УДК 621.396.961.1

Е. Е. Федоров

МЕТОД ОПРЕДЕЛЕНИЯ ГРАНИЦ ИЗОЛИРОВАННОГО СЛОВА В РЕЧЕВОМ СИГНАЛЕ

1. Введение

Постановка проблемы. В современной отечественной и мировой практике активно ведутся разработки естественно-языковых систем общения «человек–компьютер», одной из составных частей которых может быть система распознавания речи. При создании системы распознавания необходимо решить задачу определения границ речи.

Анализ исследований. В работах [1–3], посвященных распознаванию речи, рассматриваются математические модели и методики, не учитывающие при определении границ речи шумы аппаратной части и параметры речи диктора.

2. Постановка задачи и предварительная обработка сигнала

Постановка задачи. Разработать метод определения границ изолированного слова в речевом сигнале.

Основной материал. Для границ речи в статье реализован авторский метод ДАРФ, основывающийся на результатах работы [4].

Акустический сигнал, поступающий через микрофон и звуковую карту в систему распознавания речи, проходит усиление и фильтрацию, а затем оцифровывается. В процессе преобразования сигнала осуществляется подавление внешнего шума, а также шума микрофона и звуковой карты. Во время работы с диктором система распознавания получает звуковые данные $x_\Delta(i)$, оцифрованные с частотой дискретизации f_d, через интервалы времени $\Delta N/f_d$, где ΔN — длина $x_\Delta(i)$. Эти данные помещаются в буфер x базы данных. Для звуковых данных x на каждом текущем интервале $[l+n_1^*, l+n_2^*]$, где l — текущая позиция, n_1^* и n_2^* — левая и правая границы изменения длины основного тона конкретного диктора, производится интервальная оценка периода основного тона T_{0T} аналогично [5].

Согласно работе [6], выделяются четыре группы звуков речи — шумные шипящие согласные, шумные нешипящие согласные, тональные согласные и гласные. Выделение особенностей этих звуков предусматривает определение интервалов речи в отличие от интервалов шума. Методологически осуществляется выделение левых и правых границ этих интервалов. Выделение левой границы производится согласно классификации звуков на шумные шипящие и тональные. Выделение правой границы осуществляется согласно классификации звуков на шумные нешипящие и звуки речи, относящиеся к тональным и шумным шипящим.

3. Выделение левой границы слова

Структура выделения левой границы схематично представлена на рис. 1.

Звук в текущем интервале сигнала является тональным, если выполняется условие $T_{0T} > 0$, и является шумным, если выполняется условие $T_{0T} = 0$. При оценивании речи на первом этапе, подчиняющемся условию $T_{0T} > 0$, выделяются тональные звуки, которые в последующем классифицируются на высокоамплитудные (речь) и низкоамплитудные (шум). При выделении левой границы методом ДАРФ предусмотрен анализ тональных звуков и, соответственно, условий по характеристикам этих звуков, а затем выделение левой границы для особенностей шумных шипящих звуков.

Вначале рассматриваются первые составляющие сигнала — тональные звуки. Такие звуки могут характеризовать речь (т. е. быть тональными согласными или гласными) или быть тональным низкоамплитудным шумом. Для определения смысловой речи и ее отличия от шума используется соотношение

