УДК 534.78 : 519.254

А. Н. ГАВРАШЕНКО, М. Ф. БОНДАРЕНКО

О МЕТОДЕ АВТОМАТИЧЕСКОГО ВЫДЕЛЕНИЯ ГРУППЫ СОГЛАСНЫХ РУССКОГО ЯЗЫКА

Одним из главных направлений научных исследований автоматического распознавания речи является пофонемное распознавание [1]. При разработке этого направления возникают трудности, которые объясняются двумя причинами:

1) сложным характером речевых сигналов, зависящих не только от произносимых звуков, но и от контекста, отчетливости, громкости, темпа и интонации произнесения, а также индивидуальных особенностей и состояния говорящего;

2) трудностью использования необходимых для опознавания речи лингвистических данных.

Поэтому многие исследователи в области речи переходят от попыток опознавания всех фонем к опознаванию некоторых отдельных фонем.

Настоящая работа посвящена вопросу автоматического распознавания группы шумных согласных звуков русского языка (С, З, Ц, Ч, Ш, Ж, Ф, Щ, Х'), содержащихся в составе слов, произносимых произвольными дикторами (без особой подготовки). Предлагаемый алгоритм опознавания работает в реальном масштабе времени. Назовем указанную группу согласных классом распознавания.

Большинство работ по исследованию шумных звуков основано на спектральном анализе речи с привлечением информации об интенсивности речевого сигнала (РС). Исследователи Хьюз и Халле строили метод опознавания звуков [2] на использовании отношений энергии в трех парах частотных полос речевого сигнала. Цемель Г. И. [1] упоминает о методе опознавания начальных щелевых *s*, *j*, *f*, *h*, *v*, *z* посредством разделения акустического сигнала на 19 каналов с помощью полосовых фильтров с низкой добротностью. Анализ шумных звуков осуществлялся также по клиппированной речи с использованием таких характеристик, как кривые плотности распределения длительностей интервалов между нулевыми пересечениями, усредненная плотность нулей [1]. Многие зарубежные исследователи в своих системах по автоматическому опознаванию речи широко используют параметр скорости перехода сигнала через временную ось как одну из важнейших акустических характеристик [3]. Надежность опознавания исследуемых шумных звуков в указанных работах изменялась в пределах от 86 до 99,2 %.

Данные по опознаванию щелевых, приведенные в опубликованных работах, основаны на анализе сравнительно небольшого числа реализаций звуков в словах. Для получения более достоверных данных необходимо при анализе и проверке информативности признаков применять более разнообразный речевой материал, включающий многосложные слова и все типы встречающихся сочетаний щелевых с другими звуками (в том числе в неударных слогах), используя большое количество дикторов.

В данной работе предлагается метод выделения группы указанных выше шумных звуков, основанный на использовании временных параметров РС таких, как частота перехода РС через временную ось, средняя плотность нулевых переходов в слове и в фонеме, длительность звука. Данный алгоритм реализован на ЭВМ серии ЕС и не требует сложных специализированных устройств как для первичной обработки акустического сигнала, так и для последующего его анализа с целью опознавания шумных звуков.

РС вводится в ЭВМ с микрофона типа МД-68 в условиях шумов машинного зала ВЦ. Параметрический код дискретизи-

рованного с частотой 38 кГц акустического сигнала запоминается в оперативной памяти. Первичная обработка речевого сигнала, заключающаяся в его нормировании и преобразовании в бинарный код, выполнена программно.

Амплитудно-временное описание РС содержит почти всю необходимую информацию о речи [1]. Поэтому амплитудные отсчеты, взятые на каждом шаге квантования, и частота переходов через временную ось могут служить признаками акустического сигнала. В предлагаемом алгоритме широко используются данные, содержащиеся в нулевых переходах РС. Как известно [1], большая часть речевой информации содержится в согласных звуках, большинство которых обладает шумовой составляющей. Акустическая характеристика скорости перехода речевого сигнала через нуль хорошо характеризует шумные звуки и качественно, и количественно, обеспечивает их существенное отличие от всех остальных звуков русского языка.

