Ф

НОВЫЕ ПОДХОДЫ К ПРОБЛЕМАМ ОПРЕДЕЛЕНИЯ ГЛУХИХ ВЗРЫВНЫХ ЗВУКОВ В КОНЦЕ ЗАПИСАННОГО СЛОВА

Акопян Артем Геннадиевич[1], Костенко Александр Владимирович[2],
Шелепов Владислав Юрьевич[3]

Сформулирована одна из актуальных проблем, связанная с сегментацией речевого сигнала на отдельные звуки в компьютерных системах распознавания речи. Рассмотрены основные алгоритмы сегментирования речевого сигнала. Предложен метод выделения глухих взрывных звуков русского языка в конце записанного речевого сигнала, что помогает более точно определить конец записи некоторой группы слов.

Общая постановка проблем

Речевой сигнал затухает постепенно. Поэтому компьютер может ошибаться в вопросе заканчивается ли слово гласным или звонким согласным. Далее, сигнал, содержащий глухой взрывной звук (к, п, т) в середине слова, имеет характерный паузообразный отрезок, поскольку при его произнесении происходит полное перекрытие речевого тракта и не участвуют голосовые связки. Если глухой взрывной будет находиться в конце слова, то определить правильные границы записанного слова и выделить конечный звук становится весьма сложной задачей, так как конец слова мало отличается от молчания. В настоящей работе предлагается способ решения этой проблемы.

Запись речевого сигнала

Алгоритм записи речевого сигнала описан в работе [1]. Сейчас мы хотим несколько видоизменить его. А именно, мы сохраним в сигнале 10000 отсчетов после момента, который в [1] описан как конец сигнала. В результате получим сигнал следующего вида (рис. 1).

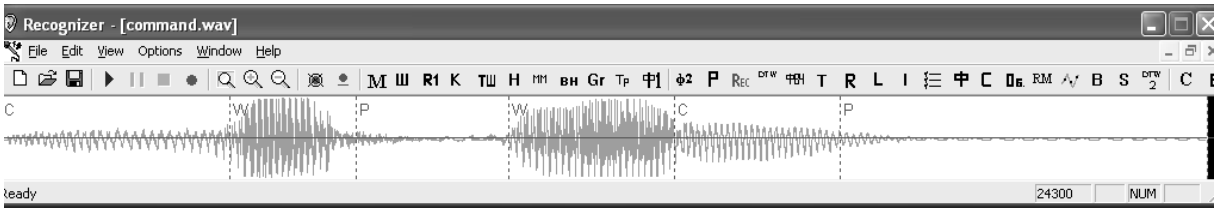

Рисунок 1. Визуализация записи слова «ЗАКОН»

Сигнал на рисунке отсегментирован. Использование паузы в конце сигнала позволяет с помощью алгоритмов сегментации [1-2] надежно различать случаи, когда слово оканчивается гласным или звонким согласным звуком. Приведем упомянутые алгоритмы.

«В-Н» - обработка числового массива

Пусть имеется одномерный числовой массив и задан некоторый порог p. Построим

[1] магистрант, тел. +38(050)705-56-41
[2] магистрант, тел. +38(050)950-75-90
[3] д.ф.-м.н., проф. кафедры систем искусственного интеллекта, Донецкий национальный технический университет Институт проблем искусственного интеллекта НАН и МОНМС Украины

символьную последовательность S, поставив в соответствие членам массива, которые больше p, символ «В» (выше порога), остальным – символ «Н» (ниже порога).

Для того чтобы устранить случайные единичные включения, для каждого промежуточного i-го элемента полученной символьной последовательности S выполняются две дополнительные обработки. Обработка «тройками», если $s[i\text{-}1] = s[i+1]$ и $s[i] \neq s[i\text{-}1]$, то полагается $s[i] = s[i\text{-}1]$. Обработка «четверками», если $s[i] = s[i+3]$ и $s[i+1] \neq s[i]$, $s[i+2] \neq s[i]$, то полагается $s[i+1] = s[i]$ и $s[i+2] = s[i]$.

