Секция «Математические модели в космосе»

# Сравнение и анализ алгоритмов обнаружения речи для задач голосового управления

**Научный руководитель – Сазонова Софья Викторовна**

***Литвинов Иван Борисович***
*Студент (специалист)*
Московский государственный университет имени М.В.Ломоносова, Факультет космических исследований, Москва, Россия
*E-mail: litvan007@gmail.com*

Рассматривается задача создания механической робо–руки, управление которой осуществляется посредством голосовых команд. Первым из этапов разработки голосового интерфейса является выделение участков звукового сигнала с речью. Для решения подобных задач используется семейство алгоритмов Voice Activity Detection (VAD).

Входные данные представляют собой дискретизированный звуковой сигнал: звуковой сигнал разбивается на кадры (фреймы) достаточно малой длины, чтобы в пределах одного кадра сигнал можно было считать постоянным.

Для алгоритмов VAD характерны кадры длительностью 25 мс с шагом 10 мс∼[3]. После дискретизации сигнала происходит априорная оценка пороговых значений выбранных ключевых характеристик на эталонных кадрах шума. Затем проводится сравнение значений признаков на каждом кадре входного сигнала с полученными пороговыми значениями, а затем классификация каждого кадра как шум и перерасчет пороговых значений.

Результатом нашей работы стало сравнение и анализ работы двух алгоритмов из семейства VAD и последующая их модификация в связи со спецификой рассматриваемой задачи. Один из них, применяемый в системах распознавания казахской речи, использует комбинацию энергии и среднего числа перехода через нуль значений сигнала∼[2]. Описанный алгоритма предполагает выбор фреймов длины 16 мс с шагом 16 мс, что приводит к ошибочным результатам. Чтобы улучшить его работу, сигнал разбивался фреймы длины 25 мс с шагом 10 мс в соответствии с∼[3]. Другой алгоритм основывается на рассмотрении энергии и энтропии сигнала∼[1], который также был доработан, так как предложенная последовательность действий не приводит к нужному результату.

На (рис. 1) показаны результаты работы до модификации двух алгоритмов при длине фрейма 16 мс и с шагом разбиения 16 мс, на (рис. 2) показаны результаты работы после модификации двух алгоритмов при длине фрейма 25 мс с шагом разбиения 10 мс.

Оба алгоритма показали высокую эффективность в обработке зашумленных сигналов. Однако если использовать в качестве комбинации признаков — энергию и среднее число перехода через нуль значений сигнала, то такой подход имеет значительный выигрыш по времени и поэтому в дальнейшей работе планируется использовать именно модификацию этого алгоритма.

## Источники и литература

1) Ермоленко Т. В., Тихончук А. П. Определение голосовой активности в речи // Проблемы искусственного интеллекта. 2017. №2(5). С.45–46.

2) Калимолдаев М. Н., Мамырбаев О. Ж., Мусабаев Р.Р., Тусупова Б.Б. Методы применения VAD в системах распознавания казахской речи // Проблемы информатики. 2013. №1(18). С.65–66.

3) Jurafsky D., Martin J., Martin L. Speech and Language Processing. New Jersey. Prentice-Hall, Inc. 2008.

## Иллюстрации

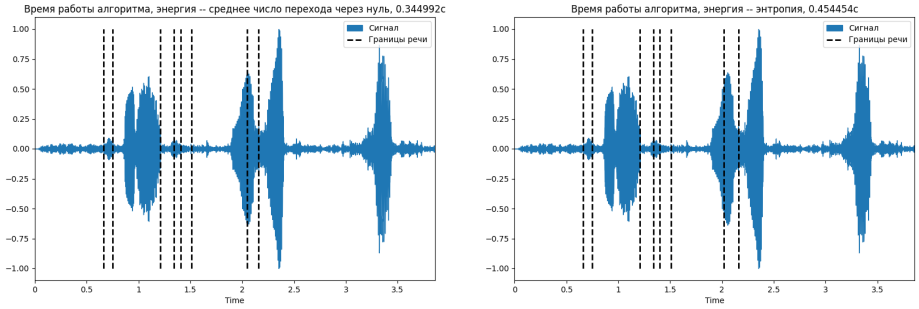


**Рис. 1.** Результаты до модификации алгоритмов

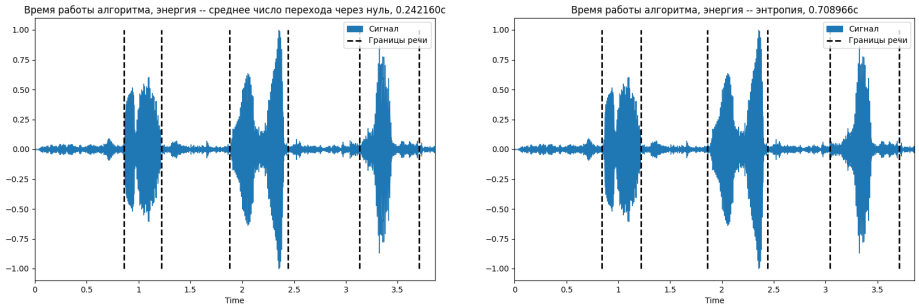


**Рис. 2.** Результаты после модификации алгоритмов