

Д. А. ЛЬВОВ

Научный руководитель И. А. ЕВСЕЕНКО, канд. техн. наук, доц.
БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ

Актуальность работы заключается в том, что в настоящее время наиболее перспективными являются более естественные для человека методы управления компьютером. Одним из таких методов является речевой метод.

Процесс произнесения звуков речи имеет несколько основных стадий. Человек создаёт звуковые волны, которые распространяются в пространстве. Звуки могут формироваться при участии истинных голосовых связок и без их участия и от этого существенно меняется их образ. Звуковые колебания воспринимаются микрофоном, и как результат преобразования имеется аналоговый сигнал, что дает возможность применить аналоговые методы анализа сигнала. Как правило, на этой стадии могут применяться системы фильтров. Однако, если рассматривать распознавание речи в приложении к компьютерным технологиям на уровне программного обеспечения, то необходимо провести этап преобразования из аналогового сигнала в непрерывно-дискретный. Зная, что диапазон частот человеческого голоса составляет примерно от 200 Гц до 2000 Гц применяем систему фильтров для входного сигнала, чтобы избавиться от помех и шумов.

После того, как сигнал отфильтрован можно приступить ко второму этапу обработки. На втором этапе осуществляется программный анализ поступающего сигнала. Он состоит из нескольких частей: анализ отдельных звуков, анализ отдельных фонем (используя разделение непрерывного сигнала на домены), анализ по смыслу отдельных слов и предложений.

Новизной является применение скрытой Марковской модели (фиксирование перехода из одного состояния в другое и расчет вероятности события последующего после перехода) для распознавания речи. Возможно также использование искусственных нейронных сетей, применение которых эффективно повышает мощность Марковской модели, однако требует большее количество ресурсов. При этом параллельно производится корректировка по времени, обучение алгоритма с помощью некоторой корреляционной функции, также применяются словари морфем слов.

На завершающем этапе производится семантический анализ поступающей непрерывной речи и превращение ее в некоторые команды.

Таким образом, по ходу анализа входной непрерывной речи осуществляется подстройка под диктора, однако существует несколько препятствий: разные акценты, сложная структура семантики языка, несовершенная система подавления шумов.