Выделение глухих согласных

Этот этап сегментации осуществляется с помощью обработки сигнала полосовым фильтром с полосой пропускания от 100 до 200 Гц. Глухие звуки отличаются от всех остальных тем, что после такой фильтрации их участки становятся подобными паузе и содержат большое число точек постоянства (в следующий дискретный момент значение сигнала не меняется). Таким образом, на этих участках разность между числом точек непостоянства и числом точек постоянства будет отрицательной, что позволяет выделить их в массиве таких разностей, построенном для последовательности окон в 256 отсчетов.

Распознавание в паре классов «шипящая-пауза»

Рассмотрим для произвольно выделенного участка речевого сигнала численный аналог полной вариации «с переменным верхним пределом»:

$$V(0) = 0, V(n) = \sum_{i=0}^{n-1} \left| x_{i+1} - x_i \right| . \quad (1)$$

Пусть N_1 – максимальное число, такое, что $W(N_1) \leq 255$. Полагаем

$$W(n) = V(n) \text{ при } 0 \leq n \leq N_1,$$
$$W(N_1 + 1) = 0, W(n) = \sum_{i=N_1+1}^{n-1} \left| x_{i+1} - x_i \right| \text{ при } N_1 + 1 \leq n \leq N_2, \quad (2)$$

где N_2 – максимальное число, такое, что $W(N_2) \leq 255$ и так далее. Возникает массив чисел

$$N_1, N_2 - N_1, N_3 - N_2, \ldots \quad (3)$$

На сегменте шипящей величина (1) быстро растет, поэтому участки возрастания величины $W(n)$ от 0 до 255 относительно коротки, то есть числа (3) относительно малы. На сегменте паузы величина (1) растет медленно, и поэтому числа (3) относительно велики. Для различения шипящей и паузы введем порог p (для нашего оборудования 200). Возьмем выделенный сегмент глухих согласных и построим для него последовательность чисел (3). Те участки, для которых числа (3) превосходят p, относим к паузе (их объединение маркируем символом P), остальные – к шипящей (маркируем ее символом F). В результате компьютер расставит маркированные границы шипящих и пауз.

Сегментация чисто голосового сигнала

Рассмотрим случай слова, не содержащего глухих звуков. Разобьем сигнал на окна по 256 отсчетов, и на каждом из них вычислим значение вариации

$$V = \sum_{i=0}^{254} \left| x_{i+1} - x_i \right| . \quad (4)$$

Далее от начала слова берется интервал из 20 таких окон и вычисляется среднее значение соответствующих величин (4), которое принимается за порог. Производится «В-Н»-обработка числового массива с этим порогом. Затем интервал, на котором выполняются описанные процедуры, сдвигается вправо на одно окно и так далее. В результате возникает таблица вида, изображенного на рис. 2.

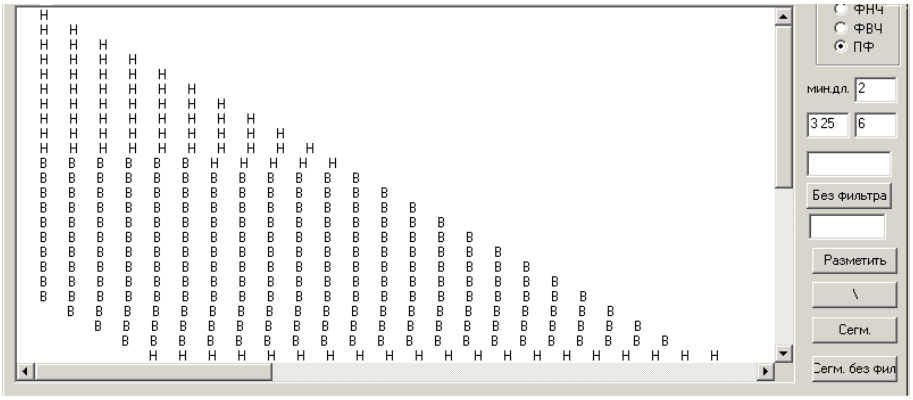

Рисунок 2. Таблица, используемая при сегментации

Затем просматриваются все строки полученной таблицы и создается новая символьная последовательность *S*. Если текущая *i*-я строка таблицы начинается и заканчивается одним и тем же символом («Н» или «В»), то в S на *i*-ю позицию записывается соответствующий символ. Иначе считается количество вхождений каждого из символов в данной строке. Если количество «В» превышает количество «Н» или равно ему, то в S на соответствующую позицию записывается «В», иначе «Н». К полученной последовательности применяется «В-Н»-обработка. Метки сегментации ставятся там, где происходит смена символов «Н» на «В», или «В» на «Н». В-участок считается соответствующим гласному (возле левой метки проставляется символ *W*). Н-участок считается соответствующим звонкому согласному (возле левой метки проставляется символ С).

