**2.2.2. Информационный критерий оценки фонетической** неопреде­ленности. При распознавании устной речи необходимо стремиться к тому, чтобы все фонемы классифицировались правильно, поэтому нас интересует распознавание полной последовательности фонетических единиц, составляющих высказывание. При этом основным источником неопределенности при распознавании речи является сам акустичес­кий сигнал. Еще большую неопределенность представляет параметри­ческое описание речевой волны. Рассмотрим неопределенности аку­стического сигнала и приведем меру оценки фонетической неопре­деленности. Используя эти мерь, можно оценить лексическую и фра­зеологическую неопределенности. Слитная речь расчленяется на пос­ледовательность сегментов по признакам способа образования зву­ков. *К* этим признакам добавляются признаки места образования, ко­торые изменяются непрерывно как внутри сегментов, так и через **их г**раницы [91,97]. С некоторыми дискретными единицами-звуками ре­чи - фонемами или квазифонемами сегменты связаны таким образом, что смысловые единицы речи (слова) представляются цепочкой фонем.

Большинство систем автоматического распознавания речи [79] преобразует речевой сигнал в такую фонемную цепочку, которая за­тем сравнивается с ожидаемыми в слове звуками. Процесс преобразования речевого сигнала в последовательность фонем включает нахождение признаков, сегментацию и маркировку сегментов.

Опишем модель фонетической неопределенности, позволяющую оце­нивать результаты неправильного распознавания фонем. Далее будем использовать матрицу ошибок распознавания фонем и фонетическую структуру слов словаря при оценке лексической неопределенности.

Лексическая неопределенность будет иметь место тогда, когда слова неверно классифицируются из-за близости **их** фонетической структуры, т.е. последовательности параметров, определяющих эту структуру, на конкурирующих словах. Например, в словах "слезать" и "срезать" первичные параметры звуков, входящих в эти слова, сходны. Когда оба эти слова входят в один и тот же словарь, их точ­ная классификация затруднена, поэтому их можно считать лекси­чески неопределенными. В реальных системах, если позволяет зада­ча, следует подбирать слова, чтобы такой ситуации не возникло. При­ведем критерии сложности словаря для того, чтобы можно было оце­нить степень различимости словарей [63].

Рассмотрим распознавание речи как процесс передачи речевой информации через канал с шумом и оценим информацию, теряющуюся в канале. Потерянная информация является мерой неопределенности или сложности распознавания фонем. В идеальном канале число вход­ных идеальных, полученных после сегментации высказывания экспер­тами-фонетистами, и выходных фонетических единиц должно быть оди­наковым, а последовательность фонем **на** выходе должна соответ­ствовать входной последовательности. Если же это условие не соб­людается, в канале теряется информация, и в зависимости от величины потерь можно говорить о большей **или** меньшей неопределенно­сти классификации фонем. При практической оценке фонетической не­определенности в данной работе использовались система призна­ков [73] и алгоритм сегментации речи **на** семь типов сегментов:

*V -* гласный, Т - переходный, *М -* сонорный, *L -* низкоча­стотный, *Н -* высокочастотный, R - шумный, *П -* пауза. Затем алгоритм маркировки ставил в соответствие каждому сегменту не­который фонетический символ, используя априорно полученные гисто­граммы параметров. От надежности маркировки сегментов во многом зависит точность работы CPP.

Так как СРР рассматривается здесь как канал передачи инфор­мации, предположим, что имеются *R* возможных входных символов алфавита *А* и s возможных выходов алфавита В . Таким об­разом, СРР описывается канальной матрицей.

Канал передачи информации, используемой для описания **сис­темы** распознавания речи, представленной цепочкой фонем, преобразу­ет не зашумленную последовательность звуков в выходную последова­тельность "машинных " фонем, содержащую ошибки пропуска, вставки слияния и замены звуков.

