

УДК 510.62

*З. Ю. МАЛЕНЧЕНКО, Г. А. ПРАСОЛ*

**О МОДЕЛИРОВАНИИ СВЯЗЕЙ МЕЖДУ ЗВУКОВОЙ  
И ТЕКСТОВОЙ ФОРМАМИ РЕЧЕВОГО СООБЩЕНИЯ**

Человек обладает способностью превращать текстовые сообщения в звучащую речь, он также может записать речь с голоса. Задача автоматизации этой способности человека имеет большое народнохозяйственное значение. Ее решение позволило бы автоматизировать труд стенографа, создать аппаратуру, озвучивающую сообщения, дало бы возможность оборудовать вычислительные системы устройствами речевого ввода и вывода

информации. Создание аппаратов подобного назначения обеспечило бы заметное продвижение вперед по всему фронту работ автоматизации труда человека. Представляется реальной постановка задачи о создании в обозримом будущем специального прибора с размерами транзисторного приемника, выполняющего анализ и синтез звучащей речи.

Анализ речи заключается в автоматическом восприятии разборчиво звучащего речевого сообщения, продуцируемого любым человеком — носителем русского языка. В результате осуществляется световая индикация отрезка текста этого сообщения длиной в несколько слов в виде бегущей строки. Кроме того, производится передача двоичного кода текста сообщения к вычислительному устройству или печатающему аппарату по специальному каналу связи. Формируемый прибором текст сообщения не будет безупречным в орфографическом отношении. В нем будут учтены лишь основные фонетические и орфографические правила (например, звучание *синевы, дуп, солнце* будут преобразованы в тексты *синего, дуб, солнце*). Прибор не сможет преобразовать звучание *чек* и *Сан Саныч* в тексты *человек* и *Александр Александрович* и сохранит их в первоначальном виде. Синтез речи состоит в формировании звучащего сообщения по коду текста, вводимого в аппарат извне (например, от вычислительной машины), голосом того диктора, который непосредственно перед этим обращался к аппарату с речью. Голос диктора будет имитироваться весьма приблизительно: автомат не сможет «комкать» слова, управлять интонацией речи и выполнять другие тонкие действия, доступные человеку.

Создание прибора с описанными выше функциями должно основываться на работах по моделированию способности человека к восприятию речи и к речеобразованию. Это моделирование, в свою очередь, должно базироваться на изучении и математическом описании связей, имеющих место между звучащей и текстовой формами речевого сообщения. Звук и текст чрезвычайно далеки друг от друга, их можно уподобить противоположным берегам очень широкой реки. Выявить и описать отношения, связывающие звук и текст, не проще, чем перебросить мост через такую реку. Одним пролетом широкую реку не перекрывать нужны промежуточные опоры. Подобно этому, для успеха моделирования нужно ввести между звуком и текстом несколько промежуточных объектов: 1) первичный портрет звука; 2) вторичный портрет звука; 3) акустические признаки звука; 4) фонетические признаки звука; 5) фонему звука.

При переходе от звучащей речи к ее первичному портрету используется эффект сглаживания в слухе [1]. Осциллограмма звука преобразуется в некоторую последовательность биполярных импульсов. Вся информация о звуке речи в процессе этого преобразования сохраняется: озвучивание первичного портрета звука приводит к звучанию, неотличимому на слух от исходного звучания. В роли первого объекта моделирования будет высту-

пать отношение, связывающее осциллограмму звука с его первичным портретом. При переходе от первичного портрета звука ко вторичному происходит отстройка портрета от громкости и определение уровня громкости речевого сообщения, отстройка от высоты тона и определение высоты тона речевого сообщения, отстройка от тембра звука, формируемого голосовыми связками говорящего, и определение характеристик тембра, отстройка от индивидуальных особенностей резонирования в речеобразующем тракте человека и определение характеристик этого резонирования, отстройка от индивидуальных особенностей шумообразования в органе речи и определение характеристик этого шумообразования. В процессе обратного перехода от вторичного портрета к первичному происходит настройка портрета на нужные громкость, высоту тона, тембр голоса, на нужные режимы резонирования звуков и генерирования шумов. В роли второго объекта для моделирования выступает отношение, связывающее первичный и вторичный портреты речевого сообщения.

