

УДК 004.934.1

АЛГОРИТМ ПОСТРОЕНИЯ ИНТЕРФЕЙСА ПРОГРАММЫ ДЛЯ РАСПОЗНАВАНИЯ ДЕФЕКТОВ СЛОВ

В. А. Столер, К. А. Гурин, С. В. Арешко

*Белорусский государственный университет информатики и
радиоэлектроники, Минск, Республика Беларусь*

Аннотация. Предложен алгоритм построения интерфейса программы для распознавания дефектов слов при их произношении, используемой в устройствах звукового управления электронными устройствами, например в системах информационной безопасности. Рассмотрены способы оптимизации программы и пути повышения точности распознавания речи.

Ключевые слова: обработка звука; дефекты произношения слов; интерфейс программы; алгоритмы обработки.

ALGORITHM FOR CONSTRUCTING A PROGRAM INTERFACE FOR RECOGNITION OF WORD DEFECTS

V. A. Stoler, K. A. Gurin, S. V. Areshko

*Educational Institution "Belarusian State University of Informatics and Radioelectronics",
Minsk, Republic of Belarus*

Abstract. An algorithm for constructing a user interface for a program for recognizing word defects during their pronunciation, used in sound control devices for electronic devices, such as in information security systems. Methods for program optimization and ways to improve speech recognition accuracy are considered.

Keywords: sound processing; pronunciation defects; program interface; processing algorithms.

Введение

Известно множество различных программ распознавания речи, в том числе и автоматическая обработка звука, однако большинство из них являются коммерческими программами. Поскольку коммерческие распознаватели речи доступны для определенных приложений, таких как произношение или транскрипция, то многие проблемы автоматического распознавания речи (ASR) при дефектах речи, в шумной среде, низкое качество записи, еще предстоит решить [1].

Основная часть

В данной статье предложен алгоритм построения программы для распознавания ключевых слов, используемых при голосовом управлении, содержащие дефекты произносимой речи. В терминах цифровой обработки сигналов процесс шумоочистки представляет собой преобразование входного сигнала, содержащего как полезный сигнал – речь, так и аддитивный сигнал-помеха – шум, в выходной сигнал, содержащий только речь. Поскольку создание систем, в точности удовлетворяющих данному условию, является довольно сложной задачей, проблему очистки сигнала от шума приходится решать следующим образом [2].

Во-первых, необходимо задать погрешность для шума, оставшегося после обработки. Определяем 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 (рис. 1).

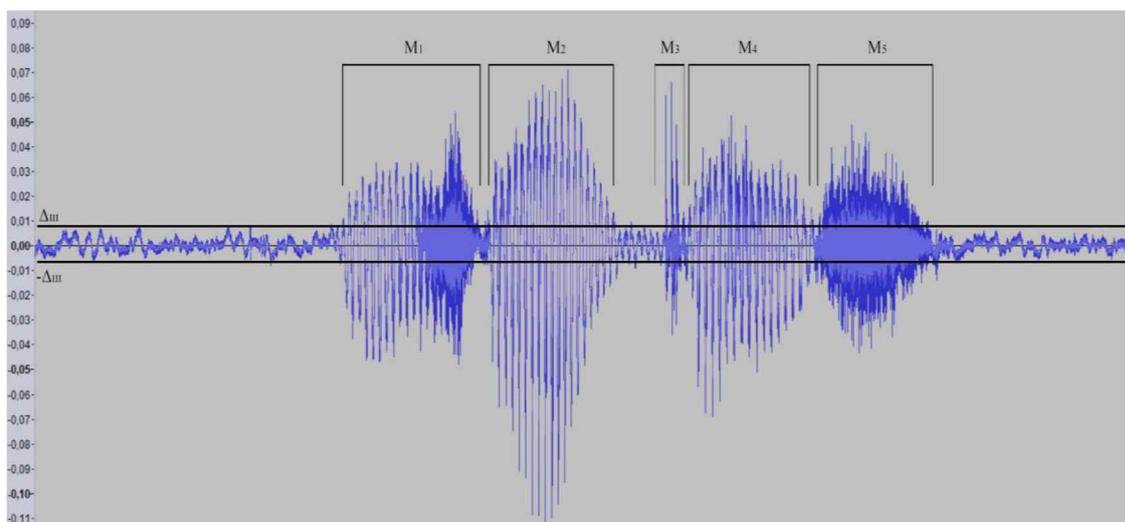


Рис. 1. Выбор интервалов M_i на графике записанной речи
Fig. 1. Selecting M_i intervals on the recorded speech graph

Во-вторых, 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 этом эталонными считаются графики слов, записанные на профессиональном оборудовании, и содержащие минимальное количество шумов. Данные условия необходимы для минимизации числа ошибок при сравнении слов.