Пусть элемента входного фонетического алфавита {Ai} появля­ются на входе с некоторой априорной вероятностью p*(A1 ),р(A2* ),.,p(Ar), а элементы алфавита {Bj} на выходе - с вероятностью *P(B1,)*, *p(B2),*..., *р(Bs).*  Как отмечено ранее, работу канала пере­дачи входного алфавита {Ai} характеризует канальная матрица, поэтому

P{Bj}=∑ri=1P(Ai)\*P(Bj/Ai)

|  |  |  |  |
| --- | --- | --- | --- |
| Символ | А | О | И |
| А | 0,89 | 0,1 | 0,01 |
| O | 0,15 | 0,75 | 0,1 |
| И | 0,01 | 0,1 | 0,89 |

2.2 Пример матрицы условных вероятностей распознавания изолированных звуков*(2.2)*

Информация I(Ai,Bj), получаемая от канала, когда на его вход поступила фонема Ai *,* а на выходе распознавалась как Bj, , определяется [91]

I(Ai,Bj)=LOG(P(Ai/Bj)/P(Ai)) (2.3)

Средняя информация, получаемая на выходе канала с потерями при передаче (распознавании) входного алфавита фонем A(Ai)*,* который распознается как алфавит B=(Bj) , будет

I(A, B)=∑A, BP(Ai,Bj)\*I(Ai,Bj)= ∑A, BP(Ai,Bj)\*LOG2 (P(Ai/Bj)/P(Ai)=

=-∑A, BP(Ai,Bj)\*LOG2 P(Ai)+ ∑A, BP(Ai,Bj)\*LOG2 (P(Ai/Bj);

I(A, B)=H(A)+∑A, BP(Ai,Bj)\*LOG2 (P(Ai/Bj); (2.4 )

Отметим, что Н(A)- энтропия, характеризующая степень неопределенности входного алфавита *А*=(Ai) *.* Из (2.4) получаем, что

H(A)- I(A, B)=-∑A, BP(Ai,Bj)\*LOG2 P(Ai/Bj)=

=-∑A, BP(Ai,Bj)\*P(Bj)LOG2 P(Ai/Bj)=-∑ BP(Bj)∑AP(Ai/Bj)LOG2 P(Ai/Bj)=H(A/B)

Н(А/B)- **апостериорная энтропия входного** алфавита фонем, которая

характеризует меру информации, теряемой в системе распознавания при передаче входного алфавита (Ai) . Апостериорная энтропия и является мерой, оценивающей сложность входного словаря для авто­матического распознавания при фиксированном параметрическом описании.

При наличии значений энтропии входного алфавита фонем можно вычислить размер (объем), равный 2H(A)*,* а значения 2 H|(A/B) ха­рактеризуют среднее количество возможных альтернативных (конку­рентных) элементов алфавита (Ai) на входе СРР после того, как на выходе получили множество (Bj) , т.е. меру сложно­сти распознавания входного алфавита фонем. Назовем эту меру эквивалентным размером алфавита фонем. Значение 2 H|(A/B)  можно назвать энтропийным критерием оценки фонетической неопределенно­сти, который является обобщенной характеристикой сложности рас­познавания алфавита фонем (Ai) данной системы распознавания. Если СРР работает без ошибок, условная энтропия Н(А/В)=О и эквивалентный размер алфавита фонем 2 H|(A/B) =1. Естественно, что если Н(А/В)=0, то 2 H|(A/B) =1, а в случае, когда СРР не рас­познает Н(А/В)=Н(А), то эквивалентный размер алфавита фонем равен 2 H|(A) .

Эквивалентный размер алфавита фонем дает возможность коли­чественно оценить среднее число возможных конкурентных фонем (имеющие близкие параметрические описания), и для его определе­ния необходимо знать апостериорные вероятности P(Ai/Bj) вхо­дного алфавита.

Для решения конкретных проблем автоматического распознава­ния ограниченных наборов слов все многообразие фонем можно свести к двум-трем рабочим фонетическим единицам (например, к классам длительных шумных, звонких и смычных звуков), которые при использовании простой системы признаков и несложных алго­ритмов распознавания дают нулевую апостериорную энтропию. Однако при решении задачи распознавания относительно сложных словарей и/иди требование надежной фонетической верификаций произнесенно­го слова такого количества рабочих фонем сказывается явно недо­статочно. Работать же с полным набором фонем "ложно из-за оши­бок их автоматического распознавания. Поэтому и приходится идти **на** компромиссные решения - искать какой-то оптимум при фонетичес­ком описании рабочих словоформ. **Эти** проблемы будут частично рассмотрены в п. 2.2.3.

Условные вероятности распознавания фонем P(Ai/Bj), опреде­ляющие эквивалентный размер фонетического алфавита, можно определить несколькими методами.