Под понятием частоты переходов через нуль акустического сигнала $F_{\text{п}}$ рассматривается количество пересечений речевым сигналом временной оси за какой-либо интервал времени. В точке пересечения сигнал изменяет свой знак на противоположный. Частота переходов через нуль РС вычислялась на сегментах длиной 10 мс. Этот интервал был выбран из следующих соображений. Речевой тракт — это акустическая труба с неодинаковой по продольной оси площадью поперечного сечения. Изменив продольный профиль трубы, можно получить другую звуковую волну на выходе у рта. Профиль акустической трубы меняется медленно. Его можно считать неизменным за интервал 10 мс. Шумовой источник тоже является неизменным. В результате вычисления значений $F_{\text{п}i}$ для каждого сегмента на интервале всего слова получаем представление любого, сказанного в микрофон слова, в виде числовой последовательности

$$F_{\text{п}1},\ F_{\text{п}2},\ F_{\text{п}3},\ \ldots,\ F_{\text{п}N},$$

где N — количество сегментов в слове.

Для наглядности было использовано графическое представление частоты переходов через нуль в слове $F_{\text{п}}$ как функцию времени. В качестве примера такое представление слова «СУША» показано на рис. 1.

В описываемом алгоритме опознавания шумных звуков используется метод сечения, заключающийся в том, что график $F_{\text{п}}$ в слове разрезается на определенном уровне сечением. Для определения уровня проведения сечения определяется среднее значение частоты переходов через нуль в слове

$$\text{МО} = \frac{\sum_{i=1}^{N} F_{\text{п}i}}{N}. \qquad (1)$$

Данный уровень (далее будем обозначать его символом МО) выбран не случайно, так как значение МО является усредненной

характеристикой всего слова, и относительно этого уровня можно судить о том, присутствуют ли в слове фонемы, отличающиеся от остальных фонем величиной параметра $F_п$ на своем протяжении (рис. 1). Как показали исследования, если в слове присутствует шумный звук (один или несколько) из состава класса распознавания, то значение $F_п$ на всей длине звука превосходит уровень сечения и ни на одном сегменте не опускается ниже этого уровня. Значение параметра $F_п$ на каждом сегменте этих звуков по величине превосходит 60. Это характерно для всех фонем рассматриваемой группы и является стабильной харак-

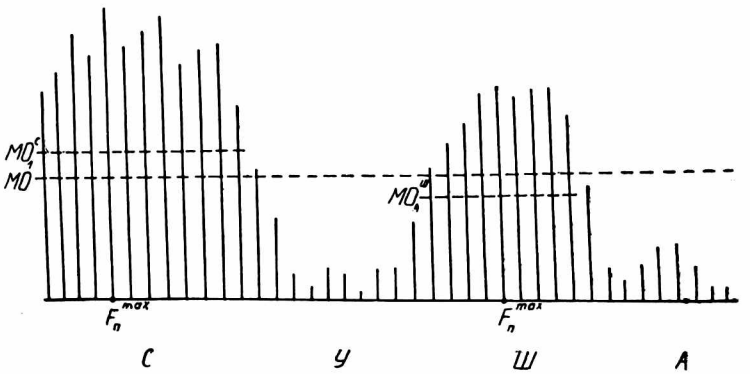

Рис. 1. Географическое представление слова «СУША»

теристикой, позволяющей отличать эти фонемы от всех остальных звуков русского языка.

Однако существует ряд фонем (не входящих в указанную последовательность), которые на определенном участке своего звучания дают всплеск частоты переходов через нуль, при этом значение $F_п$ на сегментах может достигать 60 и выше. Это такие фонемы, как Д', К, Т', Х, сочетание фонем ВТ. Эти всплески имеют довольно случайный характер и в зависимости от произношения могут иметь место или отсутствовать. Как указывалось выше, всплески значения параметра $F_п$ в указанных фонемах возникают лишь на отдельных, рядом стоящих сегментах этих фонем.

Таким образом, длина участка фонемы не сравнима с протяженностью самой фонемы. Эту особенность можно использовать в классификации всплесков частоты переходов через нуль, соответствующих различным фонемам в слове. Помимо звуков Д', К, Т', Х и сочетания ВТ, случайные всплески $F_п$, превышающие уровень сечения МО, могут давать и некоторые другие фонемы. Исследовав характер этих выбросов на интервале фонемы в словах с шумными звуками, пришли к выводу, что такие всплески существуют одновременно не более чем на трех

рядом расположенных сегментах, затем следует провал значения $F_п$, опускающийся ниже уровня МО в слове (рис. 2).