При равенстве точек $[x_j; y_j]$ графика записанной речи точкам $[x_k; y_k]$ 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. В результате, обобщенный алгоритм программы для распознавания ключевых слов в записанной речи можно представить следующим образом (рис. 2).

Для оптимизации алгоритма необходимо нормализовать входной сигнал, т. е. задать верхнюю границу значения y . 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. о неравенстве слов [3].

Для корректного определения M_i необходимо учесть, что значение y локального максимума символа много больше заданного значения погрешности шума $\Delta_{ш}$. Изменение значений погрешности шума $\Delta_{ш}$ и погрешности точки Δ_T позволяет задать необходимую точность сравнения.

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

Кроме того, необходимо обратить внимание на такую ситуацию, при которой длина интервала M_i в записанном слове больше длины соответствующего интервала M_i в эталонном слове, т. е. количество вмещаемых интервалом точек N различно. Такое различие может возникнуть не только при неравенстве символов, но и при более

длинном произношении буквы (например, «Маамаа»). 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 .

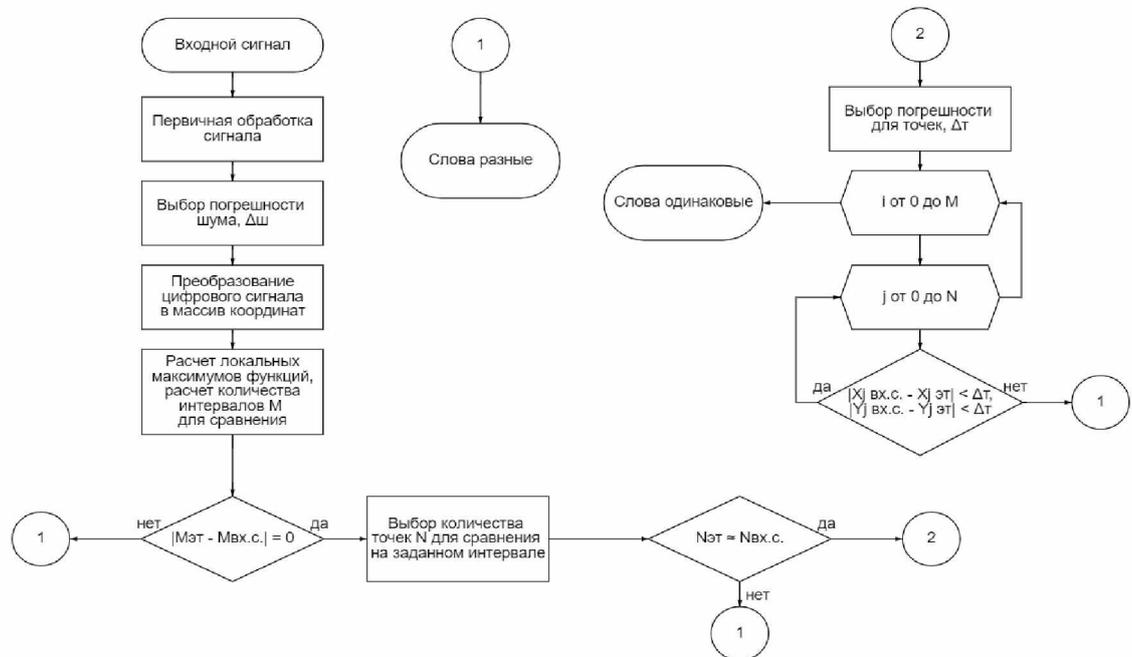


Рис. 2. Алгоритм программы для распознавания ключевых слов
Fig. 2. Algorithm of the program for recognizing keywords

Заключение

В результате исследований был выполнен анализ основных способов взаимодействия пользователя с интерфейсами, используемых в программах распознавания речи. Предложен алгоритм обработки звука, учитывающий дефекты произношения слов, для построения компьютерной программы распознавания речи, используемой в различных программах и системах, например, при звуковом управлении электронными устройствами. Рассмотрены способы оптимизации программы и повышения точности распознавания речи.

Список использованных источников / References

1. Springer Handbook of Speech Processing / Jacob Benesty, M. Mohan Sondhi, Yiteng Arden Huang. – Springer-Verlag Berlin Heidelberg, 2008. – 1161 p.
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.
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.

Сведения об авторах

Столер В.А., канд. техн. наук, доц., профессор, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», stoler@bsuir.by.

Гурин К.А., магистрант, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники».

Арешко С.В., магистрант, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники».

Information about the authors

Stoler V.A., Ph.D. (Eng.), Assoc. Prof., Professor, Educational Institution "Belarusian State University of Informatics and Radioelectronics", stoler@bsuir.by.

Gurin K.A., Master's student, Educational Institution "Belarusian State University of Informatics and Radioelectronics".

Areshko S.V., Master's student, Educational Institution "Belarusian State University of Informatics and Radioelectronics".