83

Статистический метод позволяет получать вероятности распоз­навания фонем, используя реальную СРР. Это осуществляется путем сравнения результата распознавания системы с точной ручной сегментацией и маркировкой речевого сигнала (иди его параметрического представления), поступающего на вход системы распознава­ния. В результате получается классическая матрица правильной и ошибочной классификации входного алфавита фонем.

Акустико-параметрический метод, когда матрица ошибок клас­сификации фонем получается путем прямого сравнения их параметрического описания. При этом эталон фонемы выбирается из мно­жества реализации данной фонемы. Расстояние между фонемами исполь­зуется для оценки условных вероятностей ошибочной классификации фонем. Точность этого метода зависит от выбранного эталона и объе­ма исследовательского материала.

Кроме этих методов, оценку вероятности ошибочной классификации фонем можно произвести **на** основе моделирования речеобразующего тракта человека [7].

**2.2.3. Оценка сложности распознавания слов по их фонетической структуре.** Рассмотрим неадаптивную систему распознавания слов как канал передачи информации. Слова входного словаря V=(V1,V2,..,Vr,..,VR) можно представить последовательностью фонетических символов Vr=(Ai1,Ai2,..,Ain) , а слова выходного словаря канала W=(W1,W2,..,Ws,..,WS) –цепочками квазифонетических эталонов Ws=(Bj1,Bj2,…,Bjr) ,где Ai∈A , Bj∈B – соответственно входной и выходной алфавит фонем канала ; r= 1,R ; s= 1,S ; n = n (r ) ; l= l(s)*.* Тогда оценку сложности распознавания слов, производимого сравнением входной реализации с цепочками квааифонетических эталонов, можно осуществить на основании анализа матрицы ошибок, подученной при представлении эталонов слов Ws∈W поверхностными формами Wsk ∈ Ws , K=1, Ks каждого выходного слова. Фактически сложность распознавания входного словаря *V* определяется наличием сходных эталонных поверхностных форм Wsk выходного словаря W и частотой встречаемости этих поверхностных форм P(Wsk). Основная проблема при построении матрицы ошибок для каждого словаря заключается в формировании эталонов поверхностных форм Wsk ∈ Ws , для реализация каждого слова и получения квазифонетического графа f(Ws), учитывающего все поверхностные формы в вероятностями **их** появления. Все множество квазифонетических по­верхностных форм слова Ws*,* записать в виде эталонного графа трудно, так **как** при аппаратурно-программном методе распознавания появляются не только поверхностные формы слова, обусловленные особенностями произношения, но и формы, включающие случайные сегменты, маркированные квазифонетическими метками, появление которых связано с не идеальностью автоматической фонетической сегмен­тации и маркировки нашим аппаратурно-программным методом, вызванной, например, изменением интенсивности речевого сигнала.

В дальнейшем будем рассматривать влияние двух обстоятельств на формирование эталонных поверхностных форм слов рабочего словаря, учитывая, что поверхностные формы, связанные о особенностями произношения и матрицей ошибок квазифонемной классификации, можно построить вручную (или автоматически, используя таблицу акустико-фонологических правил, хранящуюся в памяти, и прилагаемых к базовой квазифонетической цепочке), а поверх­ностные формы Wsk *.* , обусловленные особенностями аппаратуры вы­деления информативных признаков, можно получить, анализируя статистику реализации квазифонетических цепочек слов рабочего сло­варя, полученных с помощью ЭВМ. Получение этой статистики не всегда обязательно, особенно если рассматриваются слова, контраст**ные** по своим акустическим свойствам. Предварительную оценку сложности распознавания слов можно сделать аналогично оценке сложности фонетического алфавита - по фонетической структуре слов, вычисляя апостериорную словесную неопределенность и не исследуя статистики реализации.

Все эталоны слов Wsk рабочих словарей должны быть представлены последовательностью маркированных фонетическими метками отрезков, где квазифонемы должны делиться на опорные, обязательные для данного слова (определяющие базовую форму **и,** как Правило, присутствующие во всех поверхностях), и "вспомогатель­ные", трудно классифицируемые. Трудно классифицируемые сегменты должны быть расчленены (хотя бы грубо) на несколько квазифонети­ческих элементов, если длина этих сегментов выше пороговой (это делает на первом этапе человек на основании знаний фонетической структуры возможных форм каждого слова). Опорными сегментами слова следует считать маркированные отрезки которые при их маркировке квазифонетическими метками допускают суммарную ошибку ниже эвристически определенного порога.