Из всего сказанного выше вытекает, что для фонем С, З, Ц, Ч, Ш, Ж, Ф, Щ, Х' всплеск частоты переходов через нуль, превышающий уровень сечения МО, имеет стабильный характер на протяжении всей фонемы и ни на одном сегменте не опускается ниже указанного уровня, независимо от того, сколько слогов имеет слово и сколько шумных звуков входит в его состав. При этом значение $F_п$ на сегментах фонемы превосходит 60. Для всех остальных звуков русского языка, входящих в состав слова,

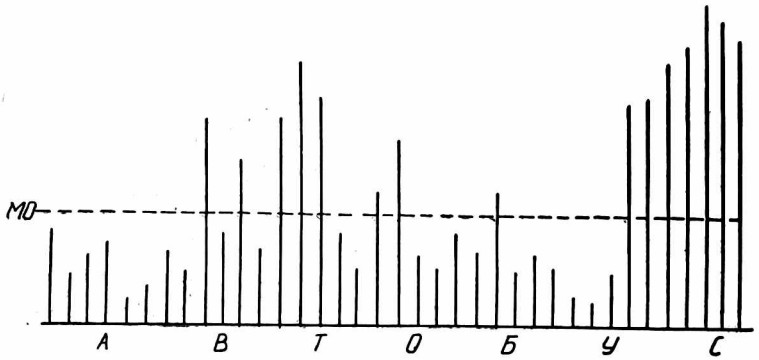

Рис. 2. Географическое представление слова «АВТОБУС»

значение параметра $F_п$ на их протяжении либо вообще не превышает уровень сечения МО, либо, если это случается, происходит не более чем на трех рядом стоящих сегментах фонемы, после чего обязательно следует провал $F_п$, опускающийся ниже уровня сечения МО в слове (рис. 2).

Такая довольно несложная операция выделения шумных звуков в словах выполняется алгоритмом, описываемым в данной работе, если в слове присутствуют шумные звуки из класса распознавания. Если же в анализируемом слове указанные звуки отсутствуют, то, естественно, уровень сечения в слове МО проходит на довольно незначительной высоте, так как фонемы слова имеют малые значения $F_п$ и не отличаются по этому параметру друг от друга. Таким образом, сечение МО может пересекать значения $F_п$ на протяжении всей фонемы. Этот случай иллюстрируется рис. 3. Алгоритм выполняет следующие шаги анализа предложенного слова. По признаку превышения уровня сечения МО в слове выделяются границы шумных звуков, причем выбираются только те звуки, на сегментах которых значение $F_п$ не опускается ниже МО. Осуществляется подсчет количества рядом стоящих сегментов, имеющих значение $F_п$ большее, чем уровень МО. Если это количество не меньше 7, то выделяется этот участок слова. Начальной границей участка

будет начало первого из выделенных сегментов, конечной — конец последнего из выделенных сегментов. Затем находится максимальное значение $F_п$ на этом участке $F_п^{max}$. На выделенном отрезке слова на уровне

$$МО_1 = \frac{F_п^{max}}{2} \qquad (2)$$

проводится еще одно сечение (рис. 1). Алгоритм выполняет операцию анализа, подобно той, что и на интервале слова, что-

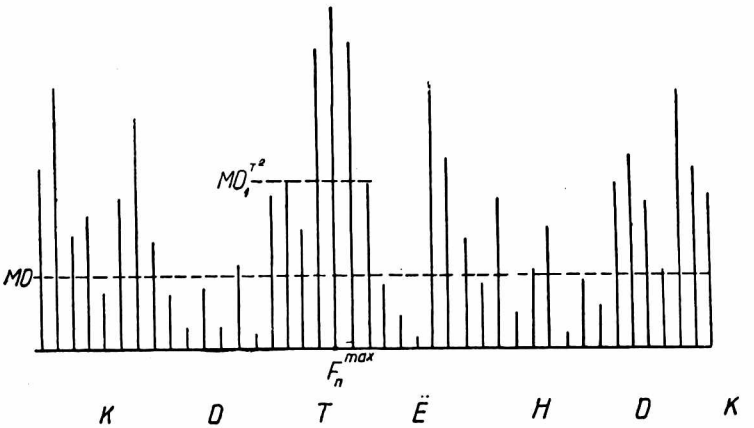

Рис. 3. Географическое представление слова «КОТЕНОК»

