Классификация звуков русской речи с помощью бинарных решающих деревьев

В.Б. Кузнецов,
кандидат филологических наук

В.Я. Чучупал,
кандидат физико-математических наук

Рассматривается вопрос контекстно-зависимой классификации звуков русской речи с помощью построения бинарных решающих деревьев. В качестве речевого материала использовалась обучающая выборка базы данных TeCoRus, предназначенная для приложений, использующих телефонный канал связи. В работе приводится описание результатов экспериментальной классификации русских гласных и согласных в зависимости от контекста.

В последние годы традиционная фонетика претерпевает значительные изменения под напором идей, методов, инструментария и достигнутых результатов в области речевых технологий. В частности, это относится и к такой проблеме, как спектральная классификация звуков речи. Ведётся острая полемика между теми исследователями, которые считают, что фонетическое качество гласного определяется его формантной структурой на стационарном сегменте, где достигается или по крайней мере осуществляется максимальное приближение к акустической цели данного звука, и теми учёными, для которых характер гласного задаётся динамикой его спектральных параметров.

Проблема инвентаризации звуков речи приобрела на сегодняшний день исключительное значение прежде всего в речевой технологии [4]. Так, решая задачу синтеза речи путём «склеивания» сегментов, соизмеримых с отдельным звуком, нужно определить, сколько реализаций звука [а] необходимо взять, чтобы обеспечить приемлемое качество звучания. Если не учитывать фонотактические ограничения, то только для одного ударного [а] необходимо рассмотреть более 1000 различных контекстов. При построении систем распознавании речи также важно располагать оптимальным инвентарём звуков, для которых будут строиться акустические модели. Оптимальность здесь

подразумевает, в частности, компромисс между количеством акустических моделей, точностью представления речевого материала и возможностью оценки параметров моделей на доступных обучающих выборках.

Решение этого вопроса на практике приводит к необходимости ввести понятие обобщённого (типичного) аллофона. В работах создателей системы синтеза русской речи (филологический факультет МГУ) обобщённый аллофон понимается как «акустически и перцептивно различаемая контекстная реализация фонемы» [1]. Формирование множества обобщённых аллофонов осуществляется экспертным путём на основе акустико-фонетических знаний. В одном из последних вариантов синтеза число обобщённых аллофонов гласных достигало 1100, а при разработке тем же коллективом исследователей речевой базы данных предполагается, что в ней будет содержаться как минимум 1800 аллофонов.

Очевидно, что экспертный подход к определению инвентаря обобщённых аллофонов наряду с достоинствами обладает рядом существенных ограничений. Во-первых, принятие экспертом решений во многих случаях не удаётся формализовать (особенно это относится к перцептивным оценкам), во-вторых, объём акустико-фонетических знаний — величина непостоянная, в третьих, эксперт не в состоянии провести на высоком уровне сопоставительный слуховой и акустический анализ большого числа аллофонов.

Альтернативой экспертному подходу могут служить методы вероятностного моделирования, играющие сегодня ведущую роль в автоматическом распознавании и синтезе речи. В настоящей работе для моделирования спектральной динамики гласных используется скрытая марковская модель [6], которая позволяет представить звук в виде последовательных состояний, соотносимых с членением звука на сегменты (субаллофоны). В нашем случае гласный разделяется на три отрезка одинаковой длины (начальный и конечный формантные переходы плюс вокалическое ядро). В качестве алгоритма классификации состояний СММ применяется метод кластеризации сверху-вниз бинарного решающего дерева [10]. Суть этого метода применительно к нашей задаче заключается в следующем. На первом шаге построения дерева акустические наблюдения (вектора из параметров) всех звуков объединены в одном корневом узле, для которого строится общая акустическая модель. Затем из заранее сформированного списка бинарных вопросов (то есть вопросов, которые допускают только два ответа — да и нет), которые могут относиться как к самому гласному, так и окружающим его звукам, выбирается вопрос, дающий наилучшее в некотором смысле разбиение множества наблюдений корневого узла на два дочерних. На следующем шаге среди всех возможных пар «узел-вопрос» ищется та, что обеспечивает последующее оптимальное ветвление дерева. При выполнении определённых условий построение дерева считается завершённым, а листья этого дерева (терминальные узлы) являются искомыми классами субаллофонов, из которых конструируются исходные аллофоны.