При автоматическом распознавании выбор эталонов (из словаря эталонов) должен быть в первую очередь обусловлен наличием в поступившей на вход реализации опорных, обязательных маркированных сегментов о. с учетом того, что за счет не идеальности сегментации общее число сегментов входной реализации может не совпадать с возможным числом сегментов эталонного графа, за счет не опорных сегментов, образующихся **или** выпадающих случайно.

Ошибки классификации дают появление "путающихся" поверхностных форм (представленных последовательностью казифонемы для раз­личных слов словаря. Будем считать, что матрица ошибок при распознавании слов априори формируется таким образом, что (при сходстве поверхностных форм различных слов словаря) более часто встре­чающиеся поверхностные формы слов одного класса считаются относящимися к словам только этого класса, а редко встречающиеся сходные поверхностные формы для других слов словаря дают ошибки распознавания. Впрочем, используя синонимию или семантико-синтаксические ограничения при распоз­навании пословно произносимых фраз. Всегда следует добиваться того, чтобы подобные случаи не происходили (трудности представ­ляют слова, входящие в одну семантико-синтаксическую группу, которые нельзя заменить синонимами, например, названия цифр).

Следует отметить, что принятые решения о принадлежности поступившей **на** вход реализации к тому или иному классу следует делать но эталонам с одинаковым числом опорных сегментов и с учетом верификации слова, всякий раз используя эвристически выбранные пороги достоверности, в общем случае разные для различных слов. Так, для принятия окончательного решения о принадлеж­ности входной реализации Vx к классу Ws необходимо выбрать два наиболее вероятных кандидата Ws1 И Ws2 , которым соответствуют вероятности P(Vx/Ws1) и P(Vx/Ws2) ,и проверить, удовлет­воряются ли условия:

P(Vx/Ws1)> Δ s 1;

P(Vx/Ws1)-P(Vx/Ws1))> Δ s1s2

где Δs1 *-* пороговое значение вероятности того, что входная реа­лизация соответствует слову Ws1 , Δ s1s2 *-* пороговые значения разности условных вероятностей принадлежности входной реализации Vx классам Ws1 и Ws2 *,* при которых принимается решение о классификации Vx.

Пороговые значения Δs1, Δ s1s2 выбираются экспериментально по заданной системе используемых фонетических признаков, а также требуемых точности распознавания и вероятности отказов от рас­познавания. В случае, если подбором порогов заданные требования к системе распознавания не удается выполнить, следует провести более детальный анализ не опорных сегментов, иди попытаться улуч­шить систему признаков. В ряде случаев для удовлетворения заданных в системе требований следует использовать синонимию. Рассмотрим далее более конкретно, как оценить лексическую неопределенность словаря V языка речевого общения неадаптивной системы автоматического распознавания. Аналогично тому, как оценивалась неопределенность алфавита фонем, можно определить сложность распознавания входного словаря *V ,* состоящего из *R* слов, и вычислить эквивалентный размер входного словаря. При этом необходимо получить вероятности P(Vr/Ws) близости областей признакового описания слов Vr∈ V , Ws∈ W , r= 1R , s= 1S , которые представляются в виде последовательности фонетических единиц (фоне­тической транскрипции слов). Далее оценим вероятности P(Vr/Ws).

Как уже отмечалось, на основе лингвистических знаний, эталоны слов Ws∈ W представляются в вида фонетических ( вер­нее, квазифонетических) цепочек, совокупность которых описывается графом с конечным числом состояния, а каждая фонема - признака­ми способа и места образования. Слову Ws соответствует одна **или** несколько траекторий (цепочек поверхностных форм) на графе (количество траекторий зависит от метода произношения и харак­теристики диктора). Направленный граф f (Ws) представляет все фонемы эталона слова Ws∈ W *,* который имеет Wsk*,* поверхностных форм, K = 1,2,3..,Ks; каждая поверхностная форма Wsk∈Ws содержит L=L(s,k) опорных квазифонемы ,

Необходимо отметить, что количество опорных сегментов в поверхностных формах слов выходного словаря различно, т.е. предел изменения индекса L зависит как от номера слова, так и от его поверхностной формы L=L(s,k).

