



Горшков Андрей Сергеевич,

Студент 4 курса, СПбГУАП, г. Санкт-Петербург

МОДЕЛИРОВАНИЕ АЛГОРИТМА СЕГМЕНТАЦИИ РЕЧЕВОГО СИГНАЛА В СРЕДЕ MATLAB

Аннотация: В статье рассматривается алгоритм распознавания речи, её сегментации на участки активности и пауз. Существует несколько алгоритмов для распознавания речи, таких как алгоритм с применением частотной обработки, с применением машинного обучения. Рассмотрен алгоритм пороговой обработки в различных условиях. Результатом моделирования является алгоритм, написанный в среде MATLAB.

Ключевые слова: MATLAB, фильтрация, речь, VAD, распознавание речи, сегментация речевого сигнала.

Алгоритм VAD (Voice Activity Detection) определяет границы речевой активности, обнаруживая наличие сигнальных признаков, которые характерны только для речи в записи. Обычно энергия сигнала выбирается в качестве такого признака. Однако много факторов может ухудшить работу алгоритма, таких как фоновый шум, шумные звуки, а также наличие малой энергии в начале или конце записи со звуком, которые могут постепенно нарастать или убывать.

Поэтому большинство алгоритмов регулярно модифицируются, чтобы повысить свою эффективность и оставаться конкурентоспособными на рынке разработок.

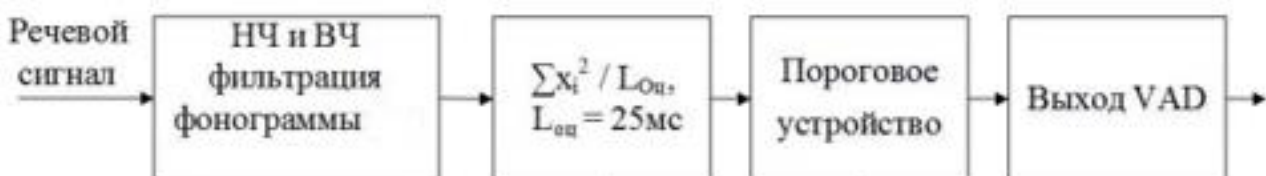


Рис. 1 – Схема разрабатываемого алгоритма



При исследовании спектров речевых сигналов было установлено, что частотный диапазон речи можно ограничить значениями от 150 до 3500 Гц. Следовательно, для ФНЧ выбирается частота среза, равная 150 Гц, а для ФВЧ - 3500 Гц.

Затем фрагмент сигнала делится на интервалы длительностью 25 мс, на которых вычисляются средние значения мощности сигнала с использованием формулы:

$$P_{cp} = \frac{\sum_{i=1}^N x_i^2}{N}, \quad (1)$$

где x_i — i -й элемент выборки, N — объем выборки.

На каждом интервале измерения $L_{оц}$ средние значения квадратов нормируются на длину этого интервала. Затем строится гистограмма, по которой определяется уровень порога активности в соответствии с заданным выражением:

$$П_{акт} = Y_0 K_{упр}, \quad (2)$$

где $K_{упр}$ — коэффициент управления порогом, Y_0 - базовый порог, который определяется как первый минимум при движении слева направо на гистограмме.

Из-за особенностей произношения шумных звуков их мощность значительно ниже, чем мощность вокализованных звуков. Это требует понижения порога обнаружения речевой активности с помощью коэффициента управления порогом. Это позволяет также определить участки шумных маломощных звуков как активную речь, что предотвращает потерю информации из РС. Однако, если порог будет слишком сильно понижен, возможно ложное срабатывание алгоритма во время пауз.

Для участка речевой активности или паузы используется пороговое устройство со значением равным среднему квадрату РС, которое определяется по заданному выражению:



$$Activity[s] = \begin{cases} 0, Y[s] \leq \Pi_{акт} \\ 1, Y[s] > \Pi_{акт} \end{cases}, \quad (3)$$

где $Y[s]$ сигнал, поступающий на вход порогового устройства.

Для проверки эффективности выбранных границ, необходимо сравнить работу алгоритма с и без фильтрации. Для этого предлагается использовать фонограмму слова «Горшков», применив границы фильтрации до 150 Гц и от 3500 Гц. Это слово включает в себя шумные звуки, такие как два фрикативных [в] и [ш] и взрывное [г] в начале слова, взрывной звук [к] по середине, которые могут повлиять на качество работы алгоритма. Однако, запись этой фонограммы была произведена в условиях низкого уровня шума, что позволяет VAD-алгоритму успешно выполнить задачу.