При переходе от вторичного портрета звука к набору его акустических признаков производится разложение звука на его физически значимые компоненты. В результате производится членение портрета звука на акустически однородные участки, оценка относительной длительности этих участков, оценка на этих участках амплитуды сигналов, их частотного состава, выявление характерных особенностей рисунка сигнала и т. д. В процессе обратного перехода по значениям акустических признаков синтезируется вторичный портрет звука. Третьим объектом моделирования служит отношение, связывающее вторичный портрет с набором акустических признаков звука. При переходе от акустических признаков звука к фонетическим производится разложение звука на его психологически значимые компоненты. В результате формируются фонетические характеристики звука: гласный — согласный, сонорный — шумный, твердый — мягкий, долгий — краткий, смычный — щелевой и т. д. В процессе обратного перехода определяются значения акустических признаков по заданным значениям фонетических признаков. Четвертым объектом моделирования служит отношение, связывающее между собой акустические и фонетические признаки звука.

При переходе от значений фонетических признаков к фонетическому представлению звука формируется фонема, т. е. знак, обозначающий класс звуков, тяготеющих к определенной букве текста. Обратное преобразование состоит в замене фонетического знака набором соответствующих ему значений фонетических признаков звука. Отношение, связывающее фонему с набором признаков звука, может служить пятым объектом для моделирования. На последнем этапе вводится отношение (шестой объект для моделирования), связывающее фонетическую запись текста с его орфографической записью. Переход от фоне-

тической записи текста к орфографической и обратно совершается на основе правил языка, регулирующих произношение (фонетические правила) и написание (орфографические правила) слов. Ниже описывается понятие фонемы, предназначенное для целей моделирования связей между звуковой и текстовой формами речевого сообщения.

Анализ литературных данных из области фонетики показывает, что приемлемого для наших целей понятия фонемы не существует. Поэтому нам придется ввести понятие фонемы несколько отличное от уже известных. Определяя понятие фонемы, мы будем исходить из следующего критерия: фонетические знаки должны быть максимально приближены к орфографическим знакам, т. е. к буквам текста. Удовлетворяя этому критерию, мы максимально сократим различие между фонетическим и орфографическим текстом речевого сообщения и этим облегчим моделирование весьма сложного отношения, связывающего эти тексты. Лучше всего было бы в качестве фонем принять все буквы русского алфавита. Однако в точности это сделать невозможно: буквы ь, ъ при всем желании нельзя трактовать как звуки речи, а буква щ — это либо мягкое ш, либо последовательность из двух звуков шч. Остальные 30 букв русского алфавита допускают вполне определенную фонетическую интерпретацию, их мы и примем для обозначения фонем.

В *понятии фонемы* можно выделить три стороны. Во-первых, это просто знак; во-вторых, это класс некоторых звуков: в-третьих, это нечто, тяготеющее к одноименной букве орфографического текста. Полностью отождествить фонему с орфографической буквой нельзя (например, звучит *дун*, а пишется *дуб*). Полную математическую характеристику фонемы как знака мы получим, вводя переменную фонему  $x$  со значениями из множества {а, б, в, г, д, е, ё, ж, з, и, й, к, л, м, н, о, п, р, с, т, у, ф, х, ц, ч, ш, ы, э, ю, я}. Это определение фонемы можно записать в виде следующего индивидуального предиката:  $\text{Фон}(x) = x^a \vee x^b \vee x^в \vee x^г \vee x^д \vee x^е \vee x^ё \vee x^ж \vee x^з \vee x^и \vee x^й \vee x^к \vee x^л \vee x^м \vee x^н \vee x^о \vee x^п \vee x^р \vee x^с \vee x^т \vee x^у \vee x^ф \vee x^х \vee x^ц \vee x^ч \vee x^ш \vee x^щ \vee x^ь \vee x^ъ \vee x^я$  (1).

Формула, стоящая справа от знака равенства, записана на языке алгебры конечных предикатов [2]. Утверждение, что  $x$  есть фонема, эквивалентно уравнению  $\text{Фон}(x) = 1$ .

