в жизни, пожалуй, только в тех случаях, когда используемый словарь сокращен до 10 знаков, например. Так что насущная задача – распознавание, по крайней мере, 20 тысяч слов естественного языка – остается пока недостижимой. Эти возможности пока недоступны для широкого коммерческого использования. Однако ряд компаний своими силами пытается использовать уже существующие в данной области науки знания.

Для успешного распознавания речи следует решить следующие задачи:

- обработку словаря;
- обработку синтаксиса;
- сокращение речи;
- выбор диктора;
- тренировку дикторов;
- выбор особенного вида микрофона;
- условия работы системы и получения результата с указанием ошибок.

Существующие сегодня системы распознавания речи основываются на сборе всей доступной информации, необходимой для распознавания слов. Исследователи считают, что таким образом задача распознавания образца речи, основанная на качестве сигнала, подверженного изменениям, будет достаточной для распознавания, но тем не менее 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.

Вместо этого проводится процесс, первым шагом которого является первоначальное трансформирование вводимой информации для сокращения обрабатываемого объема так, чтобы ее можно было бы подвергнуть компьютерному анализу. Примером является "техника сопоставления отрезков", позволяющая сократить вводимую информацию с 50'000 до 800 битов в секунду. Следующим этапом является спектральное представление речи, получившееся путем преобразования Фурье. Результат преобразования Фурье позволяет не только сжать информацию, но и дает возможность сконцентрироваться на важных аспектах речи, которые интенсивно изучались в сфере экспериментальной фонетики. Спектральное представле-

ние достигнуто путем использования широко-частотного анализа записи.

Хотя спектральное представление речи очень полезно, необходимо помнить, что изучаемый сигнал весьма разнообразен. Разнообразие возникает по многим причинам, включая:

- различия человеческих голосов;
- уровень речи говорящего;
- вариации в произношении;
- нормальное варьирование движения артикуляторов (языка, губ, челюсти, нёба).

Для устранения негативного эффекта влияния варьирования голосового тракта на процесс распознавания речи было использовано множество методов. Первым делом рассматривалась характеристика пространства траектории артикуляторных органов, включая гласные, используемые говорящим. Техника динамичного искажения используется для временного вытягивания и сокращения расстояния между искаженным спектральным представлением и шаблоном для говорящего. Использование данной техники дало улучшении точного распознавания . Метод динамичного искажения используют практически все коммерчески доступные системы распознавания, показывающие высокую точность сообщения при использовании.

Вначале сигнал преобразовывается в спектральное представление, где определяется немногочисленный, но высокоинформативный набор параметров. Затем определяются конечные выходные параметры для варьирования голоса и производится нормализация для составления шкалы параметров, а также для определения ситуационного уровня речи. Вышеописанные измененные параметры используются затем для создания шаблона. Шаблон включается в словарь, который характеризует произнесение звуков при *передаче информации говорящим,* использующим эту систему. Далее в процессе распознавания новых речевых образцов, эти образцы сравниваются с шаблонами, уже имеющимися в словаре, используя динамичное искажение и похожие метрические измерения. В настоящее время этот метод изучается и дополняется.

Очевидно, что спектральное представление речи позволяет характеризовать особенности голосового тракта человека и способ использования его говорящим. Самый обычный способ моделирования специфических эффектов "модель-источник" - использование фильтров. Речевой аппарат моделируется с использованием источников, вызывающих резонанс, ведущий к пиковым точкам интенсивности звука в соседстве с отдельными частотами, называемыми формантами. При произнесении звуков вибрация голосовых связок является источником возбуждения, и эти короткие импульсы вызывают резонанс между голосовыми связками и губами. Так как язык, челюсть, губы, зубы и альвеолярный аппарат двигаются, размер и место этих резонансов меняются, давая возможность воспроизведения особых параметров звуков.

Возможно построить очень точную модель, также прямо смоделировать движения артикуляторов физиологически реальным путем. Использование этих моделей привели к пониманию пути, в котором происходит речевой сигнал. Но так как наблюдение над артикуляторами затруднено, остаются недостатки. Возможно все аспекты влияния акустической структуры контролируют сигналы и форму звукового выхода речи.

Аспекты влияния акустической структуры включает в себя:

- природу сегментов индивидуального звука (гласные/согласные);
- структуру слога;
- структуру морфем (приставки, корни, суффиксы);
- лексикон;
- уровень синтаксиса фраз и предложений;
- долгосрочные ограничения речи (*long-term discourse constraints*).

Ниже рассматривается влияние ограничений и способ их воздействия про-

изводство сигнала речи. Необходимо также принять во внимание тот факт, что человеческий аппарат восприятия также должен быть смоделирован, он сам по себе накладывает на процесс восприятия дополнительные ограничения. Недавно процесс восприятия был изучен с помощью метода сигнального подавления барабанных перепонок через возбуждение нервных клеток, которые образовывают примерно 30 тысяч нервных окончаний слухового нерва. Но изучение нервных окончаний способно только прояснить формирование простых синтетических гласных. Перед исследователями встало новое главное направление в области изучения воспроизводства речи, связанное с интеграцией всей физиологии восприятия человека. В настоящий момент появляются некоторые модели явлений, происходящих в ухе, и не без оснований можно ожидать дальнейшего улучшения понимания процесса распознавания речи из-за более полного понимания характеристик этого влияния.