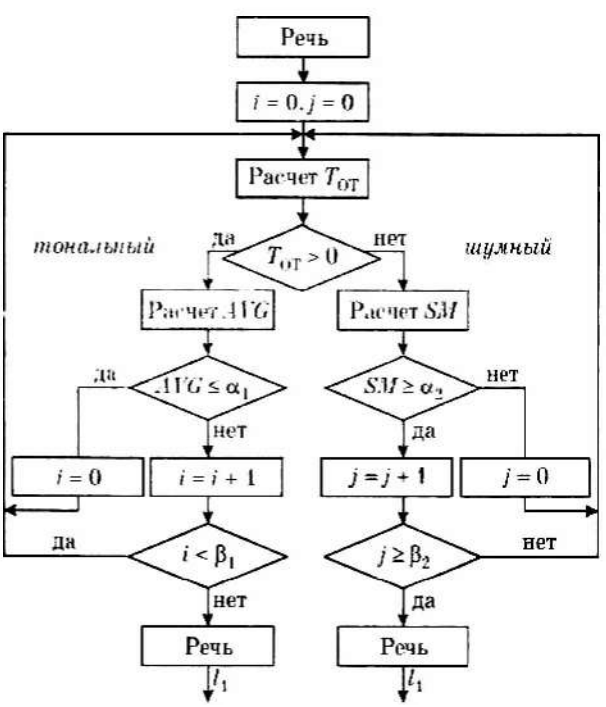

Рис. 1. Структура выделения левой границы

амплитуд, т. е. Вычисляется текущее среднее значение речевого сигнала AVG

$$AVG = \sum_{n=l}^{l+T_{OT}-1} |x(n)| \Big/ T_{OT},\ l \in \overline{0, N}. \qquad (1)$$

Среднее значение шума звуковой карты рассматривается как параметр α_1, определяемый в режиме обучения, и, соответственно, последующий анализ речи диктора осуществляется путем сравнения с этим параметром.

Если выполняется условие $AVG > \alpha_1$, то звук в текущем интервале сигнала является тональным и высокоамплитудным, и в этом случае для него выделяется левая граница слова $l_1 = l$. Затем в последующих процедурах вся временная длина сигнала разбивается на интервалы, удовлетворяющие этому условию, и фиксируется количество таких интервалов i. При этом допускаем, что на этих интервалах сигнал представляет речь, а не шум, если удовлетворяется условие $i \geq \beta_1$, где $\beta_1 = T^1_{\min}/T_{OT}$ — целочисленный параметр речи диктора, характеризующий тональные составляющие речи конкретного диктора; $T^1_{\min}$ — минимальная длина тонального звука, табулированная в [7]. Если выполняется условие $AVG \leq \alpha_1$, то звук в текущем интервале сигнала является низкоамплитудным шумом, тогда осуществляется сдвиг $l = l + T_{OT}$ и оценивание производится на следующем интервале.

Этот подход характеризует одну из составных частей первого этапа распознавания звуков в сигнале (левая часть рис. 1). Поскольку, кроме тональных звуков, в речи могут присутствовать шумные шипящие, удовлетворяющие условию $T_{OT} = 0$, то необходимо оговорить условие разделения особенностей этого вида сигнала и шума (правая часть рис. 1). Для определения смысловой речи, представленной шумными шипящими, и ее отличия от функционального шума может использоваться количество строгих минимумов SM на текущем интервале анализа сигнала $[l, l + n_1^*]$. Количество строгих минимумов шумного шипящего звука для речи любого диктора рассматривается как параметр α_2, определяемый в режиме обучения, и, соответственно, последующий анализ речи диктора осуществляется путем сравнения с этим параметром.

Если выполняется условие $SM \geq \alpha_2$, то звук в текущем интервале сигнала является шумным шипящим, и в этом случае для него выделяется левая граница слова $l_1 = l$. Затем вся временная длина сигнала разбивается на интервалы, удовлетворяющие этому условию, т. е. выделяются интервалы, на которых сигнал в последующем подвергается проверке на слитную речь, и фиксируется количество таких интервалов j. В последующей проверке допускаем, что на этих интервалах сигнал представляет речь, а не шум, если удовлетворяется условие $j \geq \beta_2$, где $\beta_2 = T^2_{\min}/n_1^*$ — целочисленный параметр речи диктора, характеризующий шипящие составляющие речи конкретного диктора; $T^2_{\min}$ — минимальная длина шумного шипящего звука, табулированная в [7]. Если выполняется условие $SM < \alpha_2$, то звук в текущем интервале сигнала относится к шуму, тогда оценивание производится на следующем интервале со сдвигом $l = l + n_1^*$.