Может вызвать возражение то обстоятельство, что в числе фонетических знаков не включены мягкие фонемы б<sup>б</sup>, в<sup>в</sup>, г<sup>г</sup> и т. д., как это обычно делается в фонетике. Мы, однако, считаем, что такое решение отдалило бы алфавит фонем от алфавита букв орфографического текста. Вместе с тем удваивать число согласных фонем для наших целей нет никакой необходимости, этого можно достичь иным путем, вводя признак палатализации звука со значениями — твердый и мягкий. Нет никаких препятствий для использования этого признака при описании связи между фонетической и орфографической записями

текста. Другое возможное возражение состоит в том, чтобы не вводить десять знаков для обозначения гласных звуков, а оставить, к примеру, только пять, как это часто принимается авторами фонетических систем [3]. Действительно, вводя десять гласных фонем, мы допускаем, по классическим понятиям, некоторую избыточность в числе фонем, но зато максимально приближаем алфавит фонетических знаков к алфавиту орфографических букв.

Во введенной системе фонетических знаков можно усмотреть еще и тот недостаток, что она игнорирует часто встречающиеся в русской речи звуки: щелевое г ( $\bar{г}$ ) и звонкие ц, ч ( $\bar{ц}$ ,  $\bar{ч}$ ), присутствующие, к примеру, в таких текстах: *Учгиз, Гинцбург*. Этот недостаток системы нейтрализуется тем, что фонема г, ч, ц трактуются в ней расширительно — и как звуки г (смычное), ч, ц, (глухие), и как звуки  $\bar{г}$ ,  $\bar{ч}$ ,  $\bar{ц}$ , а выбор конкретного звука из каждой пары регулируется значениями признаков смычной — щелевой (для звуков г —  $\bar{г}$ ) и глухой — звонкий (для звуков ч —  $\bar{ч}$ , ц —  $\bar{ц}$ ).

Еще один недостаток вводимой фонетической системы можно усмотреть в том, что в ней отсутствуют специальные знаки для обозначения гласных звуков в безударном положении, имеющих стертое, промежуточное звучание по сравнению со звучанием ударных гласных. Это возражение отводится благодаря тому, что в случае нечеткого звучания гласного звука прибор, автоматически анализирующий звуки речи, будет выдавать сразу все множество (или его часть) возможных в данном случае гласных фонем, поскольку такая возможность предоставляется алгеброй конечных предикатов. Например, первое о в слове *домой* будет распознано прибором как фонема из множества {а, о}. Вряд ли прибор сможет сделать это точнее, поскольку ухо принимает как правильный тот и другой звук, а значит будут встречаться на практике оба варианта произношения как допустимые. Вместе с тем, этой неполной информации, как мы полагаем, будет достаточно, чтобы однозначно выбрать орфографический знак, удовлетворяющий правилам фонетики и орфографии русского языка.

Остановимся на звуковой характеристике некоторых из введенных фонем. Гласные фонемы а, о, у, ы, э — это непалатализованные звуки, которые звучат, например, в словах *там, торт, тут, тыл, пэр*. Гласные фонемы я, ё, ю, и, е — это палатализованные звуки, присутствующие, к примеру, в словах *дядя, тётя, ключ, синь, тень*. Фонема й звучит, например, в начале слова *яма* и в конце слова *сарай*. Гласные фонемы, а также й и сонорные л, м, н, р могут звучать как звонко (речь голосом), так и глухо (речь шепотом). Согласные б, в, г, д, ж, з при шепоте превращаются в фонемы п, ф, к, т, ш, с, поэтому при речи шепотом число фонем сокращается до 24. Формальное определе-