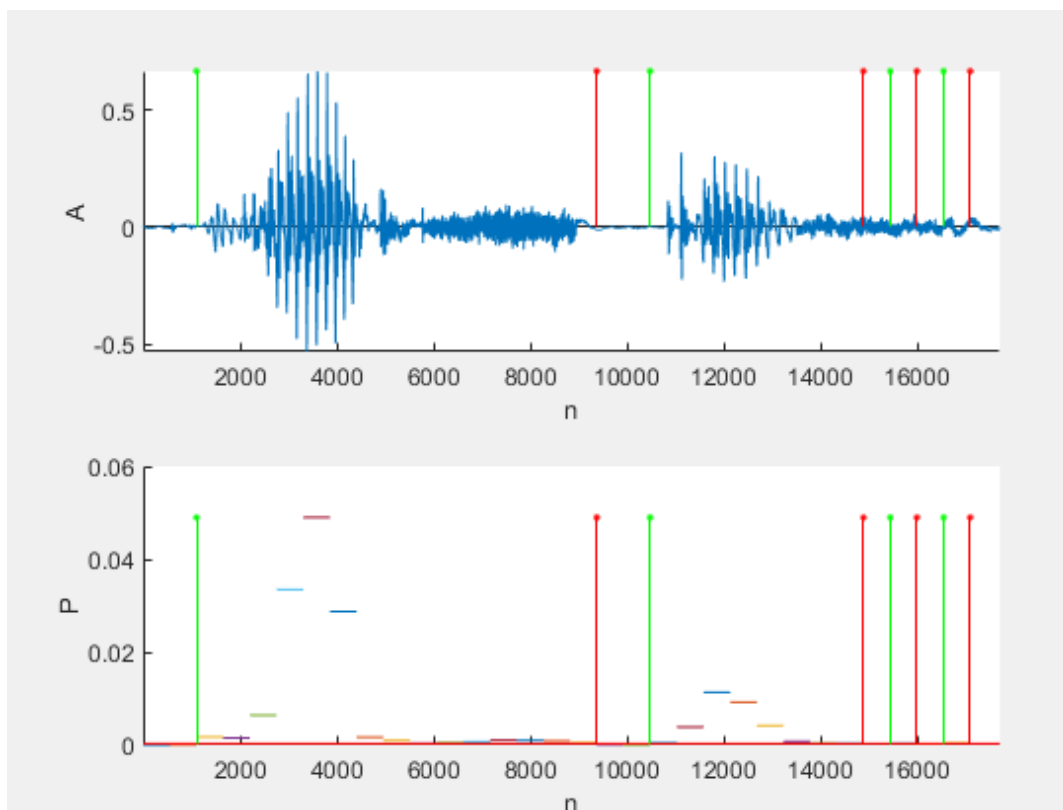


Рис. 2 – определение границ речевой активности с помощью VAD-алгоритма слова «Горшков», без фильтрации

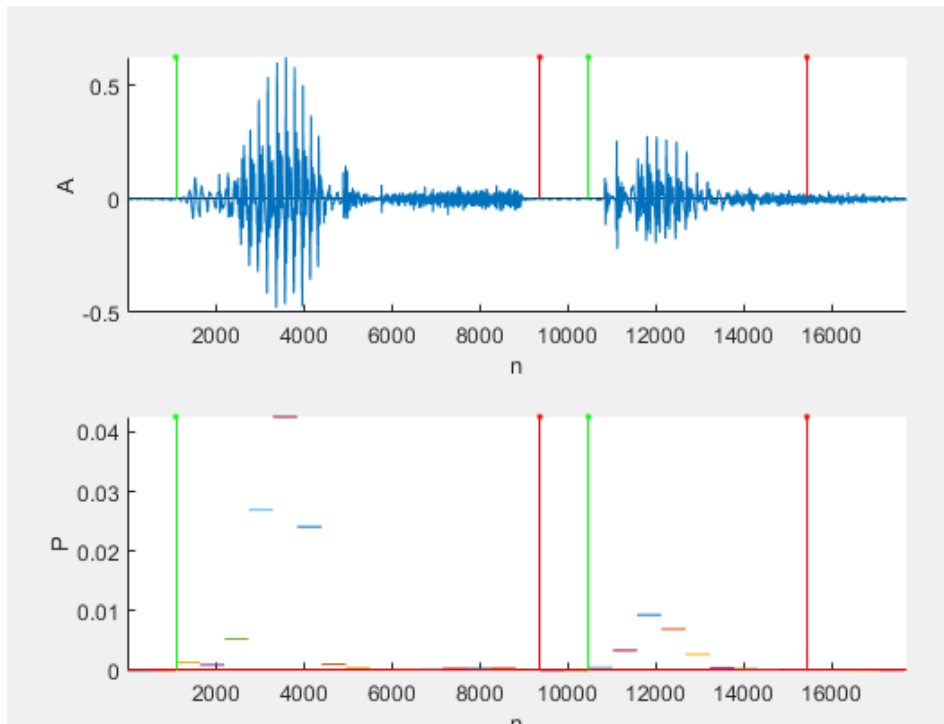


Рис. 3 – определение границ речевой активности с помощью VAD-алгоритма слова «Горшков», с фильтрацией

