



9. Kohonen, T. *Self-Organizing Maps (Third Extended Edition)*. – New York, 2001, 501 p.

А.С. Подгорный

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА РАСПОЗНАВАНИЯ РЕЧИ В РОБОТЕ АССИСТЕНТЕ

(Уфимский государственный авиационный технический университет)

Создание ассистента, способного распознавать человеческий голос и отвечать на него всегда было сложной задачей, по большей части из-за несовершенства технологий распознавания речи. Но с тех пор интеллектуальные технологии сильно продвинулись, благодаря чему создание собственного ассистента стало под силу даже студенту, причём абсолютно бесплатно.

Размещение ассистента на базе робота позволяет сделать его мобильным, а использование полноценного микрокомпьютера позволяет значительно расширить набор датчиков и устройств, которыми можно дополнять функционал робота.

Робот перемещается на гусеничном шасси (рис.1). В его основе стоит платформа *Nvidia Jetson*[2], содержащий в себе средство для ускорения работы нейросетей, что не обязательно для работы распознавания голоса, зато позволяет тратить меньше энергии на обработку и выполнять более сложные задачи в будущем.



Рис. 1. Внешний вид робота



Интеллектуальная система распознавания голоса позволяет роботу распознавать человеческую речь и отвечать на команды человеческим голосом. Для распознавания речи используется библиотека *Vosk* [1], а для генерации голоса — *RHVoice* [3]. Последняя отличается тем, что она изначально создавалась для русского языка, благодаря чему качество голоса выше чем у систем, изначально построенных для озвучивания английского языка.

Роботами может выполняться следующие наборы команд:

- Общие команды: Привет, пока, как дела, скажи что-нибудь, расскажи анекдот,

- Показания датчиков: Температура, влажность, давление, углекислый газ, органические вещества;

- Команды движения: Вперёд, назад, влево, вправо, танцуй.

Система выполняет команду в среднем через 1 секунду после произнесения. Наибольшая точность достигается при общении с одним человеком. Когда несколько людей говорят одновременно, система не всегда распознаёт команды. Это связано с тем что основой для обучения служат записи, где текст озвучивает один человек, например, в аудиокниги. При этом система очень устойчива к шумам и громкой музыке.

Одной из проблем такой системы, специфичной именно для ассистента является эхо. Когда робот произносит фразу, внутренний микрофон улавливает её и пытается распознать. Для решения данной проблемы в данном робот-ассистенте используется программный модуль эхоподавления на основе технологии *WebRTC*.

Таким образом робот-ассистент был снабжён локальной системой распознавания голоса, что позволяет использовать его в помещениях без доступа к интернету, а также делает робота-ассистента полностью мобильным.

Литература

1. Библиотека для распознавания речи «*Vosk*». URL: <https://alphacephei.com/vosk/index.ru> (дата обращения: 05.01.2022).

2. *Jetson Nano 2GB Developer Kit User Guide*. URL: <https://developer.nvidia.com/embedded/learn/jetson-nano-2gb-devkit-user-guide> (дата обращения: 05.08.2021).

3. Документация к библиотеке *RHVoice*. URL: <https://github.com/RHVoice/RHVoice/blob/master/doc/ru/index.md> (дата обращения: 10.01.2022).