ПОВЫШЕНИЕ НАДЕЖНОСТИ РАСПОЗНАВАНИЯ СЛОВ СЛИТНОЙ РЕЧИ

Дегтярев Н.П., Левков Е.Я.

Минск, 220068, СССР

Надежность распознавания слов слитной речи, при прочих равных условиях, существенно зависит от того, насколько адекватно отображены закономерные вариации слитной речи в используемых эталонах слов и насколько точно определяются границы наилучшего подобия реализации с эталонами распознаваемых слов. В настоящей работе обсуждаются возможные пути приближения к решению названных задач. Рассматривается также одно из возможных решений задачи распознавания слов в условиях, когда к началу и (или) концу реализации слова может примыкать помеха или "чужое" слово, не входящее в заданный словарь.

ВВЕДЕНИЕ

При слитном произношении слов во фразе акустическое взаимовлияние (коартикуляция) граничных фонем может проявляться с той же силой, что и между фонемами внутри слова. Поэтому отсутствие удовлетворительной коартикуляционной модели стыковки эталонов является одним из источников ошибок при распознавании слитной речи[1]. Одним из радикальных подходов к решению этой проблемы представляется использование дифонов в качестве эталонных элементов слитной речи [2,3].
Тем не менее, использование слов в качестве единиц распознавания сегодня также привлекательно по ряду причин. Прежде всего потому, что разработано много эффективных алгоритмов автоматического формирования эталонов раздельных слов, основные принципы которых могут быть использованы и при решении задач формирования эталонов слитной речи. Но принципиально важный аргумент в пользу такого выбора заключается в том, что при использовании в качестве единиц распознавания более крупных частей речи уменьшается относительное число стыковок эталонных элементов и, как следствие, уменьшается количество ошибок, порождаемых несовершенством той или иной коартикуляционной модели стыковки эталонов.
Один из подходов к решению задачи получения эталонных элементов слитной речи состоит в их извлечении из слитно произнесенных обучающих последовательностей [4,5]. Понятно, что полученные таким способом эталоны при распознавании наиболее эффективны в окружении элементов, заданных обучающей последовательностью.
В настоящей работе обсуждаются возможности автоматического получения эталонов слов, которые учитывают краевые эффекты, возникающие при раздельном и слитном произнесении слов, и которые поэтому допускают произвольный порядок слов в контрольной последовательности без снижения эффективности распознавания.
Следующим важным вопросом является выбор задачи распознавания и связанной с ней стратегией ее решения. Известны постановка и решение обобщенной задачи распознавания слитной речи, составленной из слов заданного словаря[6,7]. Решение этой задачи основывается на двухступенчатой процедуре оптимизации сходства реализации заданной длины с эталонной последовательностью по числу, составу и порядку следования составляющих ее слов. Однако, на наш взгляд, ближе к практике такая ситуация, когда входные фразы содержат не только "свои", но и "чужие" слова, не входящие в заданный словарь, а также помехи различного рода.
В этом случае, а он и рассматривается в нашей работе, становится необходимым текущее распознавание слов заданного словаря по мере того, как они реализуются в непрерывном входном сигнале. При этом стратегия распознавания строится также, как и для распознавания раздельных слов, но допускается, что: 1) они могут примыкать друг к другу при слитном произнесении; 2) их могут разделять "чужие" слова или паузы, в том числе и такие, которые заполнены помехами различного рода. В этих условиях основной стратегии распознавания становится определение границ наилучшего подобия эталонов слов с соответствующими им участками непрерывного входного сигнала. По оценке наилучшего подобия может быть принято решение "свой-чужой" или отобраны ранжированные по значениям меры сходства претенденты на окончательное решение с учетом синтаксического анализа.