Речевой материал и модель звука

В качестве речевого материала использовалась обучающая выборка базы данных TeCoRus [9], предназначенная для приложений, использующих телефонный канал связи. Обучающая выборка представляет собой шестичасовую запись чтения шестью дикторами (из них 3 женщины) фонетически представительного множества 510 отдельных предложений, отсегментированных вручную. В экспериментах по классификации гласных использовались записи только дикторов-мужчин.

В качестве параметрического описания речевого сигнала выбрана наиболее распространённая сейчас система признаков — мел-кепстральные коэффициенты и их первые производные. Эти параметры оценивались на 25 мс окне анализа. Вычислялось 16 коэффициентов и их первых производных. Таким образом, элементарное наблюдение представляло собой вектор из 32 параметров. Последовательность таких векторов использовалась для построения двух кодовых книг, отдельно для мел-кепстральных коэффициентов и их производных. В качестве кодовой книги применялась нейронная сеть — трёхмерная карта признаков Кохонена из 1000 элементов. Фонетическая модель звука представляла собой дискретную скрытую марковскую модель, обычно из трёх состояний.

Бинарные решающие деревья

Во многих случаях дискретной классификации или распознавания образов типичной задачей является оценка значения некоторого (конечного и дискретного) параметра по имеющимся наблюдениям. Решающее дерево — это граф, который задаёт соответствие между наблюдениями и искомыми значениями параметров. Листья решающего дерева соответствуют возможным значениям параметров, а ветви — некоторым комбинациям признаков, в соответствии с которыми наблюдения группируются в различные классы. Таким образом, решающее дерево можно рассматривать как представление алгоритма классификации наблюдений. Если каждый узел решающего дерева имеет ровно два потомка, тогда дерево называется *бинарным решающим деревом*.

Построение бинарного решающего дерева для классификации звуков речи состоит из следующих этапов [5]:

- формирование набора потенциальных корневых узлов;
- создание множества вопросов к звукам, их левым и правым контекстам, идентифицирующих их принадлежность к классу звуков или к конкретному фону;
- определение критериев ветвления узлов, включая оценку приращения логарифма коэффициента правдоподобия и минимальное количество наблюдений (заселённость) в терминальном узле (классе).

Построение общего дерева для звуков (например, всех гласных или всех согласных) начинается с единственного корневого узла, в котором без учёта контекста и состояний СММ объединены все наблюдения из выборки. Для узла q строится вероятностная модель распределения параметров наблюдений $q(x)$ и оценивается её качество. В данном случае выборка наблюдений для узла разбивалась на два равномощных подмножества, на одном из которых вычислялись эмпирические частоты распределения параметров, другое подмножество (контрольная выборка) служило для оценки качества модели, которое определялось как логарифм правдоподобия контрольной выборки относительно модели:

$$L(q) = \sum_{x_1, x_2, \ldots, x_N} \log q(x) = \sum_{x \in \Omega_q} \overline{q}(x) \log q(x),$$

где $x_1, x_2, \ldots, x_N$ — список наблюдений, составляющих контрольную выборку, а $\overline{q}(x)$ — эмпирическая вероятность появления наблюдения x в контрольной выборке Ω_q.

К корневому (родительскому) узлу ищется оптимальный вопрос из конечного множества вопросов, обеспечивающий такое расщепление родительского узла на два

дочерних, которое даёт максимальное приращение оценки качества моделирования. Расщепление узла означает, что принадлежащие этому узлу векторы параметров разделяются на два подмножества в соответствии с тем, удовлетворяют они или нет поставленному вопросу.

Пусть в результате применения некоторого вопроса узел *q* стал родителем для узлов *r* и *r'* . Мера пригодности вопроса для узла *q* определялась как величина приращения качества моделирования, то есть:

$$\Delta L(q) = L(r) + L(r') - L(q) \ ,$$

где $L(q)$ — качество модели родительского узла, $L(r)$ и $L(r')$ — качество модели первого и второго дочерних узлов.

На каждом последующем шаге для текущих терминальных узлов ищется такая пара «узел-вопрос», которая обеспечивает максимальное значение .

Если найденная величина $\Delta L(q)$ превышает заранее заданное пороговое значение и число обучающих фреймов в потенциальных узлах соответствует критерию минимальной заселённости, родительский узел расщепляется на два дочерних.