Таким образом, подход, изображенный на рис. 1, позволяет последовательно определять, что сигнал содержит интервалы тональных или шипящих звуков, т. е. в сигнале есть хотя бы один тональный или шипящий звук. В этом случае производится фиксация левой границы l_1.

После фиксации левой границы временного интервала оценивания слова необходимо определить его правую границу. Методологически при выделении правой границы на первом этапе осуществляется разделение на нешипящие шумные звуки с количеством нестрогих минимумов $NSM \geq \alpha_3$ и звуки, принадлежащие к остальным трем типам, с количеством нестрогих минимумов $NSM < \alpha_3$, где α_3 — параметр речи диктора. Выделение правой границы производится для звуков, классифицированных как шумные нешипящие (паузы).

4. Выделение правой границы слова

Структура выделения правой границы схематично представлена на рис. 2.

Согласно этой структуре осуществляется нахождение количества нестрогих минимумов NSM на текущем интервале анализа сигнала $[l, l + n_2^*]$. Количество нестрогих минимумов шумного нешипящего звука для речи любого диктора на стадии обучения системы рассматривается как параметр α_3, определяемый при идентификации речи диктора, и, соответственно, текущий анализ звуков речи диктора осуществляется путем сравнения с ним. Это позволяет реализовать уточнение интервальных оценок речи с классификацией ее на шумные нешипящие звуки и звуки речи, относящиеся к тональным и шумным шипящим.

Если выполняется условие $NSM < \alpha_3$, то текущий интервал сигнала является тональным или шипящим звуком, маркируется с учетом рис. 2, и оценивание правой границы производится на следующем интервале со сдвигом $l = l + n_2^*$. Если выполняется условие $NSM \geq \alpha_3$, то текущий интервал сигнала является шумным нешипящим звуком или