Сегментация при наличии шипящих и пауз

Если слово содержит шипящие или паузы, то мы выделяем их, как описано выше, после чего значения величины (4) для соответствующих им окон полагаем равными нулю и сегментируем сигнал только что описанным способом (шипящие и паузы автоматически попадают в число Н-участков). Для надежного выделения звонкого согласного непосредственно после шипящего или паузы порядок формирования *S* непосредственно после шипящего или паузы меняется: если в строке появляется «В», но она заканчивается на «Н», то ей сопоставляется «Н». Аналогичная ситуация с голосовым согласным непосредственно перед шипящей или паузой.

Эволюционный метод выделения глухого взрывного звука в конце слова

В разработанном нами программном обеспечении предложенный метод, описанный ниже, показал следующие результаты. Пусть произнесено слово «ЗАКОН», заканчивающееся звонким согласным. Визуализация соответствующего сигнала

приведена на рис. 1 с сегментацией в соответствии с выше описанными алгоритмами. Построим функцию $W(n)$ (рис. 3).

На рис. 5 показан результат вычисления массива (3).

Разработанное нами программное обеспечение поддерживает соответствие между выделением строки в списке рис. 5 и положением курсора на рис. 4 (где представлен тот же график, что и на рис. 3). Большие числа в конце списка соответствуют участку молчания, записанного в конце сигнала. Движемся по списку снизу вверх, проходя строки, числа в которых больше порога p_1 (мы берем этот порог равным 1000). Выделяем строку, для которой число в предыдущей строке уже меньше p_1. Выделенной строке соответствует положение курсора на рис. 4. Это предполагаемый конец речевого сигнала.

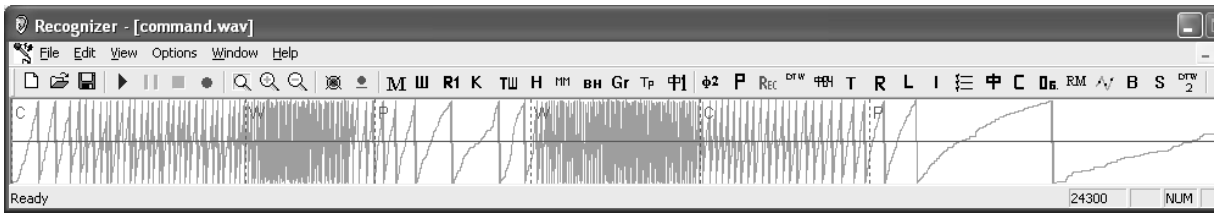

Рисунок 3. График функции *W(n)*, соответствующей сигналу на рис. 1

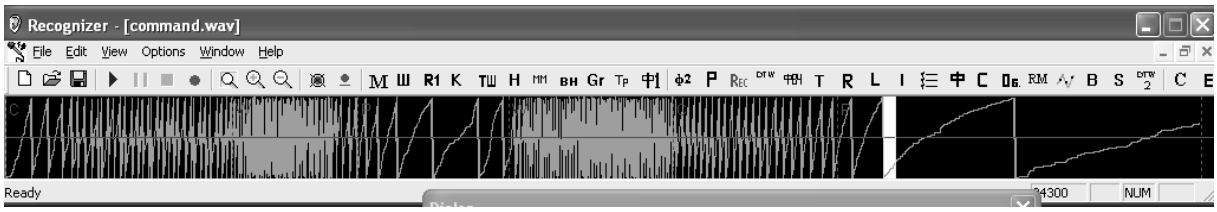

Рисунок 4. Положение курсора, определяющее предполагаемый конец сигнала

Продолжаем движение по списку снизу вверх пока левый край курсора на рис. 5 не совпадет с меткой *P,* или впервые не окажется левее нее. Суммируем все промежуточные числа списка и сравниваем вычисленную сумму *Sum* с порогом p_2 (мы берем его равным 3000). В данном случае она оказывается меньше p_2. Поэтому мы считаем метку *P* концом сигнала и удаляем маркировку *P*. Размеченная визуализация сигнала приведена на рис. 6:

Теперь произнесем слово «РОТ». Его визуализация с окончательной разметкой приведена на рис. 6, а рис. 7 содержит график функции $W(n)$ с курсором в позиции предполагаемого конца сигнала.