Когда ни один из терминальных узлов не может быть расщеплён (например, выигрыш от расщепления данных в узле становится меньше пороговой величины или заселённость в терминальном узле становится ниже допустимого минимального значения) или число терминальных узлов достигает заранее установленного порогового значения, процедура ветвления останавливается и дерево считается построенным.

Одно из основных преимуществ, связанных с применением процедуры кластеризации «сверху-вниз» на базе решающих деревьев, состоит в том, что при классификации аллофонов, не представленных в обучающей выборке, мы можем справиться с этой ситуацией, привлекая экспертные знания о классах фонетически близких аллофонов, для которых на этапе обучения уже были получены соответствующие статистические модели.

Таким образом, задача фонетиста состоит в том, чтобы определить классы звуков, которые оказывают на своих соседей в речи сходное коартикуляционное воздействие, и выразить эти экспертные знания в форме множества бинарных вопросов (требующих ответа «да» или «нет»), которые затем будут использованы для расщепления узлов дерева.

Классификация гласных звуков

Для системы русских гласных, известных своей высокой контекстуальной вариативностью и степенью редукции, было предложено в качестве потенциальных корневых узлов 53 иерархически организованных класса. На вершине классификации находится класс «Все гласные»; классы низших уровней могут состоять как из единичных элементов (например, ударный [о]), так и из группы схожих гласных. Большое количество классов объясняется, в частности, тем, что их исходное число было практически удвоено, чтобы учесть назализацию гласных в соседстве с носовыми согласными. Результаты сегментации базы данных показали, что назализация, не являясь в русской речи смыслоразличительным признаком, регулярно проявляется в речи дикторов.

Согласные были разделены на 29 классов, в ряде случаев пересекающихся.

К самому узлу дерева (центральному элементу трифона) могло быть задано 57 вопросов. Два из этих вопросов идентифицировали принадлежность наблюдений в узле одному из состояний СММ.

Вопросы к левому и правому контексту были идентичными: 40 вопросов проверяли принадлежность звуков к широким фонетическим классам и 58 вопросов идентифицировали конкретный звук (заметим, что взрывные и аффрикаты трактовались в настоящем исследовании как сочетание двух отдельных звуков — смычки и взрыва). Последний тип вопросов был ориентирован на те случаи, когда отдельный звук был представлен в данном контексте достаточным количеством фреймов в обучающей выборке.

При построении дерева решений использовались следующие пороговые величины: минимальное приращение логарифма правдоподобия — 6.0, минимальная заселённость терминального узла ≥ 150 фреймов.

Результаты

На первом шаге построения дерева (*см. рис. 1*) основой для расщепления корневого узла «Все гласные» послужило не фонетическое качество гласного или характер контекста, а противопоставление конечной трети любого гласного его предшествующей части. В предварительном эксперименте, когда обучающая выборка была увеличена на три часа за счёт привлечения речевого материала, записанного тремя дикторами-женщинами, разбиение корневого узла произошло аналогичным образом.

Фрагмент дерева решений (верхушка)

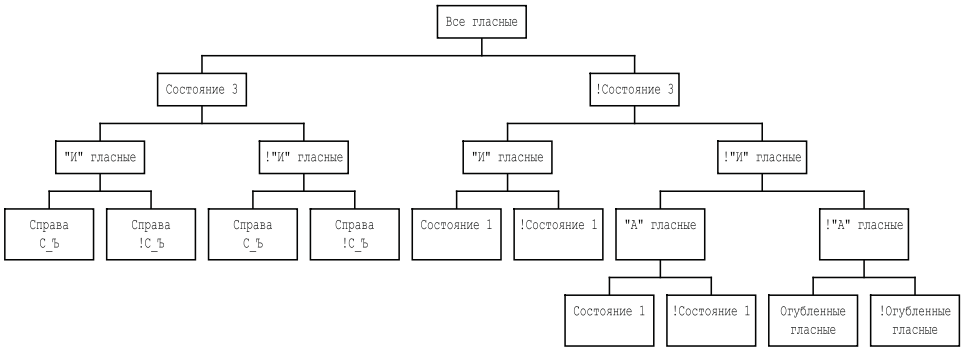

Рис. 1. Фрагмент построения глобального дерева решений (начальные шаги)

На рисунках восклицательный знак перед словом означает логическое отрицание, т.е. выражение «!Состояние 3» равноценно «Состояния 1 и 2»; С_ъ — означает любой твёрдый согласный.