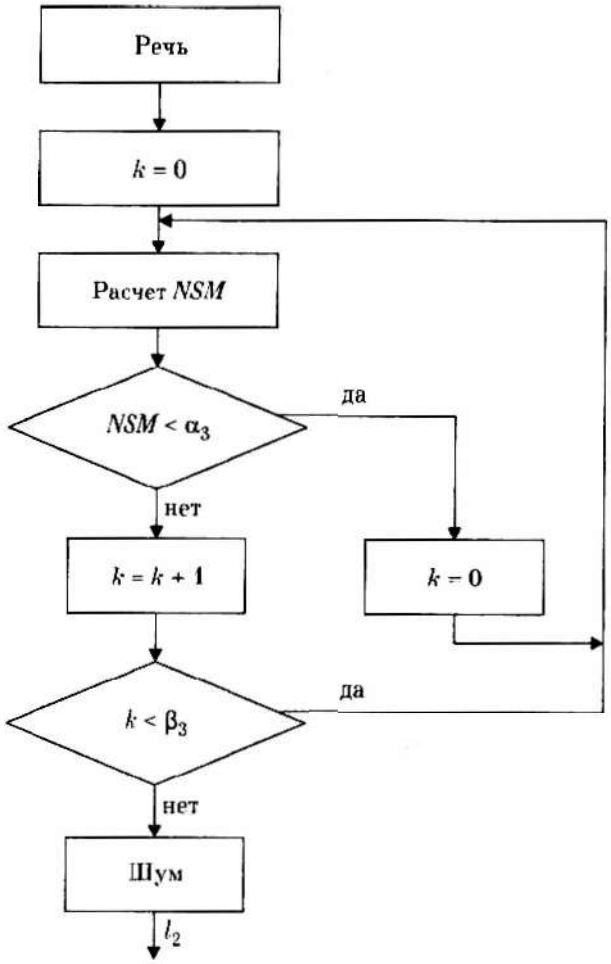

Рис. 2. Структура выделения правой границы

паузой между словами, тогда в этом случае фиксируется правая граница слова $l_2 = l$. Согласно изложенному методу ДАРФ выделения левых и правых границ слова, вся длина сигнала разбивается на фиксированные интервалы, удовлетворяющие этому условию, т. е. выделяются те, на которых сигнал в последующем подвергается проверке на слитную речь. Пауза между словами удовлетворяет условию $k \geq \beta_3$, где $\beta_3 = T^3_{\max} / n^*_2$ — параметр речи диктора, характеризующий шумные нешипящие составляющие речи конкретного диктора; $T^3_{\max}$ — максимальная длина шумного нешипящего звука, табулированная в [7]; k — количество интервалов. Если это условие не выполняется, интервал маркируется как шумный нешипящий.

Таким образом, подход, изображенный на рис. 2, позволяет последовательно определять, что сигнал содержит интервалы шумных нешипящих звуков или пауз между словами. В этом случае фиксируется правая граница l_2.

В результате такого анализа выделяется каждое изолированное слово в сигнале $x(n)$ с границами l_1 и l_2.

5. Оценка параметров речи диктора и алгоритм выделения границ слова

Реализация основных положений, приведенных в методе ДАРФ, в виде алгоритмов обучения позволяет определить численные значения параметров, характеризующих особенности речи конкретного диктора.

Согласно разработанному методу, в режиме обучения по выборке из 1000 произнесенных команд были идентифицированы параметры речи, характеризующие специфику указанного диктора (табл. 1).

Таблица 1

Оценка параметров речи диктора

$P(\beta_1 \geq 4)$	$P(\beta_1 \geq 10)$	$P(\beta_1 \geq 12)$	$P(\alpha_2 \geq 3)$	$P(\alpha_3 \geq 2)$
0,957	0,951	0,954	0,952	0,958