Таблица I. Характеристики начальных и конечных сегментов граничных фонем эталонов слов, отображающие краевые эффекты при их раздельном и слитном произнесении

Разбиение граничных фонем на группы		Характеристика и требуемое число вариантов начальных сегментов эталонов	Характеристика и требуемое число вариантов конечных сегментов эталонов
Гласные заднего ряда	У, О, А	1. Переходное сонорное начало 2. Стационарное сонорное начало 3. Переходное мягкое сонорное начало 4. Стационарное мягкое сонорное начало	1. Стационарное сонорное окончание
Сонорные	Э, Ы, И, М, Н, Л, Р М′, Н′, Л′, Р′	1. Переходное сонорное начало 2. Стационарное сонорное начало	1. Стационарное сонорное окончание
Дифтонги	Ю, Я, Е, Ё	1. Стационарное звонкое или фрикативное начало	1. Стационарное сонорное окончание
Звонкие фрикативные	З, З′, Ж, Й	1. Стационарное звонкое или фрикативное начало	1. Переходное звонкое окончание 2. Стационарное фрикативное окончание
Глухие фрикативные	С, Ш, Ц, Ч, Ф С′, Ш′, Х′, Ф′	1. Стационарное фрикативное начало	1. Стационарное фрикативное окончание
Взрывные	Б, Д, Г, П, Т, К, В	1. Переходное звонкое начало	1. Переходное звонкое окончание 2. Стационарное аспиративное окончание
Мягкие взрывнфе	Б′, Д′, Г′, П′, Т′, К′, В′	1. Переходное звонкое начало 2. Стационарное фрикативное начало	1. Переходное звонкое окончание 2. Стационарное фрикативное окончание
Аспиративные	Х	1. Стационарное аспиративное начало	1. Стационарное аспиративное окончание

ПРИНЦИПЫ ФОРМИРОВАНИЯ ЭТАЛОНОВ, УЧИТЫВАЮЩИХ СВОЙСТВА ГРАНИЧНЫХ СЕГМЕНТОВ СЛОВ ПРИ ИХ РАЗДЕЛЬНОМ И СЛИТНОМ ПРОИЗНЕСЕНИИ

Как мы уже отметили выше, при слитном произнесении слов имеют место неизбежные изменения параметров пограничных сегментов их реализаций, отражающих взаимовлияние (коартикуляцию) граничных фонем. Отсюда следует принципиальная необходимость учета названных явлений в эталонных описаниях слов слитной речи. Одним из возможных путей решения этой задачи является использование априорных знаний о закономерностях изменений формантно-фонемных связей под влиянием соседних фонем для автоматизации процесса формирования эталонов слитной речи. Очевидно, что наиболее подходящим для этих целей является формантно-параметрическое описание речевого сигнала[8]. В этом случае, например, оказывается возможным автоматическое формирование нужного числа эталонов слова, отображающих краевые эффекты, имеющие место как при слитном, так и при раздельном произнесении слов, по одной реализации отдельно произнесенного слова[9]. Основные элементы работы такого алгоритма состоят в следующем.
Входной информацией является формантно-параметрическое описание реализации слова и его фонемная транскрипция. Алфавит граничных фонем разбивается на группы (см. табл. I), характеризующиеся одинаковым способом образования или общими акустическими свойствами в пространстве параметров описания, связанными с позицией фонемы - начальная/конечная. Соответственно на первом этапе определяется, к каким группам принадлежат начальная и конечная фонемы в транскрипции слова.
Алфавит фонем, сочетаемых с граничными фонемами, разбивается на группы, характе-

ризующиеся единством акустического влияния на граничные фонемы: 1) паузы, взрывные; 2) мягкие согласные, дифтонги; 3) звонкие; 4) глухие. Поэтому на втором этапе в зависимости от групповой принадлежности граничных фонем и возможного влияния на них сочетаемых фонем, определяются варианты начальных и конечных сегментов эталонов (см. табл. I) и соответствующие правила их формирования.
На последнем этапе в соответствии с заданными правилами из принятой реализации слова формируются эталоны, отличающиеся требуемыми вариантами начальных и конечных сегментов. Требуемые модификации граничных сегментов эталонов получаются путем фильтрации и отбрасывания части отсчетов или части параметров на определенном временном интервале описания реализации или путем изменения значений части параметров реализации слова. На одно слово может быть сформировано один, два или четыре эталона. Сформированные таким образом эталоны учитывают наиболее выраженные краевые эффекты реализаций слов как при раздельном, так и при их слитном произнесении и могут поэтому использоваться в едином алгоритме распознавания раздельно и слитно произносимых слов, основные принципы построения которого излагаются в нижеследующем разделе.