На следующем шаге узлы состояний в свою очередь были расщеплены на узлы: [и]-образные гласные и все остальные. В класс [и]-образных гласных входят ударные и безударные аллофоны фонем /и/ и /е/, а также акустически нечленимые

сочетания безударных гласных, как, например, в окончании слов «Эстонии», «многие» и т.п. Последующее разделение узлов ветви «Состояние 3» зависело от твёрдости/мягкости согласного справа.

Продолжение формирования ветви, исходящей из родительского узла с характеристикой: «Состояние 3», [и]-образные гласные, правый контекст — любой твёрдый согласный», — представлено на *рис. 2*. Для краткости в описаниях узлов во всех случаях опущено, что вопрос, относящийся к согласному, адресован к правому контексту. Отметим, что один из применённых вопросов идентифицирует качество самого гласного, а именно, является ли он ударным [е]. Прямоугольники, нарисованные пунктирными линиями, являются терминальными узлами.

Построение дерева было остановлено по критерию приращения коэффициента правдоподобия. Заметим, что и зафиксированная минимальная величина заселённости классов (155 фреймов) приблизилась к критическому значению. Результирующее дерево имело 156 терминальных узлов[1]. По состояниям СММ они распределились практически равномерно: 46 узлов — «Состояние 1», 43 узла — «Состояние 2», 48 узлов — «Состояние 3» и в 19 случаях терминальный узел был построен на первых двух состояниях совместно. В последнем случае большинство элементов этого множества составляли гласный [ы] и безударные гласные после твёрдого согласного, традиционно транскрибируемые с помощью знака [ъ], а также ряд назализованных гласных.

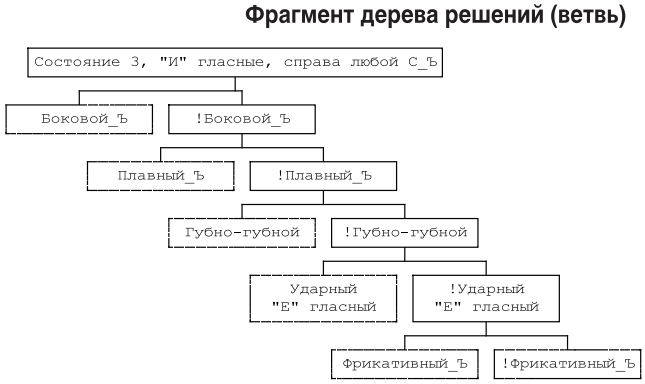

Рис. 2. Фрагмент построения глобального дерева решений (ветвь для родительского узла «Состояние 3», [и]-образные гласные, правый контекст — любой твёрдый согласный»)

В ходе построения дерева было использовано 64 различных вопроса, употреблённых в сумме 155 раз. 77% заданных вопросов относились к окружению гласного.

Левый контекст оказался значимым в 64 случаях при использовании 29 вопросов, из которых 16 относились к твёрдым согласным (39 употреблений), 12 — к мягким (24 употребления) и один — к классу [а] образных гласных (одно употребление).

Правый контекст оказался определяющим в 56 случаях при использовании 20 разных вопросов, из которых 12 относились к твёрдым согласным (41 употребление), пять — к мягким (девять употреблений) и три — к классам [а, у, е]-образных гласных (три употребления). Причём в семи случаях вопросы к левому контексту проверяли наличие конкретных звуков: [в, м, д, т, н, з', р',]; в правом контексте идентифицировался звук [с'].

[1] При увеличении критического значения приращения коэффициента правдоподобия до 11.0 построение дерева было завершено, когда число терминальных узлов равнялось 102. Подробное описание полученного инвентаря субаллофонов дано в [3, 8].

Качество самого гласного оказалось значимым в 26 случаях при использовании 13 вопросов. Эти вопросы относились как к классам гласных, так и отдельным звукам. Так, например, оказалось важным, является ли гласный [а]-образным ротовым или назализованным, или же ударным [а]. Заметим, что с назализованными гласными было связано четыре вопроса. В девяти случаях идентифицировалось состояние гласного: один раз «Состояние 3» и восемь раз «Состояние 1».

Как и ожидалось, для узла «Состояние 3» релевантными были вопросы к правому контексту. Только в одном случае был использован вопрос о мягкости левого согласного для класса огубленных гласных, справа от которых находился твёрдый согласный. В результате был образован соответствующий терминальный узел. Для узла «Состояние 1» вопросы адресовывались за редким исключением к левому контексту, и для узла «Состояние 2» имели значение вопросы к обоим контекстам.