Вычисляем сумму *Sum,* так же, как это сделано выше. В данном случае она оказывается больше порога p_2 (курсор на рис. 8 отстоит от метки *P* много дальше, чем в предыдущем примере). Поэтому истинным концом речевого отрезка мы полагаем позицию левого края курсора на рис. 8. Сегмент от метки *P* до этой новой метки конца речевого отрезка – порождение глухого взрывного звука в конце слова.

Выводы

В работе решена задача связанная с определением глухих взрывных согласных звуков русского языка находящихся в конце слова.

Научная новизна работы заключается в том, что предложен новый эволюционный

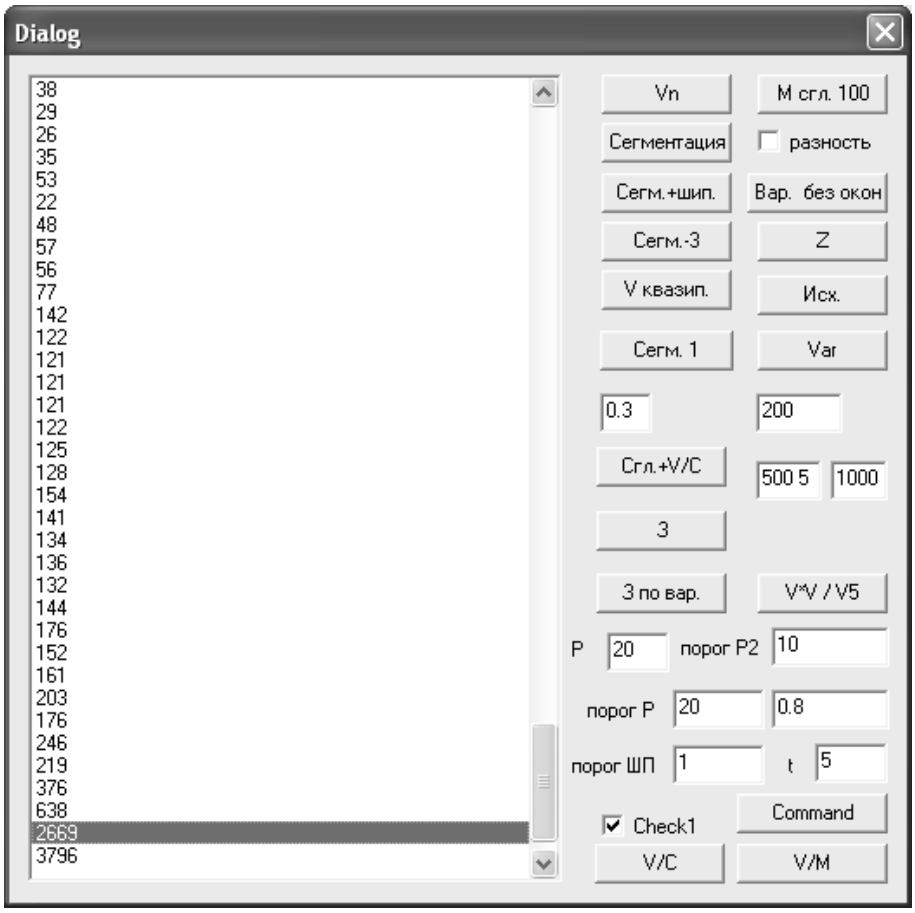

Рисунок 5. Список в левой части окна представляет массив (3)