АЛГОРИТМ РАСПОЗНАВАНИЯ СЛОВ В ТЕКУЩЕМ РЕЧЕВОМ СИГНАЛЕ

Исходя из обсуждавшихся во введении предпосылок ставится задача распознавания слов из данного словаря N по мере их реализации в неизвестном текущем сигнале

$$X=\{x(i)\};\ i=\overline{1,I}, \qquad (1)$$

где $x(i)=\overline{p}(i)$ - отсчет вектора p-параметров описания сигнала в i-й момент дискретного $i\Delta t$ времени. Каждый из эталонов $n\in N$ описывается последовательностью $y(j)$ отсчетов тех же p-параметров описания

$$Y(n)=\{y(j)\};\ j=\overline{1,L(n)} \qquad (2)$$

где $L(n)$ - длина (число отсчетов) n-го эталона.
В качестве основы для конструирования мер сходства между эталонами и текущим сигналом в данной работе используется симметричный DP-алгоритм, предложенный в [I0] для расчетов частичного расстояния $g(j,i)$ в точке (j,i) на фазовой плоскости сопоставления образов. Основные принципы решения поставленной задачи приведены в [II] и сводятся к следующему.
На первом этапе формируются N функций $R(n,v(i),i)$, отражающих текущие расстояния начальных сегментов длиной $\ell<L(n)$ каждого из $n\in N$ эталонов к сигналу на его i-ом отсчете

$$R(n,v(i),i)=g(n,\ell,v(i),i), \qquad (3)$$

где $g(\cdot)$ - расстояние, определяемое DP-методом для сигналов с неизвестным началом, когда переход к каждому следующему отсчету реализации производится с нулевой мерой сходства; $v(i)$ - отсчет, с которого начинается оптимальная траектория на фазовой плоскости сопоставления эталона и реализации на ее i-ом отсчете [I2]. Для каждого случая, когда

$$R(n,v(i))=\min_{i} R(n,v(i),i)<PR \qquad (4)$$

где PR - пороговое значение, формируется массив эталонов $\forall\{m\}\exists R(n,v(i))<PR$, а также определяется отсчет $i_н$ вероятностного начала реализации слова

$$i_н=v(i)=\arg\min_{n} R(n,v(i)) \qquad (5)$$

Тем самым предполагается, что на i-ом отсчете обнаружен отрезок сигнала $(i_н,i)$ близкий к начальным сегментам отобранных $\{m\}$ эталонов настолько, что можно предположить начало реализации одного из $n\in\{m\}$ слов.
Прежде чем перейти ко второму этапу алгоритма заметим, что используемая нами на первом этапе мера (3) для оценки близости начальных частей эталонов к текущему сигналу не может быть использована для оценки близости к сигналу эталонов слов в целом, т.к. мера (3) оказывается близкой для таких пар слов, когда одно из них является конечной частью другого. На втором этапе алгоритма, начиная с отсчета $i_н$ реализации для $\forall n\in\{m\}$ эталонов определяются текущие интегральная

$$D(y(n),X)=D(n,i)=g(L(n),i) \qquad (6)$$

и локальная [I3]

$$Q(n,i)=\min_{j} g(j,i);\quad j=\overline{1,L(n)} \qquad (7)$$

меры сходства. По найденным $Q(n,i)$ для тех же эталонов вычисляются интегральные меры удаления (различия) эталонов от реализации

$$G(n,i)=G(n,i-1)+[Q(n,i)-P(i)], \qquad (8)$$

где $P(i)$ - адаптивный к числу $(i-i_н)$ порог. Определяются также претенденты на конечный отсчет $i_к$ реализации слова

$$i_к(n)=\arg\min_{i} D(n,i) \qquad (9)$$

и оценки значений интегральной меры близости

$$D(n)=\min_{i} D(n,i). \qquad (10)$$

Второй этап алгоритма заканчивается, когда

$$\forall n\in\{m\}\exists\ G(n,i)<PG,$$

где PG - пороговое значение, или найдена неречевая пауза.
На третьем, последнем, этапе алгоритма по полученным на предыдущем этапе данным принимаются результирующие решения. При выполнении условия

$$\min_{n} D(n)<PD(n), \qquad (11)$$

где $PD(n)$ - пороговое значение, определяется номер эталона

$$\hat{n}=\arg\min_{n} D(n), \qquad (12)$$

который указывает распознанное слово, и по найденным $\hat{n}$ и $i_к(n)$ - отсчет $i_к$, указывающий конец распознанного слова, после чего осуществляется переход на $(i_к+1)$ отсчете реализации к первому этапу следующего цикла распознавания. В случае не выполнения условия (II), принимается решение об отказе, возврат к $(i_н+1)$ отсчету реализации и переход к первому этапу следующего цикла распознавания.