Терминальные узлы с характеристикой «Состояние 1» и «Состояние 2» распределены между классами гласных следующим образом. Наибольшее число узлов приходится на гласный [а]: 15 на «Состояние 1» и 14 на «Состояние 2». Причём преобладают [а] в ударном и первом предударном слогах.

Следующие по числу занимаемых узлов идут [и]-образные гласные (передние, средне-верхнего подъёма): 13 — на «Состояние 1» и 10 — на «Состояние 2». В последнем случае из [и]-образных гласных исключены ударные [е], которые образовали три отдельных узла на «Состояние 1» и один на «Состояние 2». На [ы]-образные гласные пришлось по два узла на «Состояние 1» и «Состояние 2» и 14 на их объединение.

Огубленные гласные образовали на «Состояние 1» 11 узлов и на «Состояние 2» — семь узлов.

Следует упомянуть ещё об одном классе гласных, на который пришлось пять узлов «Состояние 1», шесть узлов — «Состояние 2» и четыре на оба состояния совместно. Этот класс — фонетически неоднородный и образуется в основном назализованными гласными, из которых в ряде контекстов исключаются огубленные назализованные.

По сравнению с первыми двумя состояниями классификация гласных на последнем состоянии характеризуется с традиционной точки зрения большей неопределённостью. Так, например, 16 узлов «Состояние 3» приходятся на такой класс, как «все гласные, за исключением [и]-образных». Ещё один красноречивый пример, в некотором смысле противоположный первому: два узла образованы для всех гласных, за исключением [и]-образных, [а]-образных, огубленных гласных, назализованных [а]-образных, назализованных [и]-образных и назализованного ударного [о]. В остатке имеем ротовые и назализованные [ы]- и [у]-образные гласные в ударных и безударных слогах.

На огубленные гласные приходится восемь узлов «Состояние 3», столько же — на [а]-образные.

Следует особо отметить, что в 37 случаях терминальные узлы были образованы для отдельных гласных. В частности, 27 узлов принадлежали [а], находящемуся в первом предударном и/или в ударном слогах.

Наибольшее значение для расщепления узлов дерева имели следующие признаки согласных звуков: твёрдый/мягкий, боковой, плавный, носовой, губно-губной, глухой/звонкий и фрикативный.

Обсуждение и выводы

Как показывают полученные результаты, предложенный метод представления спектральной динамики гласных в виде комбинации субаллофонов оказался достаточно эффективным. Классификация субаллофонов опирается как на чисто акустические параметры, так и на экспертные фонетические знания. Обращает на себя внимание высокая пластичность полученных классов. С одной стороны, это могут быть очень широкие классы, как, например, класс: «Состояние 3», все гласные, кроме [и]-образных, с другой стороны, несколько терминальных узлов были образованы только для ударного [а]. В ряде случаев «Состояние 1» и «Состояние 2» объединялись в одном классе.

Полученные результаты опровергают традиционное представление о том, что для характеристики гласного основное значение имеет левый контекст. Во-первых, корневой узел «все гласные» был разделён вопросом о принадлежности векторов параметров «Состоянию 3», свойства которого определяются правым контекстом. Во-вторых, количество терминальных узлов для «Состояния 1» и «Состояния 3» практически совпадает: 46 и 48. Число вопросов к левому и правому контексту — величины одного порядка: 64 и 56 вопросов соответственно.

Образование нескольких классов для назализованных гласных подтверждает целесообразность использования этого признака для характеристики русских гласных.

Классификация согласных звуков

Классы звуков и набор бинарных вопросов. Согласные звуки были представлены 48 классами, которые могли пересекаться, и 59 звуками, в число которых входили смычки и взрывы, а также вокалический компонент и удар вибранта. Классы согласных формировались как с учётом, так и без учёта признака «твёрдость/мягкость». В отдельный класс вошли сегменты разметки, указывающие на границы предложения (например, пауза, вдох и т.п.). Число классов гласных звуков было равно 25.

Общее число проверяемых вопросов составило 373. Из них 107 были адресованы к центральному элементу трифона (к узлу дерева), остальные вопросы — к левому и правому контексту.

Как и в случае классификации гласных звуков при построении дерева решений использовались следующие пороговые величины: минимальное приращение логарифма правдоподобия — 6.0, минимальная заселённость терминального узла $\geq$ 150 фреймов.

