

АЛГОРИТМ ОБРАБОТКИ ДЕФЕКТНОЙ РЕЧИ ПРИ ГОЛОСОВОМ УПРАВЛЕНИИ ЭЛЕКТРОННЫМИ УСТРОЙСТВАМИ

Гурин К.А.

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

Научный руководитель: Столер В.А. – к.т.н., доцент, зав. кафедрой ИКТ

Аннотация. Предложен алгоритм построения программы для распознавания ключевых слов, используемых при голосовом управлении электронными устройствами и интернет-ресурсами. Указаны способы оптимизации программы и повышения точности распознавания при дефектах произношения.

Ключевые слова: дефектная речь, алгоритм обработки звука, компьютерная программа

Введение. За последние несколько десятилетий произошло огромное развитие парадигмы взаимодействия пользователя с интерфейсом программ и устройств. В частности, большую популярность набирает автоматическое распознавание речи (*ASR*), используемое в голосовом управлении устройствами.

В настоящее время известно множество различных методов и устройств распознавания речи, однако большинство из них являются коммерческими. Поскольку коммерческие распознаватели речи доступны для четко определенных приложений, таких как произношение или транскрипция, многие проблемы *ASR*, такие как распознавание в шумной среде, низкое качество записи и распознавание при дефектах речи, еще предстоит эффективно решить [1].

В данной статье предложен алгоритм обработки речи для построения компьютерной программы распознавания ключевых слов при дефектах их произношения, используемой при звуковом (голосовом) управлении электронными устройствами.

Основная часть. Алгоритм обработки записанной речи при голосовом управлении устройством имеет следующий вид:

- перевести аналоговый сигнал в цифровой вид;
- произвести обработку шумов цифрового сигнала;
- задать погрешность шума, оставшегося в записи, $\Delta_{ш}$;
- преобразовать цифровой сигнал в массив координат $[x; y]$;
- выбрать количество интервалов для сравнения M ;
- выбрать количество точек N для сравнения на заданном интервале;
- задать погрешность для точек, в пределах сравниваемых интервалов графиков, $\Delta_{г}$;
- сравнить выбранные интервалы эталонного графика и графика записанной речи.

Эталонными считаются графики слов, записанных на профессиональном оборудовании, и содержащие минимальное количество шумов. Данные условия необходимы для минимизации числа ошибок при сравнении слов.

Рассмотрим подробно предложенный алгоритм.

В терминах цифровой обработки сигналов процесс шумоочистки представляет собой преобразование входного сигнала, содержащего как полезный сигнал – речь, так и аддитивный сигнал-помеху – шум, в выходной сигнал, содержащий только речь. Поскольку создание систем, в точности удовлетворяющих данному условию, невозможно, задачу очистки сигнала от шума упрощают [2].

Зададим погрешность для шума, оставшегося после обработки. Определим локальные максимумы графика. Интервал M_i , на котором график выходит за пределы заданного

значения $\Delta_{ш}$, имеет локальный максимум и ограничен локальными минимумами, является буквой в слове (рисунок 1).

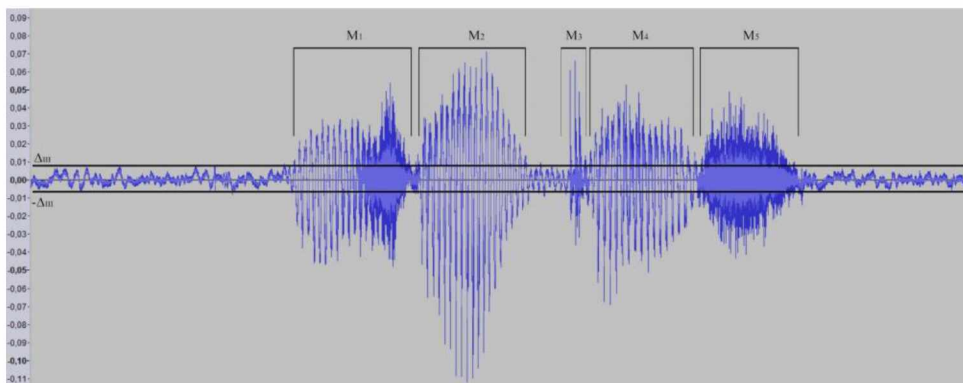


Рисунок 1 – Выбор интервалов M_i на графике записанной речи

Далее сопоставляем эталонный график и график записанной речи. Выберем количество точек N для сравнения на заданном интервале M_i и зададим погрешность для точек, в пределах сравниваемых интервалов графиков, Δ_T .

При равенстве точек $[x_j; y_j]$ графика записанной речи точкам $[x_j; y_k]$ эталонного графика с учетом выбранной погрешности Δ_T можно говорить о равенстве произнесенных букв.

Из равенства всех интервалов M_i сравниваемых звуковых сигналов следует равенство произнесенных слов.

Таким образом, общий алгоритм распознавания ключевых слов в записанной речи можно представить следующим образом (рисунок 2).

