# Новые подходы к проблемам конца речевого сигнала

 Шелепов В.Ю., Акопян А.Г., Ниценко А.В., Костенко А.В.

Общая постановка проблемы

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 предлагаются способы решения этих проблем.

1. Запись речевого сигнала

Алгоритм записи речевого сигнала описан в [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).

Рис. 1. Визуализация записи слова «ЗАКОН»

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

2. «В-Н» - обработка числового массива

Пусть имеется одномерный числовой массив и задан некоторый порог р. Построим символьную последовательность S, поставив в соответствие членам массива, которые больше р, 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», еслиs[i-1] = s[i+1] иs[i] ≠ s[i-1], тополагаетсяs[i] = s[i-1]. Обработка«четверками», еслиs[i] = s[i+3] иs[i+1] ≠ s[i], s[i+2] ≠ s[i], тополагаетсяs[i+1] = s[i] иs[i+2] = s[i].

3. 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.

4. Распознавание в паре классов «шипящая-пауза»

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

                       (1)

Пусть N1 – максимальное число, такое, что W(N1) ≤ 255. Полагаем

               (2)

где N2 – максимальное число, такое, что W(N2) ≤ 255 и так далее. Возникает массив чисел

                          (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.

5. 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

                             (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 рисунке 2.

Рис. 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).

6. 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». Дальше все как описано выше. Аналогичная ситуация с голосовым согласным непосредственно перед шипящей или паузой.

7. Определение конца слова. Обнаружение и выделение глухого взрывного звука в конце слова

Пусть произнесено слово «ЗАКОН», заканчивающееся звонким согласным. Визуализация соответствующего сигнала приведена на рисунке 1 с сегментацией в соответствии с только что описанными алгоритмами. Построим функцию (рис. 3).

Рис. 3. График функции W(n), соответствующей сигналу на рисунке 1

Рис. 4. Положение курсора, определяющее предполагаемый конец сигнала

На рисунке 5 показан результат вычисления массива (3).

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

Разработанная нами программа поддерживает соответствие между выделением строки в списке рисунка 5 и положением курсора на рисунке 4 (где представлен тот же график, что и на рисунке 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 p1 (мы берем этот порог равным 1000). Выделяем строку, для которой число в предыдущей строке уже меньше p1. Выделенной строке соответствует положение курсора на рисунке 4. Это предполагаемый конец речевого сигнала.

Продолжаем движение по списку снизу вверх пока левый край курсора на рисунке 5 не совпадет с меткой P, или впервые не окажется левее нее. Суммируем все промежуточные числа списка и сравниваем вычисленную сумму Sum с порогом p2 (мы берем его равным 3000). В данном случае она оказывается меньше p2. Поэтому мы считаем метку P концом сигнала и удаляем маркировку P. В результате размеченная визуализация сигнала выглядит так:

Рис. 6. Окончательная разметка сигнала рисунка 1 с отмеченным концом сигнала

Теперь произнесем слово «РОТ». Вот его визуализация с окончательной разметкой:

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

Вот график функции W(n) с курсором в позиции предполагаемого конца сигнала

Рис. 8. 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 p2 (курсор на рисунке 8 отстоит от метки P много дальше, чем в предыдущем примере). Поэтому истинным концом речевого отрезка мы полагаем позицию левого края курсора на рисунке 8. Сегмент от метки P до этой новой метки конца речевого отрезка - порождение глухого взрывного звука в конце слова.

Выводы

Модификация способа классификации последнего звонкого звука речевого сигнала и алгоритмы пункта 7 являются новыми. Они обеспечивают новые возможности классификации заключительного звука в сигнале по сравнению с алгоритмами, описанными в [1].

Список литературы

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 – Режим доступа к ресурсу: http://www.nbuv.gov.ua/portal/natural/ii/2010\_3/AI\_2010\_3%5C3%5C00\_Shelepov\_ Nitsenko\_Zhuk.pdf.

3. Методы пофонемного распознавания, использующие свойства языка и речи [Электронный 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/Razdel4/06\_Dorokhina.pdf.