бы определить, на каких сегментах звука параметр $F_п$ превышает уровень $МО_1$. В результате исследований установлено, что если в слове присутствует шумный звук из класса распознавания, то на его протяжении есть не менее 7 рядом стоящих сегментов, частота переходов через нуль которых больше уровня сечения $МО_1$ в звуке. Для всех остальных звуков это условие не выполняется. Это можно показать на примере слова «КОТЕНОК», представленного графически на рис. 3. Проведенное сечение на уровне МО выделяет по признаку пересечения $F_п$ на протяжении не менее 7 рядом стоящих сегментов участок слова, содержащий фонему Т'. Вычислим значение $F_п^{max}$ на этом участке и найдем уровень сечения $МО_1$ по формуле (2). На следующем шаге алгоритм выполняет анализ на выделенном участке с целью определения количества рядом расположенных сегментов, на которых $F_п$ превышает уровень $МО_1$. Как видно из рисунка, количество сегментов равно 3. Алгоритм отнес бы этот участок слова к классу шумных звуков, если бы количество этих сегментов было не менее 7 и значение $F_п^{max}$ на нем превышало бы 60. Но так как эти условия не выполняются,

то алгоритм относит выделенный отрезок слова к группе звуков, не входящих в состав класса распознавания.

В результате экспериментов оказалось, что для классификации звуков русского языка важное значение имеет параметр максимального значения частоты переходов через нуль $F_{\text{п}}^{\max}$ на интервале звука. По значению $F_{\text{п}}^{\max}$ все фонемы можно разбить на три группы: 1 — С, З, Ц; 2 — Ч, Ш, Ж, Ф, Щ, Х'; 3 — все остальные звуки русского языка.

Область значений для $F_{\text{п}}^{\max}$ первой группы лежит в пределах от 117 до 180. Для второй группы — от 60 до 116. Для третьей группы — ниже 60.

В ходе экспериментов замечено, что значения $F_{\text{п}}^{\max}$ для второй группы могут иметь случайные небольшие по величине выбросы, в результате которых $F_{\text{п}}^{\max}$ попадает в область значений первой группы звуков. В связи с этим предлагаемый алгоритм производит разбиение на группы исследуемого класса фонем не по признаку $F_{\text{п}}^{\max}$, а по некоторой усредненной характеристике, вычисляемой по формуле

$$F_{\text{п}}^{\text{ср}} = \text{МО}_1 + \frac{(F_{\text{п}}^{\max} - \text{МО}_1)\,2}{3}. \qquad (3)$$

Предлагаемый алгоритм по признаку $F_{\text{п}}^{\text{ср}}$ все фонемы русского языка однозначно разбивает на три группы (рис. 4).

На предлагаемый для анализа словарь накладывают определенные ограничения, которые заключаются в требовании, чтобы в предлагаемом слове не было рядом расположенных шумных звуков, входящих в класс распознавания.

В качестве промежуточного этапа предлагаемый алгоритм выполняет операцию уточнения границ выделяемых шумных звуков. Это связано с тем, что для звуков из второй группы значение $F_{\text{п}}$ на их протяжении бывает близко к границе 60 (рис. 4). Поэтому если в слове не имеется фонем из 1-й группы, а есть только из 2-й и рядом с ней стоит звук типа К или Т', которые имеют значения параметра $F_{\text{п}}$ в пределах $30 \leqslant F_{\text{п}} \leqslant 50$, то сечение на уровне МО может вырезать границы шумного звука, начало которого будет совпадать с началом звука из группы 2, а конец — с концом рядом стоящего с ним звука К или Т'. В связи с этим алгоритм осуществляет корректировку ранее найденных границ шумного звука. Для этого используется две ранее найденные характеристики. Первая из них $DL1$ обозначает количество рядом стоящих сегментов, соответствующих длине шумного звука, на которых значение $F_{\text{п}}$ превышает уровень МО в слове. Вторая $DL2$ — количество рядом стоящих сегментов внутри уже найденных границ, на которых значение $F_{\text{п}}$ превышает уровень МО_1 в шумном звуке. Если разность этих двух характеристик $RAZN = DL1 - DL2$ не более 4, то в

необходимо иметь признаки, на базе которых можно отличать фонемы одной группы от остальных звуков. Для этого используются следующие признаки:

f, характеризующий значение параметра $DL1$; по этому признаку на интервале анализируемого слова выделяются характерные шумные участки;

h, характеризующий значение параметра $DL2$; по этому признаку все фонемы, отнесенные в группу шумных по признаку