Для того, чтобы осуществить оценку неверной классификации слов словаря на стадии лексического распознавания по фонетичес­кой структуре этих слов, выполним операцию разбиения всех по­верхностных форм эталонов слов на *М* фонетических групп с одина­ковым количеством опорных сегментов L=L(s) *.* При этом слова, поверхностные формы которых принадлежат разным группам, не будут путаться между собой, поскольку их легко классифицировать по числу "опорных" фонем, составляющих слова.

Вообще говоря, можно представить себе фонетические группы эталонных поверхностных форм, отличающиеся не только числом опор­ных фонем, но **и их** характером, а также порядком следования. Если учесть все три фактора, позволяющие разбить эталоны на существен­но большее число фонетических групп, то дальнейшие рассуждения можно отнести к каждой из этих групп. Для простоты, однако, бу­дем считать, что мы имеем *М* фонетических групп, в каждой из которых одинаковое число опорных сегментов. В практических за­дачах при разбиении на группы следует учитывать все эти факторы, однако необходимо строго ограничивать число различных опорных сегментов, выбирая лишь те, которые не путаются между собой и ха­рактеризуются групповыми признаками места образования - ударные гласные, смычные, фрикативные [81,80] .

Итак, допустим, что существуетM фонетических групп слов W1,W2,W3,…,Wn,…,Wm *,* в каждой из которых одинаковое число опорных квазифонемы. Общее число эталонов W= Un=1m Wm *,* а ко­личество фонем, составляющих: слова (длина фонетической цепочки) каждой группы, обозначим через Lm; m=1, M.

Представляя таким образом слова словаря на входе СРР и ис­пользуя матрицы ошибочной классификации фонем, составляющих слова

P(a/b)=[Pij], (2.5)

можно оценить вероятности Pm(Vr/Ws) спутывания поверхностных форм слов внутри каждой группы слов следующим образом:

Pm (Vr/Ws)=П; (2.7)

*Где* T=1,2,..,Tmдлина фонетической цепочки группы слов Wm , Art ∈ Vr ,

Bst ∈ Ws*..*В общем случае одно и то же слово Ws может иметь Ks*,* поверхностных форм, имеющих разное число фонетических элементов и попадающих в разные группы слов Wm *.* Поэтому общую услов­ную вероятность "спутывания" слов словаря определим

P(Vr/Ws)= (2.8)

Для определения потери информации в СРР, которая рассматрива­ется как канал передачи информации, в случае распознавания слов используем выражение

I(V/W)=-, (2.9)

То тогда 2I=(V/W) определяет эквивалентный размер словаря - число альтернативных слов на входе системы распознавания, а 2I=(V) *-* фактический объем входного словаря, где

I(V)=- , (2.10)

Эти выражения, аналогичные формулам (2.4), (2.5), оцениваю­щим фонетическую неопределенность, являются критерием оценки лексической неопределенности. Они определяют сложность распознава­ния словаря и позволяют судить о качестве СРР. При автоматичес­кой маркировке, наряду с ошибками неверной классификации фонем, существуют, как уже отмечалось, ошибки неверной сегментации, при­водящие к слиянию отрезков, соответствующих смежным фонемам, в один сегмент или расчленению отрезка, соответствующего одной фонеме, на несколько смежных фонем разных классов. При выборе альтернативных слов словаря надо следить за тем, чтобы неприят­ности такого рода не вызывали подобия последовательностей фоне­тических единиц, соответствующих разным словам. Для этого необ­ходимо использовать матрицы, отражающие возможные варианты сег­ментации слов словаря и частоты встречаемости тех или иных вари­антов сегментации, соответствующих различит поверхностным формам слов. Так как информация о словах, содержащихся в фонемах, избы­точна, то часто при оценке различимости слов словаря вполне достаточно использовать опорные фонемы, допускающие минимум оши­бок расчленения и слияния. Поэтому в формуле (2.7) для прибли­женной оценки спутывания слов необходимо в первую очередь ис­пользовать вероятности ошибочного распознавания таких опорных фонем, которые в данном слове не дают ошибок слияния и расчлене­ния.