Результаты

Характеристика терминальных узлов (листьев) дерева. Построенное дерево имело 192 терминальных узла. В 142 случаях в узле содержался только один элемент (аллофон), в 26 случаях —

2 элемента и в 24 случаях число аллофонов было ≥ 3. Только правый контекст учитывался 34 раза, только левый — 40 раз, оба контекста оказались одновременно значимыми в 114 случаях. Независимыми от контекста были четыре взрыва смычных согласных: [P', B, K', G]. Более 30% аллофонов пришлось на фрикативные согласные, две трети которых были твёрдыми.

Второй по численности группой (более 25%) оказались глухие и звонкие смычки, а также удары вибранта. Доли взрывов, носовых и плавных колебались в районе 10%. Как и в случае с фрикативными согласными преобладали твёрдые аллофоны. В классы, состоящие из двух элементов, как правило, попадали фонетически близкие звуки: например, мягкие плавные [L', R'], йот и один из мягких плавных [J' + L'/R'], фрикативный и шумовой компонент аффрикаты [S, TS] и т.п.

Процесс построения дерева не противоречил принципам классификации, характерным для традиционной фонетики. Сначала весь массив наблюдений (класс «все согласные») был разбит на мягкие и немягкие согласные, которые, в свою очередь, были затем разделены на звонкие и глухие, а от группы твёрдых и мягких звонких были отделены соответствующие носовые. Начальная фаза формирования дерева представлена на *рис. 3*. Восклицательный знак обозначает логическое отрицание, апостроф после знака транскрипции обозначает мягкость согласного.

Анализ применённых вопросов. В ходе построения дерева было использовано 78 различных вопросов. Всего к центральному элементу было адресовано 49 вопросов (30 разных). Вопрос о мягких согласных был задан 20 раз, о твёрдых — 25. В 18 вопросах речь шла о твёрдых и мягких губно-губных и губно-зубных согласных.

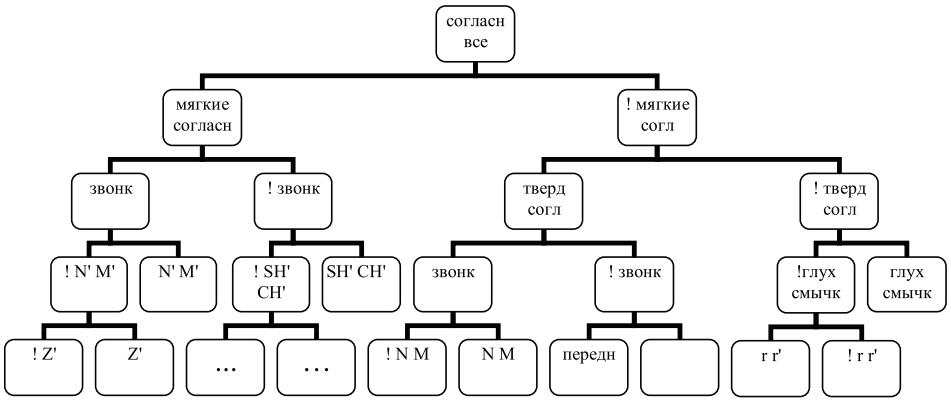

Рис. 3. Начальный этап построения дерева (обозначения см. в тексте)

Вопросы к левому контексту были использованы всего 78 раз (24 разных вопроса). В 41 случае в качестве левого контекста выступали согласные звуки и граница предложения (девять раз). Причём преобладали вопросы о твёрдых согласных. В 37 случаях левый контекст определялся гласными. Разных вопросов было пять. Класс и-образных гласных использовался 18 раз, класс а-образных гласных — 12 раз и огубленные гласные встретились пять раз.

К правому контексту распределение вопросов было следующим. Всего было применено 64 вопроса. Вопросов о согласных и границе было 53 (20 разных). В 21 случае в качестве правого контекста мог выступать любой согласный. Твёрдые согласные встретились 10 раз, мягкие — восемь раз и граница предложения — семь раз. 11 раз в качестве правого контекста выступали гласные: пять раз — огубленные гласные, четыре раза — класс а-бразных гласных.

