

С. Евчик, В. В. Касьяник
(БрГТУ, Брест)
ГОЛОСОВОЕ УПРАВЛЕНИЕ
МОБИЛЬНЫМ РОБОТОМ

Язык – это естественный способ коммуникации между людьми, однако в современной робототехнике все чаще возникает проблема удобного человеко-машинного интерфейса взаимодействия между роботом и человеком. Естественным желанием здесь является использование голосовых команд – естественного языка человека. Люди изучают язык с раннего детства без специальных инструкций, просто слушая родителей, и используют его всю жизнь для общения с другими людьми, не задумываясь о том, как работает речевой аппарат. На сегодняшний день существует множество подходов к решению задачи машинного распознавания речи: скрытые Марковские модели, статистические методы акустического и временного моделирования, гибридные модели. Среди готовых программных систем стоит отметить IBM ViaVoice, Microsoft Voice, SvoiceControl, Dragon Dictate, SPHINX. Однако все эти решения имеют как достоинства, так и недостатки, к сожалению универсального решения пока не существует. Перспективным направлением исследований в данной области является применение нейросетевого подхода к распознаванию речи. Здесь выделяют два способа использования нейронных сетей для моделирования акустического сигнала: прогнозирование и классификация. В данной работе предлагается модель нейросетевой системы распознавания голосовых команд для управления мобильным роботом, так как этот подход характеризуется независимостью от диктора, скоростью работы, фильтрацией шума, архитектурным параллелизмом.

Для решения задачи предлагается следующая структура программного обеспечения: предварительная обработка сигнала; нейросетевая обработка сигнала. Предварительная обработка сигнала состоит из: преобразования сигнала в его частотную форму и последующего сжатия полученных данных. Для перехода от амплитудной к частотной форме сигнала используется преобразование Фурье.

Фурье-преобразование позволяет не заботиться о расположении сигнала на временной оси, а только о его частотном представлении. На рисунке 1 приведен образец реального входного сигнала, а на рисунке 2 его спектр.

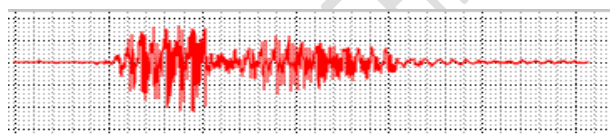


Рисунок 1 – Представление входного сигнала

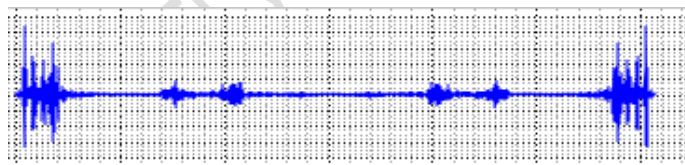


Рисунок 2 – Спектр сигнала

Входной сигнал имеет размерность 3072 отсчета с частота дискретизации 8000 Гц. Число бит на отсчет 32. Для нейросетевой обработки сигнала предлагается использовать многослойную нейронную сеть. Параметры нейронной сети: 5 выходных нейронов, 100 нейронов в скрытом слое, 1536 входных нейронов. Функция активации: сигмоидная. Для обучения нейронной сети используется алгоритм обратного распространения ошибки. Обучающая выборка сформирована следующими образом: 2000 образов, 1536 значений на входе, 5 на значений на выходе. Выходные значения представлены в бинарной кодировке: 1 — верно, 0 — не верно. Для каждого образа выходные значения представлены одной 1 и четырьмя 0. На данный момент ведется разработка алгоритмов предварительной обработки, чтобы повысить коэффициент распознавания до приемлемого значения 95 %.

В результате работы планируется получить программную систему распознавания голоса независимую от диктора для управления мобильными роботами в СНИЛ «Робототехника» БрГТУ. Кроме того, применение такой системы будет возможно и в других областях, где необходим механизм удобного человеко-машинного взаимодействия.

Литература

1. Фролов А.В., Фролов Г.В. Синтез и распознавание речи. Современные решения М.: МИР, 1976 452 с.
2. Lippman R.P., Gold B. Neural-net classifiers useful for speech recognition //IEEE Int. Conf. Neural Networks. San Diego, 1984.
3. http://speech.bme.ogi.edu/tutordemos/nnet_training/tutorial.html#2.1_Quick_Review_of_Frame-Based_Speech

4. <http://speech-text.narod.ru/chap3.html>

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