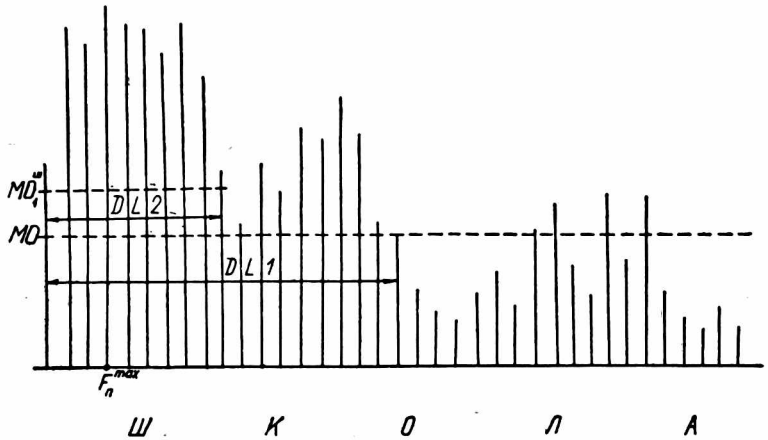

Рис. 5. Географическое представление слова «ШКОЛА»

f, делятся на две группы: *1* — фонемы из класса распознавания и *2* — фонемы, не относящиеся к классу распознавания;

d, характеризующий значение параметра $F_п^{ср}$; по этому признаку звуки русского языка делятся на три группы (рис. 4).

Используя значения параметров $DL1$, $DL2$ и $F_п^{ср}$, характерных для каждой из трех групп (рис. 4), запишем уравнения на языке алгебры конечных предикатов, описывающие каждую из групп.

Для звуков С, З, Ц, объединенных в группу *1*, параметры $DL1$, $DL2$ и $F_п^{ср}$ принимают следующие значения:

$$DL1 \geqslant 7;$$

$$DL2 \geqslant 7;$$

$$115 \leqslant F_п^{ср} \leqslant 180.$$

Уравнение, описывающее эту группу фонем, имеет вид

$$A = \overline{Men\,(f,\ 7)} \wedge \overline{Men\,(h,\ 7)} \wedge Mp\,(d,\ 180) \wedge \overline{Men\,(d,\ 115)}. \qquad (4)$$

Для звуков Ч, Ш, Ж, Ф, Щ, Х', объединенных в группу *2*, параметры $DL1$, $DL2$, $F_п^{ср}$ принимают значения

$$DL1 \geqslant 7;$$

$$DL2 \geqslant 7;$$

$$60 \leqslant F_{\text{п}}^{\text{ср}} < 115.$$

Эту группу фонем описывает уравнение

$$B = \overline{Men\,(f,\ 7)} \wedge \overline{Men\,(h,\ 7)} \wedge Men\,(d,\ 115) \wedge \overline{Men\,(d,\ 60)}. \quad (5)$$

Для группы звуков, не входящих в класс распознавания (группа 3), уравнение будет иметь вид

$$C = Men\,(f,\ 7) \vee Men\,(h,\ 7) \vee Men\,(d,\ 60). \quad (6)$$

Таким образом, получили систему математических уравнений, описывающую (в удобном для счета на ЭВМ виде) три группы фонем русского языка: *1* — С, З, Ц; *2* — Ч, Ш, Ж, Ф, Щ, Х'; *3* — все остальные звуки русского языка.

Решив эту систему на ЭВМ, получаем однозначный ответ на вопрос, к какой группе звуков относится выделенная по описанному в данной статье алгоритму та или иная фонема русского языка.

Список литературы: 1. *Цемель Г. И.* Опознавание речевых сигналов. — М.: Наука, 1971. — 142 с. 2. *Hughes G. W., Halle M.* Spectral properties of fricative consonants. — JASA, 1956, 28, p. 303—310. 3. *Paliwal K. K.* An isolated word recognition system for Hindi digits using linear time normalization. — *J.* Instn. Electronics and Telecom. engrs., 29, N 1, 1983, p. 18—22. 4. *Шабанов-Кушнаренко Ю. П.* Начала теории интеллекта: Технические средства. — Деп. ВИНИТИ № 3323—82. — 245 с.

Поступила в редколлегию 04. 01. 85.