ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1971 г. вып. 44

УДК 621.391:519.2.

МЕТОД АВТОМАТИЧЕСКОГО ВЫДЕЛЕНИЯ УДАРЕНИЯ В ПОТОКЕ РЕЧИ

А.Г. Хайретдинова

При решении таких задач, как автоматическая сегментация речевого потока, нормализация речи по темпу и другие возникает необходимость автоматического выделения ударения в потоке речи.

Целью данной работы является исследование и разработка метода выделения ударных гласных в речевом потоке с помощью ЭВМ.

Многочисленные эксперименты, проведенные нами и другими исследователями речи, позволили подробно изучить фонетические особенности ударных и безударных гласных в акустическом сигнале и указать достаточно большое число их отличительных признаков.

Выбор признаков

При составлении машинного алгоритма из всей совокупности известных признаков были выбраны те, которые, на наш взгляд, являются наиболее устойчивыми и удобными при выделении их на ЭВМ. Рассмотрим эти признаки.

Признак x_1. Ударная гласная в отличие от остальных гласных внутри одного слова характеризуется наибольшей длительностью x_1. Это свойство справедливо как для слов, произнесенных изолированно, так и для слов, произнесенных во фразе. Лингвистическая частость появления этого признака для нормального темпа произношения приближается к 0.9. Однако определение границ гласных с помощью ЭВМ осуществляется с некоторыми ошибками, которые в конечном счете снижают эту цифру до 0.8÷0.85.

Признак x_2. Ударная гласная имеет наибольшую интенсивность x_2 по сравнению с другими гласными фонемами внутри рассматриваемого слова. (здесь под интенсивностью понимается среднее значение амплитуд огибающей речевого сигнала). Это правило соблюдается почти всегда, если ударной является открытая гласная (а, о, э); если же ударной является закрытая гласная (и, ы, у), а среди безударных в слове имеются открытые гласные, то, как правило, такие ударные гласные по интенсивности уступают последним [1]. Поскольку в русской речи частость появления открытых ударных гласных значительно больше частости появления закрытых ударных [2], то признак x_2 является характерным признаком ударности.

Признак x_3. В повествовательном предложении внутри одного слова среднее значение частоты основного тона (ОТ) на участке ударной гласной имеет большую величину, чем на остальных участках. Сказанное можно проиллюстрировать таблицей 1.*) Здесь через x_3 обозначена разность между средними значениями ОТ двух соседних участков. В таблице приведены частости появления ударных и безударных гласных по признаку x_3. Исключением являются слова, стоящие в конце фразы и произносимые обычно с понижением голоса. При этом замечено: если ударение падает на первый слог, то происходит плавный спад средней частоты ОТ к концу слова; если же ударным является другой из слогов (второй, третий), то перед ударной гласной спад средней частоты ОТ происходит скачкообразно (не менее 20 гц).

Таблица 1

гласные	$x_3 < -15$ гц	-15 гц $< x_3 \leq -5$ гц	-5 гц $< x_3 < 5$ гц	5 гц $\leq x_3 <$ 15гц	$x_3 \geq 15$ гц
ударные	0,1	0,11	0,1	0,19	0,52
безударные	0,34	0,26	0,3	0,064	0,04

Признак x_4. В повествовательных предложениях наблюдается повышение частоты ОТ в течение времени звучания ударной гласной, в то время как на безударных гласных чаще всего происходит спад частоты ОТ. Это свойство можно проиллюстрировать таблицей 2. Здесь x_4 – величина, равная разности частот ОТ в

*) Таблицы 1 и 2 составлены на основе анализа гласных, встретившихся в 50 повествовательных фразах, содержащих 175 ударных и 274 безударных гласных (в произнесении одного диктора).

Таблица 2

гласные	$x_4 \leq$ -10гц	-10 гц $< x_4 <$ -2,5гц	-2,5гц $< x_4 <$ -2,5гц	2,5гц $\leq x_4 <$ 10 гц	$x_4 \geq$ 10гц
ударные	0,06	0,07	0,14	0,21	0,49
безударные	0,28	0,28	0,25	0,08	0,09

конце и начале гласной.