Обсуждение и выводы

Оценивая в целом полученную классификацию согласных, в первую очередь следует отметить, что только около 10% терминальных узлов состояли из нескольких элементов. В подавляющем большинстве случаев разбиение доходило до конкретного аллофона. Последовательность построения дерева решений хорошо согласуется с преставлениями фонетистов о значимости классифицирующих признаков. Вполне ожидаемо, фрикативные согласные по числу аллофонов оказались на первом месте и аллофонов твёрдых согласных было больше, чем мягких.

Однако с точки зрения традиционной (академической) фонетики трудно объяснить тот факт, что вторую группу по численности составляют глухие и звонкие смычки, а также удары вибранта. Дело в том, что реальные условия записи речевого материала не обеспечивали получение идеальных характеристик смычек: отсутствие сигнала для глухих или наличие только «голосовой полосы» для звонких смычек. Запись дикторов проходила в тихой обстановке в обычном рабочем помещении, что делало неизбежным присутствие эффектов реверберации в записанном материале. На *рис. 4* представлена спектрограмма и спектральные срезы фрагмента слова «апатиты», стоящего в начале предложения. Спектральные срезы были сделаны с длиной окна анализа 25 мс. Спектры были получены методом БПФ и ЛПК (гладкая спектральная огибающая). Можно видеть, что спектральные характеристики паузы (соответствующее окно анализа выделено на спектрограмме белыми вертикальным линиями) существенно отличаются от спектра смычки [p] и второй смычки [t]. Причём спектр смычек в значительной степени определяется характером предшествующего гласного. Чтобы убедиться в этом, достаточно сравнить приведённые спектральные срезы для гласного [A] и последующей смычки [p]. Можно видеть, что местоположение двух первых спектральных максимумов практически идентично. Не вызывает также сомнения и-образный характер спектра смычки [t]. Целесообразность дифференциации смычек в зависимости от контекста должна быть проверена в дополнительных экспериментах по распознаванию речи.

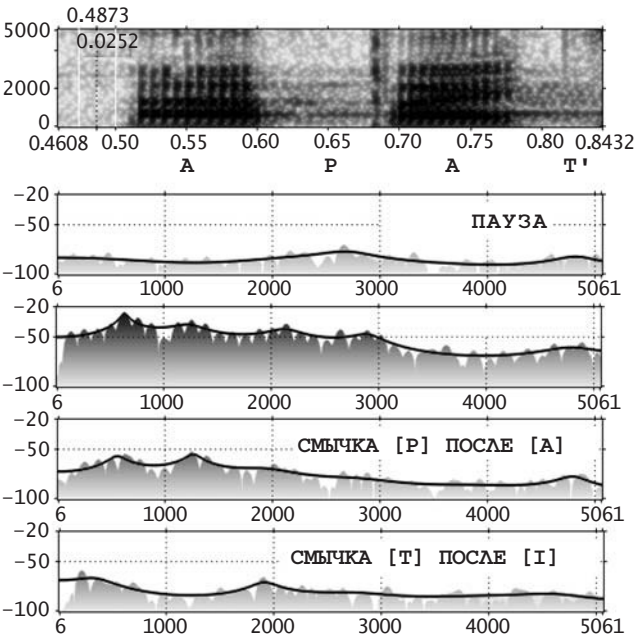

Рис. 4. Спектрограмма фрагмента слова «апатиты» и спектральные срезы, показывающие зависимость акустических характеристик глухих смычек от контекста

Как и в случае классификации гласных звуков [2, 3], для согласных оказался значимым как левый, так и правый контекст. Обращает на себя внимание тот факт, что в качестве левого контекста гласные выступили 37 раз, а в качестве правого только 11. Причём если в левом контексте преобладали и-образные гласные, то в правом они отсутствовали.

Следует отметить положительный результат, который дало отнесение взрыва аффрикат к фрикативным согласным. Результаты настоящего исследования подтверждают обоснованность использования в конкатенативном синтезе отдельных аллофонов или дифонов, когда их ближайшим соседом слева или справа оказывается пауза.

В заключение несколько общих соображений о проведённом исследовании. В настоящий момент не представляется возможным оценить устойчивость и типичность полученной классификации. Она жёстко привязана к конкретному речевому материалу, способу параметризации речевого сигнала, варианту реализации СММ, набору бинарных вопросов, критериям построения дерева решений и т.д. Проведение исчерпывающих исследований значимости этих факторов маловероятно по причине их высокой трудоёмкости. Так, в настоящей работе для построения дерева решений потребовалось около 12 часов непрерывной работы компьютера средней мощности.