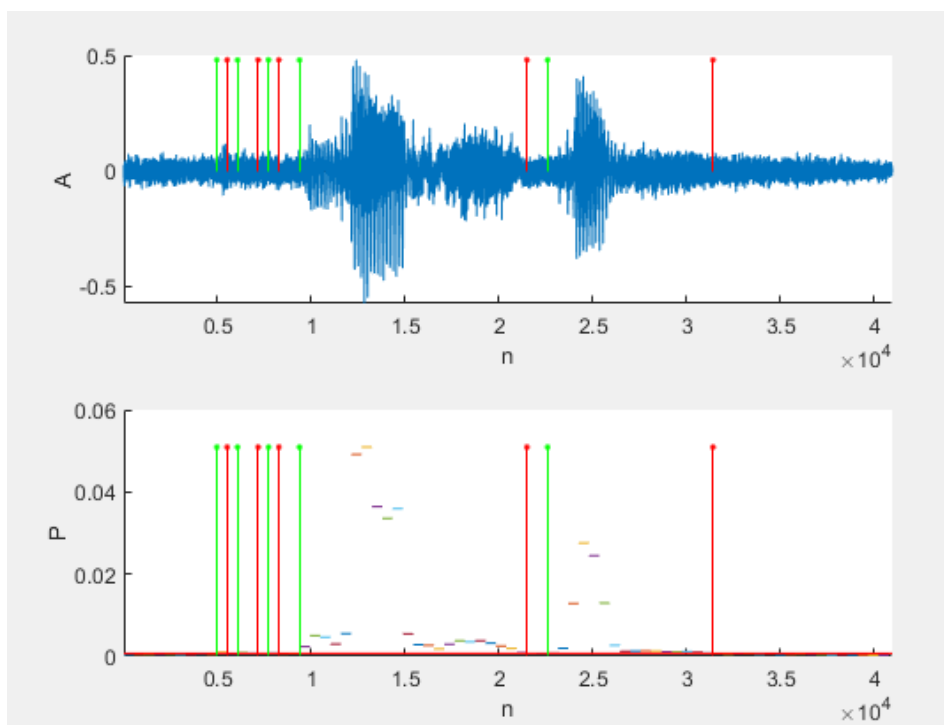


Рис. 4 – определение границ речевой активности с помощью VAD-алгоритма слова «Горшков» в шумовых условиях, без фильтрации

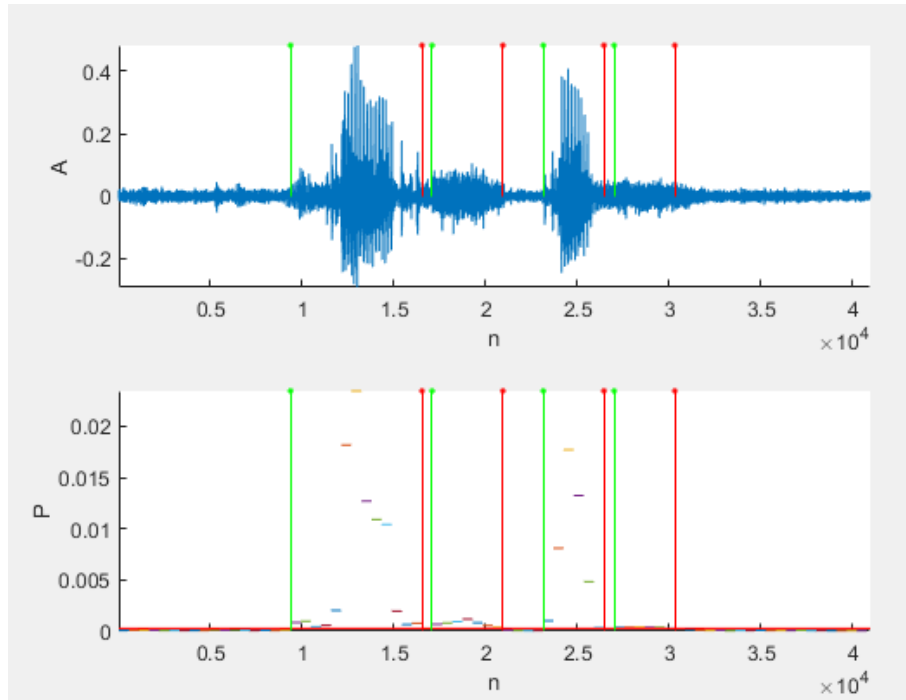


Рис. 5 – определение границ речевой активности с помощью VAD-алгоритма слова «Горшков» в шумовых условиях с фильтрацией