ЗАКЛЮЧЕНИЕ

На основе описанных алгоритмов построена аппаратурно-программная модель распознавания, характеризующаяся следующими свойствами:
I) в одном режиме работы системы реализуется распознавание как раздельно, так и слитно произносимых слов. Это стало возможным, с одной стороны, благодаря тому, что эталоны слов отображают основные характерные свойства граничных сегментов раздельно и слитно произносимых слов, а, с другой, - благодаря определению границ слов непосредственно в процессе распознавания по текущим оценкам близости элементов эталонов к сигналу;
2) распознаваемые слова могут быть разделены "чужими" словами или помехами, которые отвергаются по пороговым критериям;
3) не накладывается принципиальных ограничений на число и порядок следования слов во фразе.
В связи с последним отметим, что при необходимости данная система распознавания может быть дополнена уровнем синтаксического анализа и тогда алгоритм распознавания может быть модифицирован в алгоритм отбора последовательностей эталонов, ранжированным по оценкам меры сходства.
Проведенные экспериментальные исследования в полной мере подтверждают эффективность предложенных решений в рамках настоящей модели распознавания слов слитной речи. Дальнейшее развитие модели связывается, в первую очередь, с совершенствованием методов получения параметров описания речевого сигнала и уточнением и расширением правил учета коартикуляционных явлений на стыках слов слитной речи.

ЛИТЕРАТУРА

I. S.E.Levinson. Structural methods in automatic speech recognition. Proc. of the IEEE, 1984, vol.73, №11, pp.1625-1650.
2. Слуцкер Г.С., Старостина Э.А. Автоматическая выработка эталонов звуковых диад. Труды Акустического института, вып. XII, 1970, с. 3I-42.
3. H.Ney, D.Mergel, S.Macrus. On the automatic training of phonetic units for word recognition. IEEE Trans. Acoust. Speech and Signal Process., 34, N1, 1986, pp. 209-213.
4. L.R.Rabiner, A.Bergh, J.G.Wilpon. An improved training procedure for connected digit recognition. Bell. Syst. Tch. j., vol. 61, 1982, pp. 981-1001.
5. М.А.Абдуллаев, Ю.Н.Жигулевцев, В.И.Спорыш. Алгоритм формирования эталонов для распознавания слитных фраз. В кн.: Автоматическое распознавание слуховых образов (АРСО-I4). Ч. I, Каунас, I986, с. 84-85.
6. H.Sakoe. Two level DP-matching - A dynamic programming based patter matching algorythm for connected word recognition. IEEE Trans. Acoust. Speech, Signal Process., vol.ASSP-27, 1979, pp. 588-595.
7. Т.К.Винцюк. Обобщенная задача распознавания слитной речи. В кн.: Автоматическое распознавание слуховых образов (АРСО-I2). Киев, 1982, с. 345-348.
8. Н.П.Дегтярев. Двухформантная аппроксимация спектров речи. В кн.: Автоматическое распознавание слуховых образов (АРСО-I4). Ч. I, Каунас, I986, с. I2-I3.
9. Н.П.Дегтярев Н.П. Использование формантно-фонемных связей для формирования эталонов слитной речи. Тезисы докладов Всесоюзного симпозиума "Бионика интеллекта". Харьков, I987.
I0. Г.С.Слуцкер. Нелинейный метод анализа речевых сигналов. Труды НИИР, № 2, I968, с. I8-23.
II. Н.П.Дегтярев. Алгоритм распознавания слов в непрерывном сигнале. Тезисы докладов Всесоюзного симпозиума "Бионика интеллекта". Харьков, I987.
I2. Б.М.Лобанов, Г.С.Слуцкер, А.П.Тизик. Автоматическое распознавание звукосочетаний в текущем речевом сигнале. Труды НИИР, № 4, I969, с. 67-75.
I3. Н.П.Дегтярев, Б.М.Лобанов, Г.С.Слуцкер. О двух вариантах построения устройств распознавания речевых команд. - В кн.: Автоматическое распознавание слуховых образов (АРСО-I0). Тбилиси, I978, с. I99-200.