

Караульных Д. В.

МЕТОД ВЫДЕЛЕНИЯ ХАРАКТЕРИСТИК РЕЧИ ДЛЯ СРАВНЕНИЯ С ЭТАЛОНОМ ПРИ РАСПОЗНАВАНИИ

Работа выполнена под руководством к.т.н., доц. Орлова В. В.

*ТГТУ, кафедра «Конструирование радиоэлектронных
и микропроцессорных систем»*

В настоящее время разработчики стремятся создать интерфейсы взаимодействия с пользователем более привычные для человека, чем традиционные клавиатура и мышь. Одним из перспективных направлений в этой области является речевой ввод команд в вычислительную систему или устройство. Большинство современных систем распознавания речи основываются на двух подходах.

Первый – сравнение фрагмента звуковой волны, записанной от пользователя, с набором эталонов, выбор наиболее подходящего эталона, и определение, на основании этого, произнесённого звука. В этом случае эталоны представляют собой простой отрезок оцифрованных звуковых колебаний (сэмплов)[1].

Основными недостатками данного метода является:

- необходимость записи большого количества сэмплов, учитывающих различные сочетания предшествующих и следующих звуков;
- большой объём памяти, необходимый для хранения базы эталонов такого типа (даже при использовании какого-либо алгоритма сжатия, например, mp3, объём данных для приемлемого качества распознавания остаётся достаточно большим);
- корректная работа только с одним пользователем, который предварительно наговаривал все эталоны, что обусловлено различием частоты основного тона голоса, гармоник которого не совпадают у разных людей.

Второй подход основан на использовании частотного спектра в качестве величины для сравнения записанного сигнала с эталонами[2]. Это устраняет недостаток предыдущего метода, заключающийся в требовании большого количества памяти. Однако, остальные недостатки, связанные с необходимостью записи большого количества эталонов и привязки к голосу конкретного человека, остаются.

Избавиться от недостатков последнего метода можно, если при сравнении с эталоном использовать не весь звуковой спектр, а лишь часть. В устной речи используется диапазон частот, примерно, от 30 до 3000Гц[3], поэтому остальную часть можно убрать из рассмотрения.

Основной тон голоса не несёт никакой смысловой нагрузки и воспринимается человеком только для идентификации произносящего. Частота тона зависит от возраста говорящего, его пола, а также некоторых других параметров, и, обычно, у различных людей бывает от 100 до 350 Гц[4]. Следовательно, из спектра, используемого при сравнении, можно исключить частоты, ниже 350 Гц.

Анализируя конкретный язык, на котором говорит пользователь, можно определить такие участки в спектре, которые не участвуют при произношении любого звука. Если звуковой сигнал пропустить через полосовой фильтр, выделяющий эти частоты, то человек, общающийся на этом языке, не сможет ничего услышать. Эта особенность восприятия речи вырабатывается у каждого человека в течении жизни. Часть спектра, не несущая информации, перестаёт восприниматься нашим ухом. Эти частоты также можно исключить из сравниваемых. В речи они могут быть заполнены шумом, который мы не воспринимаем, но для анализа будут являться фактором, который вызовет неправильный результат.

Кроме того, оставшиеся частоты можно разбить на группы, шириной от 100 Гц в начале спектра и 1000 Гц в конце (3000 Гц при анализе речи). Это действие допустимо в связи с тем, что участки спектра, отвечающие за тот или иной звук, варьируются в небольших пределах и точное сравнение огибающей спектра приведёт только к увеличению погрешности оценки.

Таким образом, эталоном при данном подходе являются наборы участков спектра, которые можно выделить из сигнала, используя полосовые фильтры, настроенные на середины каждого из участков частот, и шириной, зависящей от положения в спектре относительно начала.

Применение такого подхода позволяет исключить привязку к конкретному человеку и значительно уменьшить количество эталонов за счёт использования при сравнении лишь части огибающей спектра.

Список литературы:

1. <http://www.aha.ru/~takomat/comprus.htm>
2. Сегментация речи (математическая модель) #5 май 2006 А. В. Аграновский, Д. А. Леднов, Б. А. Телеснин, КБ "Спецвузавтоматика", г. Ростов-на-Дону, Ростовский-на-Дону Государственный Университет
3. <http://forum.sources.ru/index.php?s=299df20436040932087eb792345ebb33&showtopic=143082&st=45&#entry1192808>
4. <http://www.offline.computerra.ru/1999/293/3577/>