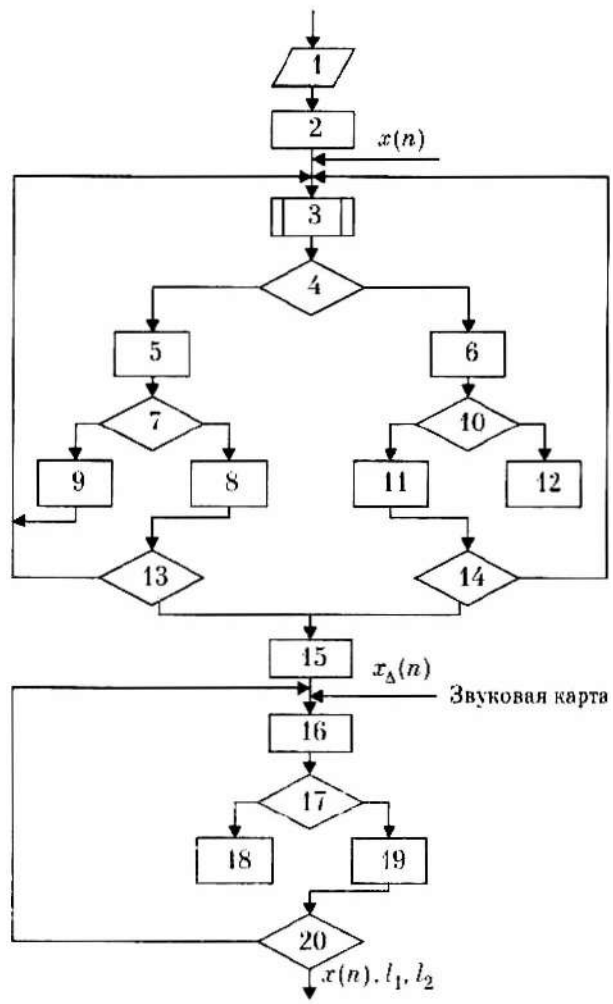

Рис. 3. Блок-схема алгоритма определения границ слова: 1 — задание констант $T^1_{\min}$, $T^2_{\min}$, $T^3_{\max}$, f_d, ΔN; 2 — $i=0$, $j=0$, $l=0$, $l_1=0$; 3 — расчет T_{OT} методом ВОТФ [5]; 4 — $T_{OT}>0$; 5 — расчет AVG (1); 6 — расчет SM; 7 — $AVG \leq \alpha_1$; 8 — $i=i+1$, $l=l+T_{OT}$; 9 — $i=0$, $l=l+T_{OT}$, $l_1=l$; 10 — $SM \geq \alpha_2$; 11 — $j=j+1$, $l=l+n^*_1$; 12 — $j=0$, $l=l+n^*_1$, $l_1=l$; 13 — $i<\beta_1$; 14 — $j<\beta_2$; 15 — $k=0$, $l_2=l$; 16 — расчет NSM; 17 — $NSM<\alpha_3$; 18 — $k=0$, $l=l+n^*_2$, $l_2=l$; 19 — $k=k+1$, $l=l+n^*_2$; 20 — $k<\beta_3$

Таблица 2

Оценка длины периода основного тона диктора

Левая граница, n_1^* (мс)			Правая граница, n_2^* (мс)		
женщины	мужчины-теноры	мужчины-басы	женщины	мужчины-теноры	мужчины-басы
$P(n_1^* \geq 2,5) = 0,952$	$P(n_1^* \geq 5,9) = 0,958$	$P(n_1^* \geq 9) = 0,953$	$P(n_2^* \geq 6,3) = 0,957$	$P(n_2^* \geq 9,1) = 0,959$	$P(n_2^* \geq 15,5) = 0,954$

В режиме обучения по эталонному произнесению звука на основании метода ВОТФ [5] определяются границы изменения длины периода основного тона n_1^* и n_2^* в зависимости от голосового тембра диктора. В табл. 2 приведены их значения, определенные экспериментально для женщин, мужчин-теноров и мужчин-басов. При выборке численностью 600 человек (по 100 человек каждой категории) с произнесением каждым из этих дикторов 10 команд вероятность оценивания границ составила не менее 95 %.

Определение границ слова в процессе распознавания представлено на рис. 3.

На выходе блок-схемы фиксируем границы слов смыслового текста.

6. Выводы

Новизна. В данной работе произведена разработка авторского метода ДАРФ, определяющего границы изолированного слова и учитывающего шумы аппаратной части и параметры речи диктора.

Практическое значение. Основные положения данной работы предназначены для реализации в интеллектуальных системах управления, в которых команды поступают на естественном языке.

Список литературы: 1. *Рабинер Л. Р., Шафер Р. В.* Цифровая обработка речевых сигналов. — М.: Радио и связь, 1981. — 496 с. 2. *Freeman D., Southcott C., Boyd I. A.* Voice activity detector for the Pan-European digital cellular mobile telephone service // IEE Colloquium «Digitized Speech Communication via Mobile Radio». — London. — 1988. — P. 61–65. 3. *Аграновский А. В., Зулкарнеев М. Ю., Леднов Д. А.* Организация иерархической модели распознавания слитной речи // Искусственный интеллект. — 2001. — № 3. — С. 17–22. 4. *Федоров Е. Е., Шелепов В. Ю.* Автоматическое определение начала и конца записи речи // Искусственный интеллект. — 2002. — № 4. — С. 295–298. 5. *Федоров Е. Е.* Выделение длины периода основного тона речевого сигнала // Искусственный интеллект. — 2004. — № 1. — С. 237–242. 6. Современный русский язык: Учеб. для филол. спец. высших учебных заведений / Под ред. В. А. Белошапковой. — М.: Азбуковник, 1997. — 928 с. 7. *Златоустова Л. В.* Фонетические единицы русской речи. — М.: Изд-во МГУ, 1981. — 108 с.

Поступила в редколлегию 07.09.2006