Однако есть возможность постепенного сбора необходимых данных. Дело в том, что на этапе обучения распознающей системы построение классификаций аналогичных нашей, является обычным делом. Но для специалистов по речевой технологии этот результат самостоятельной ценности не представляет и, как правило, не комментируется. Ситуация может измениться, если фонетисты сами проявят инициативу, чтобы получить доступ к этим данным. Потребность во всестороннем анализе этих результатов исключительно высока. По мнению крупнейших учёных в области фундаментальной и прикладной фонетики, Г. Фанта [7] и Б. Линдблома [10], фонетическая вариативность речи огромна, но не случайна и в значительной мере систематична. Необходимо длительное и кропотливое исследование реализации звуков речи во всевозможных контекстах, чтобы по мере накопления данных удалось в конце концов, используя информацию о контексте в самом широком смысле, расклассифицировать фонетическую вариативность и снять её кажущуюся неопределённость. Тогда-то и проявится зашифрованная структура речевого сигнала.

Работа выполнялась при поддержке проектов РФФИ № 07-01-00657а и № 06-08-01534а.

Литература

1. *Кривнова О.Ф., Захаров Л.М., Строкин Г.С.* Речевые корпусы (опыт разработки и использование) // Труды международного семинара «Диалог'2001 по компьютерной лингвистике и её приложениям». Аксаково, 2001.

2. *Кузнецов В.Б.* О принципах акустической классификации русских гласных // Язык и речь: проблемы и решения. М.: МГУ, 2004. С. 100–117.

3. *Кузнецов В.Б., Чучупал В.Я.* Инвентаризация гласных аллофонов русской речи методом кластеризации сверху-вниз бинарных деревьев решений // Акустика речи. Медицинская и биологическая акустика: Сб. т. XIII сессии Российского акустического общества. Т 3. М.: Геос, 2005. С. 54–57.

4. *Потапова Р.К.* Речь: коммуникация, информация, кибернетика. М.: Радио и связь, 1997.

5. *L.Breiman, J.Friedman* etc. «Classification and Regression Trees», Wadsworth, Inc, 1984.

6. Речевая связь с машинами (тематический выпуск) // ТИИЭР. 1985. Т. 73. № 11.

7. *Fant G.* On the speech code // Speech, Music and Hearing, KTH, Stockholm, Sweden, TMH-QPSR, Vol. 42, 2001, p. 61–72.

8. *Kouznetsov V., Chuchupal V.* Increasing trainability of ASR system by means of top-down clustering procedure based on decision trees (vowel data for Russian) // Proc. Intern. Workshop "Speech and Computer", SPECOM'04, St.-Petersburg, 2004, p. 289–291.

9. *Kouznetsov V., Chuchupal V., Makovkin K., Chichagov A.* Design and implementation of the Russian telephone speech database. // Proc. Intern. Workshop "Speech and Computer", SPECOM'99, Moscow, 1999, p. 179–181.

10. *Lindblom B.* Developmental origins of adult phonology. The interplay between phonetic emergents and the evolutionary adaptations of sound patterns // Phonetica Vol. 57, No. 2–4, 2000, p. 5–30.

11. *Nakajima Sh., Hamada H.* Automatic generation of synthesis units based on context clustering // Proc. ICASSP-88. 1988, Avr. N.Y. p. 659–662.

В.Б. Кузнецов —

профессор кафедры прикладной и экспериментальной лингвистики Московского государственного лингвистического университета. Сфера научных интересов — фундаментальные исследования процессов речепроизводства и восприятия речи (преимущественно на материале гласных звуков русского языка) и приложения в речевых технологиях (синтез речи, распознавание языка сообщения, распознавание речи). Автор более 100 научных и научно-методических публикаций, в том числе монографий «Автоматический синтез речи. Алгоритмы преобразования «буква-звук» управление длительностью речевых сегментов» (1989) и «Лингвистическое обеспечение систем синтеза речи по правилам: достижения, проблемы и перспективы» (1992).

В.Я. Чучупал —

закончил МГПИ им. В.И.Ленина в 1976 году по специальности «математика», в 1983 году закончил очную аспирантуру ВЦ АН СССР, научный руководитель — В.Н. Трунин-Донской, с тех пор работает в ВЦ РАН. Основная область интересов: распознавание и обработка речевых сигналов. Кандидат физико-математических наук, ведущий научный сотрудник.