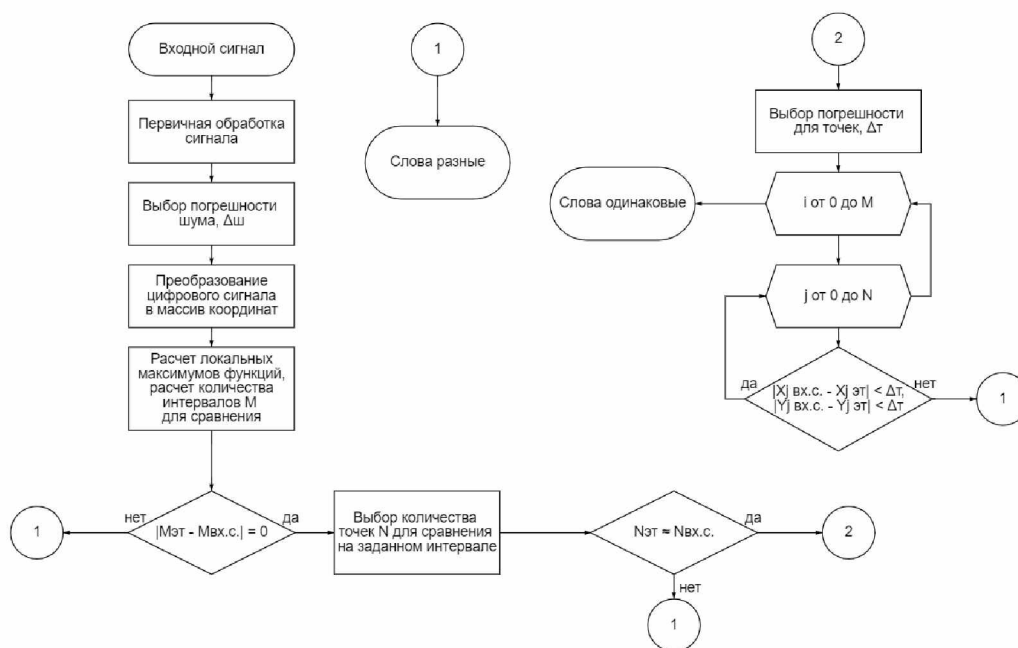


Рисунок 2 – Алгоритм обработки речи для распознавания ключевых слов

Для оптимизации алгоритма необходимо нормализовать входной сигнал, т. е. задать верхнюю границу значения y . Таким образом, если разница между минимальным (максимальным) значением y первого графика и минимальным (максимальным) значением y второго графика больше заданной погрешности точки Δ_T , можно говорить о неравенстве символов, т.е. неравенстве слов [3]. Для корректного определения M_1 необходимо учесть,

что значение u локального максимума символа много больше заданного значения погрешности шума $\Delta_{ш}$. Изменение значений погрешности шума $\Delta_{ш}$ и погрешности точки $\Delta_{т}$ позволяет задать необходимую точность сравнения.

При разработке алгоритма необходимо учесть следующие возможные ситуации: количество символов M в записанном слове больше, чем у эталонного; интервал M в записанном слове вмещает в себя большее количество точек N чем эталон.

Большее число символов M у записанного слова может означать не только более длинное слово, но и дефект произношения (например, «Привет»). Для обработки данной ситуации необходимо предусмотреть проверку, где в записанном слове будут попарно сравниваться соседние интервалы M_i и M_{i+1} . Если соседние интервалы равны между собой, значит M_i необходимо исключить из сравнения.

Также рассмотрим ситуацию, при которой длина интервала M_i в записанном слове больше длины соответствующего интервала M_i чем у эталонного, т.е. количество вмещаемых интервалом точек N различно. Такое различие может возникнуть не только при неравенстве символов, но и при более длинном произношении буквы (напр., «Ммама»). При систематическом повторении значений u в пределах рассматриваемого интервала, можно говорить о более длинном произношении символа и исключить часть интервала из сравнения для уравнивания значений N .

Заключение. Выполнен анализ основных принципов, используемых в программах распознавания речи. Установлено, что проблема распознавания речи, содержащей дефектные слова, в настоящее время ещё актуальна. Предложен алгоритм обработки речи, учитывающий некоторые дефекты произношения, для построения компьютерной программы распознавания речи, используемой при звуковом управлении электронными устройствами. Рассмотрены способы оптимизации программы и повышения точности распознавания: задаются минимально возможные значения погрешностей шума и точки; нормализуется входной сигнал для последующего попарного сравнения локальных максимумов букв в записанном и эталонном словах; в записанном слове попарно сравниваются соседние символы, а также значения u графика в пределах каждого символа.

Список литературы

1. *Springer Handbook of Speech Processing* / Jacob Benesty, M. Mohan Sondhi, Yiteng Arden Huang. – Springer-Verlag Berlin Heidelberg, 2008. – 1161 p. – ISBN 978-3-540-49125-5.
2. Vishnyakov I.E., Masyagin M.M., Odintsov O.A., Sliusar V.V. *Methods and algorithms for real time voice noise cleaning*. Proc. Univ. Electronics, 2021, vol. 26, no. 2, pp. 184–196. DOI: 10.24151/1561-5405-2021-26-2-184-196.
3. Jayashree Padmanabhan, Melvin Jose Johnson Premkumar *Machine Learning in Automatic Speech Recognition: A Survey*. Proc. IETE Technical Review, 2015, vol. 32, no. 5, pp. 240–251. DOI: 10.1080/02564602.2015.1010611.

UDC 004.934.1

ALGORITHM FOR PROCESSING DEFECTIVE SPEECH IN VOICE CONTROL OF ELECTRONIC DEVICES

Huryk K.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Stoler V.A. – Cand. of Sci., associate professor, head of the department of ECG

Annotation. An algorithm for constructing a program for recognizing keywords used in voice control of electronic devices and Internet resources is proposed. The ways to optimize the program and improve the accuracy of recognition in case of pronunciation defects are indicated.

Keywords. defective speech, sound processing algorithm, computer program