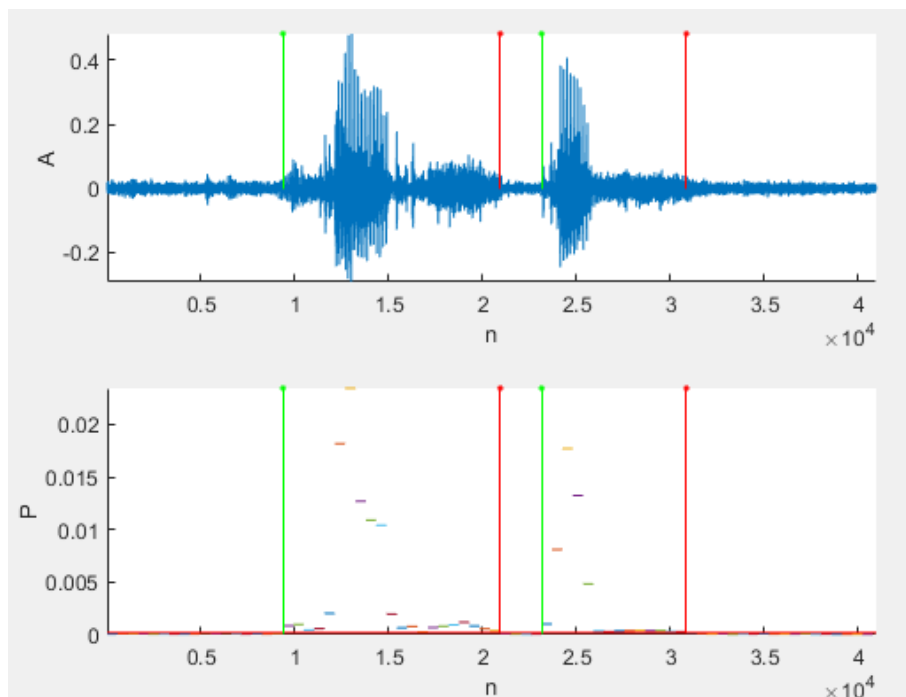


Рис. 6 – определение границ речевой активности с помощью VAD-алгоритма слова «Горшков» в шумовых условиях, с фильтрацией и понижением порогового коэффициента



Заключение:

Предложен алгоритм распознавания речевого сигнала (сегментация речевых сигналов на активные фрагменты и паузы). Результаты приведены с помощью моделирования в MATLAB при различных параметрах. На рисунках 2 и 3 заметны ошибки при определении границ речи без фильтрации (рисунок 2), с фильтрацией (рисунок 3) они практически отсутствуют, фонограмма была записана при тишине. На рисунках 3, 4 и 5 используется запись слова «Горшков» в шумовых условиях. Без фильтрации (рисунок 4) присутствуют ошибки детектирования голоса (ложное определение начала речи), с фильтрацией (рисунок 5) начало речи определено правильно, но далее обнаруживаются ложные паузы. При помощи понижения коэффициента порога из формулы (2) речь определяется в тех же границах, что и в условиях тишины. Алгоритм может быть предложен для использования в системах голосового управления.

Список литературы:

1. Акустика: Справочник / А.П. Ефимов, А.В. Никонов, М.А., Сапожков, В.И. Шоров; Под ред. М.А. Сапожкова. 2-е изд., перераб. и доп. М.: Радио и связь, 1989. – 336 с.
2. Гантов Д., Серебренников И. Основы речевых технологий. – М.: Интернет-Университет Информационных Технологий, 2004. – 80 с.
3. Джеймс Л. Фланаган. Анализ, синтез и восприятие речи: Пер. с англ./Под ред. А.А. Пирогова. М.: СВЯЗЬ, 1968. – 395 с.
4. Джейкобсон Р., Дэвис Л., Лэдлоу Д. Эффективные алгоритмы распознавания речи. - СПб.: Издательство «Питер», 2007. – 832 с.