В словах, стоящих в конце фразы, на ударной гласной имеет место спад частоты ОТ (например, из 90 конечных слов в 73 случаях x_4 больше 15 гц., в 17 случаях - меньше 15 гц).

В дальнейшем каждая гласная будет характеризоваться этими четырьмя признаками. Следует подчеркнуть, что признаки x_1 и x_2 характеризуют ударные гласные в словах, произнесенных как изолированно, так и во фразах. Признаки x_3 и x_4 относятся лишь к словам, стоящим во фразе.

Описание алгоритма

В основе алгоритма для выделения ударных гласных в речевом сигнале лежат методы распознавания образов. В пространстве выбранных признаков каждая гласная фонема, описываемая совокупностью признаков, может быть интерпретирована как точка; совокупностям всех ударных и безударных гласных соответствуют две односвязные области пространства. Предварительные эксперименты показали, что точки в каждой из областей распределены по закону, близкому к нормальному. Для разделения этих областей была использована гиперплоскость :

$$S = \sum_{i=1}^{n} w_i x_i + w_0 = 0,$$

где x_i - значение i-го признака,

w_i - вес i-го признака ; w_0 - свободный член,

n - число признаков.

Значения w_i и w_0 определялись как коэффициенты уравнения гиперплоскости, перпендикулярной к линии, соединяющей математические ожидания обоих областей и делящей эту линию на части таким образом, что проекции её на координатные оси пропорциональны средним квадратичным отклонениям соответствующих признаков этих образов [4] .

При определении коэффициентов гиперплоскости использовались точки обучающего материала. При принятии решения знак S указывает, по какую сторону от гиперплоскости находится контрольная реализация.

Определение признаков на ЭВМ

При вычислении значений выбранных признаков на ЭВМ требуется в первую очередь определить границы всех гласных фонем . При этом знание вида гласной не обязательно.

С целью определения границ гласных речевой сигнал пропускался через полосовой фильтр с граничными частотами 250 гц и 1000 гц (был использован рекурсивный цифровой фильтр, аппроксимированный по Баттерворту [5]). После фильтра осуществлялось детектирование сигнала и нахождение его огибающей (с интервалом дискретности 15 мсек.), которая представляет собой чередование "горбов" и "нулевых зон" (т.е. участков, где огибающая равна нулю). "Нулевые зоны" соответствуют высокочастотным фонемам (шипящие, взрывные и т.д.), спектральные составляющие которых в указанной полосе частот имеют низкий уровень. "Горбы " в большинстве случаев появляются там, где находятся гласные. На выходе фильтра они представлены первой или двумя первыми формантами, т.е. основной частью своей энергии. Наряду с гласными в этом сигнале присутствуют сонанты (1 форманта). Длительность "горба" на уровне 0.5 от его пиковой амплитуды принималась за длительность гласной.

На практике встречаются случаи, когда несколько соседних фонем оказываются слитыми в один "горб". Сюда относятся в первую очередь дифтонги (например, "ие", "ые"), сочетания некоторых гласных с сонантами (например, слог "умн" в слове "умножить"). Следует сказать, что даже привлечение таких тонких характеристик фонем как спектральные, не всегда помогает членению этих сочетаний.

В данной работе границы между фонемами в упомянутых случаях определялись приближенно, исходя из следующих соображений : если среди "горбов" имеется такой, длительность τ которого существенно отличается от средней длительности τ_{cp} ближайших соседних "горбов" (например, находящихся на участке длиной в одно слово), то в таком "горбе" обычно имеет место слияние нескольких фонем. В большинстве случаев отношение τ / τ_{cp} указывает число фонем, входящих в это сочетание. При этом следует

учесть, что по своей длительности сонанты близки к безударным гласным [3] .