Что касается уровня артикуляторного контроля, первым уровнем является индивидуальный фонетический сегмент, иначе говоря, – фонема. Во многих естественных языках их примерно 40. Но их набор существенно различается. Поэтому, например, английские гласные могут быть носовыми, даже ненамеренно, в то время как во французском носализация гласных является фонетическим контрастом, и поэтому влияют на значение произносимого.

На следующем уровне лингвистической структуры фонетические сегменты сгруппированы в согласные/гласные, а следовательно и в слоги. Далее, в зависимости от роли фонетического сегмента внутри этих слогов их реализация может быть сильно изменена. Так например, начальный согласный в слоге может быть реализован как абсолютно отличный от конечной позиции. Согласные очень крепко связываются между собой, что опять же влияет на последующие ограничения.

Говорящие на родном языке избегают этих ограничений или могут активно их использовать во время процесса восприятия. Из выше приведенных примеров очевидно, что хотя и существуют сильные ограничения, влияющие на слушателя, но их сила не является решающей во время произнесения речи. То есть любое моделирование процесса восприятия может быть активным и может оказать большую помощь в понимании главного смысла.

Другой пример, показывающий необходимость применения сфокусированного поиска, может быть представлен в восприятии конечного согласного. Среди многих ключевых слов для распознавания конечного согласного существует спектральная природа шума, воспроизводимого при освобождении конечной перемычки и перехода резонанса второй форманты в гласный, следующий за этой перемычкой. Многие исследователи изучали эти влияния, и результаты их исследований показали, что ограничивающее влияние обоих вышеописанных характеристик на восприятие варьируется природой следующего гласного, и следовательно, мощная стратегия распознавания должна иметь некоторые знания о твердой позиции гласного перед конечным согласным перед тем, как будет сделано само распознавание конечного согласного.

Кроме сегментного и слогового уровней существуют ограниченные влияния из-за структуры морфем, которые являются минимальными синтаксическими единицами языка.

Дополнительные ограничения на природе входа новой лексики в язык могут являться уровнем слова. Многие исследования обнаружили, что характеристика слов при введении разбиения на 5 жестких классов фонетических сегментов может быть сокращена до минимума, часто имея единственное в своем роде распознавание. Далее слишком усиливается эффект порядка двух букв и фонетических сегментов с тех пор как в изучении английских и французских словарей было обнаружено, что более 90% слов имели

единственное значение и только 0,5% имели 2 и больше альтернатив. На фонемном уровне было обнаружено, что все слова в английском словаре из 20 тысяч слов имели одно значение из-за беспорядочных фонемных пар. Этот пример помогает показать, что все еще существует ограничивающее влияние на лексическом уровне, которое еще не определено в современных системах распознавания речи. Естественно, что исследования в этой области продолжаются.

Кроме уровня слов синтаксис имеет дополнительное ограничительное влияние. Его влияние на последовательный порядок слов часто характеризуется в системах фактором, который в свою очередь характеризует количество возможных слов, которые могут следовать за предыдущим словом в процессе произнесения. Синтаксис также имеет ограничительные влияния на просодические элементы, такие как ударение. Далее, кроме синтаксического уровня ограничения доминируют над семантикой, прагматикой и речью, что плохо осознается людьми, однако имеет очень важное значение для процесса распознавания.

Закодированные представления спектральной трансформации воспроизводства речи используются для нахождения самого правильного пути через сеть, и недавно были получены очень хорошие результаты. Очень важно подчеркнуть использование такого формально- структурного подхода, который способствует автоматичному определению классов символов через структурирование и параметризацию.

При другом подходе базы данных и связанные с ними процессы обработки используются структурой контроля. Этот подход был изучен системой *HEARSAJ* 2, которая была разработана в институте *Carnegie-Mellon University*, и системой *HWIM* (*hear what I mean*). В этих системах комплексная структура данных, которая содержит всю информацию о воспроизведении звуков, изучается с точки зрения конкретных ограничений. Но как выше указано, каждое из этих ограничений имеет особую внутреннюю модель, и полный анализ не может быть произведен. Для проведения анализа в целом структура данных должна иметь взаимодействие между разными процессами, а также средства для интеграции. Несмотря на то, что структура включает в себя несколько весьма различных источников знаний и ее вклад в понимание речи очень общий, она также имеет большое количество степеней свободы, которые могут быть использованы для тщательного системного воспроизведения.

В заключение следует сделать акцент на влияние производственной технологии на эти системы. Технология интеграции не является большой проблемой для систем распознавания речи, наоборот, это является архитектурой этих систем, включая способ представления ограничений. Необходимо провести грандиозные эксперименты и найти новые способы, которые необходимы для ограничительного влияния взаимодействия.

Во многих способах распознавание речи имеет типичный пример стремительно развивающегося класса высоко интегрированных комплексных систем, которые должны использовать лучшую компьютерную технику и самые последние достижения современного математического обеспечения.

Список литературы

1. *А.С. Рылов* Анализ речи в распознающих системах. – Минск: Бестпринт, 2003. – 263 с.

2. *Т.К. Винцюк* Распознавание слов устной речи методами динамического программирования // Кибернетика. – М.: 1968. – С. 81–88.

3. *Методы* автоматического распознавания речи: В 2-х книгах. Пер. с англ./Под ред. У. Ли. – М.: Мир, 1983. – Кн. 1. – 328 с.