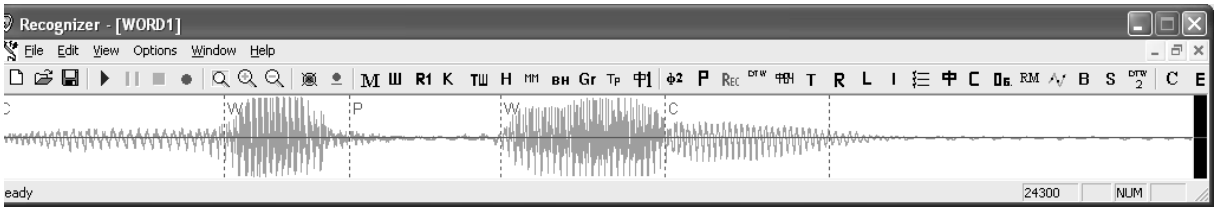

Рисунок 6. Окончательная разметка сигнала с отмеченным концом сигнала

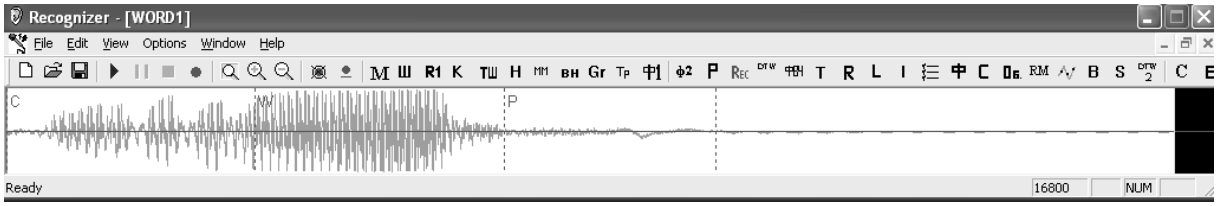

Рисунок 7. Визуализация сигнала для слова «РОТ» с окончательной разметкой

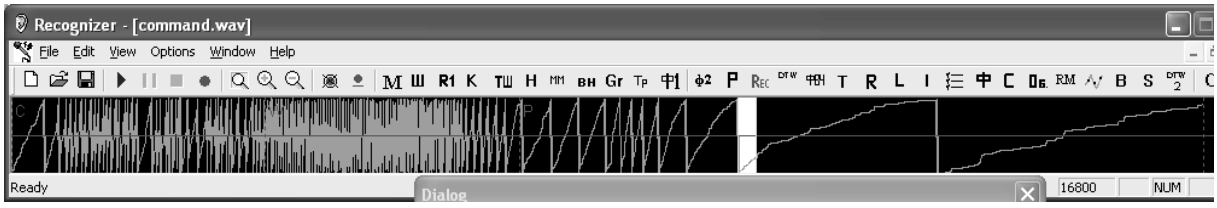

Рисунок 8. График функции *W(n)* с курсором в позиции предполагаемого конца сигнала

метод, предназначенный для различения случаев, когда слово оканчивается гласным или звонким согласным звуком и для обнаружения глухого взрывного звука в конце слова, что посредством этого позволяет также более точно определять конец слова в записанном сигнале.

Практическая ценность результатов работы заключается в том, что: 1) было разработано соответствующее программное обеспечение предназначенное для записи и сегментации речевого сигнала включившее в себя все выше указанные алгоритмы; 2) решена задача определения глухих взрывных звуков в конце слова; 3) решение сформулированной задачи было подтверждено рассмотренными в статье примерами записанных слов «ЗАКОН» и «РОТ», что позволяет увидеть переход от начально установленной метки конца слова к новой, захватывающей глухую согласную в конце слова.

Список источников

1. Шелепов В.Ю. Лекции о распознавании речи / В.Ю. Шелепов. – Д.: ІПЩІ «Наука і освіта», 2009. – 196 с.
2. Шелепов В.Ю., Ниценко А.В., Жук А.В. Построение системы голосового управления компьютером на примере задачи набора математических формул // Искусственный интеллект. – 2010. – № 3. – С. 259–267.
3. Методы пофонемного распознавания, использующие свойства языка и речи [Електронный ресурс] / Г.В. Дорохина // Искусственный интеллект – 2008. – № 4. С. 332–338 – Режим доступа к ресурсу: http://www.nbuv.gov.ua/ portal/natural/ii/2008_4/JournalAI_2008_4.