Для рассмотренных нами двух дикторов оказалось: при $\kappa < 2$ имеет место одна фонема, при $2 \leq \kappa < 3$ имеют место две фонемы, при $\kappa \geq 3$ имеют место три фонемы. Как показали дальнейшие эксперименты, такой подход к определению границ гласных является пригодным.

Итак, из речевого сигнала были выделены участки (фонемы), каждый из которых представляет собой либо гласную, либо сонанту.

В дальнейшем для каждого из участков были вычислены признаки x_1, x_2, x_3, x_4.*)

При определении x_3 предварительно подсчитывались текущие значения частоты ОТ. Для этого был использован сдвиговый метод определения ОТ [6] при интервале дискретности 15 мсек.

Нормирование признаков осуществлялось по следующему правилу:

- значения x_1 и x_2 относились к соответствующим максимальным значениям на участке фразы длиной в одно слово (для нормального темпа средняя длина слова около 500 мсек) ;
- для признака x_3 определялись средние значения соответственно по всем его положительным и отрицательным значениям ; значения x_3 в зависимости от его знака относились к модулю соответствующих средних значений.

Признак x_4 нормировался аналогично x_3 .

Для фонем, расположенных в интервале 500 мсек. (длина слова) от конца фразы, признаки x_3 и x_4 были взяты с обратным знаком.

На ЭВМ "БЭСМ-6" по описанному выше алгоритму выделения ударения были обработаны 50 повествовательных фраз, составленных из наиболее употребительных слогов. Среднее число слов во фразах составляло 5-6. Фразы произносились двумя дикторами (мужчиной и женщиной) в нормальном темпе. Общее количество гласных во фразах равнялось 898, из них 350 ударных и 548 безударных.

Обучение проводилось на ударных и безударных гласных, взятых из 20 фраз, после чего распознавались гласные, взятые из 30 новых фраз.

*) Предварительные эксперименты показали, что по указанным признакам сонанты близки к безударным гласным.

Суммарная ошибка подсчитывалась по формуле:

$$\alpha = \rho_1 \alpha_1 + \rho_2 \alpha_2 ,$$

где α_1 - ошибка отнесения ударных гласных к безударным ,
α_2 - ошибка отнесения безударных гласных к ударным ,
ρ_1 и ρ_2 - априорные вероятности появления соответственно ударных и безударных гласных.

Так как в русской речи наиболее употребительными являются двух- и трехсложные слова [7] , то было принято $\rho_1 = 0{,}35$, $\rho_2 = 0{,}65$. Суммарная ошибка получилась равной 0,08.

Результаты эксперимента показали достаточно высокую эффективность выбранной системы признаков при выделении ударной гласной из потока речи.

Предложенный алгоритм принятия решения позволяет достаточно просто и надежно (с вероятностью 0,92) решать рассматриваемую задачу.

Литература

1. Хайретдинова А.Г. Исследование свойств ударения. Отчет. Институт математики СО АН СССР, 1969.
2. Йлкина В.Н., Юдина Л.С. Статистика открытых слогов русской речи. Сб. "Вычислительные системы", Новосибирск, вып.14, 1964.
3. Дифференциальные признаки слогов. Отчет, ЛГУ, 1967.
4. Загоруйко Н.Г. Структура проблемы распознавания слуховых образов и методы её решения. В кн.: Автоматическое распознавание слуховых образов. Новосибирск, "Наука", 1966.
5. Лозовский В.С. Программа синтеза рекурсивных цифровых фильтров (Р.Ц.Ф.). Тр. ИМ СО АН СССР. "Вычислительные системы", Новосибирск, вып. 36, 1969.
6. Соболев В.Н., Баронин С.П. Исследование сдвигового метода выделения основного тона речи. "Электросвязь", № 12 , 1968.
7. Сапожков М.А. Речевой сигнал в кибернетике и связи. М., Связьиздат, 1963. .

Поступила в редакцию
7.I.1971 г.