у	л	у	к	о	м	о	р	ь	я	д	у	б	з	е	л	ё	н	ы	й
у	л	у	к	а	м	о	р	й	ё	д	у	б	з	е	л	ё	н	ы	й
у	л	у	к	а	м	о	р	й	ё	т	у	п	с	е	л	ё	н	ы	й
т	т	т	т	т	т	т	т	м	м	т	т	т	м	м	м	т	т	т	т
з	л	а	т	а		я	ц	е	п	ь	н	а	д	у	б	е	т	о	м
з	л	а	т	а	й	е	ц	е	п		н	а	д	у	б	е	т	о	м
с	л	а	т	а	й	е	ц	е	п		н	а	д	у	п	е	т	о	м
т	т	т	т	т	т	м	м	м	м		т	т	т	т	м	м	т	т	т
и	д	н	ё	м	и	н	о	ч	ь	ю	к	о	т	у	ч	ё	н	ы	й
ы	д	н	ё	м	э	н	о	ч	й	ю	к	о	т	у	ч	ё	н	ы	й
ы	т	н	ё	м	э	н	о	ч	й	ю	к	о	т	у	ч	ё	н	ы	й
т	т	м	м	т	т	т	т	т	м	м	т	т	т	т	м	т	т	т	м
в	с	ё	х	о	д	и	т	п	о	ц	е	п	и	к	р	у	г	о	м
в	с	ё	х	о	д	и	т	п	о	ц	э	п	и	к	р	у	г	о	м
ф	с	ё	х	о	т	и	т	п	о	ц	э	п	и	к	р	у	х	о	м
т	м	м	т	т	т	м	т	т	т	т	т	м	м	т	т	т	т	т	г

Введем признак вокализации звука  $y_1$  со значениями в — вокализованный, н — невокализованный. Вокализованный звук — это звук произносимый при работающих голосовых связках, т. е. с участием голоса. Невокализованный звук произносится без голоса, при выключенных голосовых связках. При речи шепотом все звуки невокализованы. При речи голосом невокализованы глухие согласные, а гласные и звонкие согласные вокализованы. Осциллограмма вокализованного звука обладает характерной особенностью — квазипериодичностью, невокализованный звук такой особенностью не обладает [4]. Поэтому имеется возможность уверенного извлечения из звучащей речи значений признака вокализации. Можем записать  $y_1^в \vee y_1^н = 1$  (3).

Человек может вести свою речь в двух режимах — голосом и шепотом. В соответствии с этим фактом вводим признак режима речи  $y_2$  со значениями г — речь голосом, ш — речь шепотом. В процессе нормальной речи голосом голосовые связки на интервале времени в одну секунду обязательно включаются в работу. Этот факт дает возможность надежного разделения речи по ее звучанию на речь голосом и речь шепотом. Если в течение секунды переменная  $y_1$  ни разу не приняла значения в, то  $y_2 = ш$ , в противном случае  $y_2 = г$ . Полагаем, что  $y_2^г \vee y_2^ш = 1$  (4).

Признаки  $y_1$  и  $y_2$  определенным образом связаны между собой. Эту связь можно формально записать в виде следующего уравнения:  $y_2^ш \supset y_1^н = 1$  (5). Это уравнение означает, что в речи шепотом все звуки невокализованы.

Каждому из допустимых наборов признаков  $y_1, y_2$  соответствует одна или несколько фонем  $x$ . Это соответствие может быть формально описано следующим уравнением:  $(y_1^в \vee y_2^ш) Гсй(x) \vee y_1^н Зшс(x) \vee y_1^н Гг(x) \vee Афф(x) = 1$  (6), где  $Гсй(x) = x^а \vee x^е \vee x^и \vee x^о \vee x^у \vee x^ы \vee x^ю \vee x^я \vee x^л \vee x^м \vee x^н \vee x^п \vee x^р$  (7),  $Зшс(x) = x^б \vee x^в \vee x^г \vee x^д \vee x^ж \vee x^з$  (8),  $Гг(x) = x^к \vee x^п \vee x^с \vee x^т \vee x^ф \vee x^х \vee x^ш$  (9),  $Афф(x) = x^ц \vee x^ч$  (10).

Здесь предикат  $Гсй(x)$  формализует понятие гласного или сонорного или звука й, предикат  $Зшс(x)$  — звонкого шумного согласного,  $Гг(x)$  — глухого согласного,  $Афф(x)$  — аффрикаты. Содержательно уравнение (6) с учетом отношений (3) — (5) означает, что: 1) гласные, сонорные и й в речи голосом всегда вокализованы, в речи же шепотом они всегда невокализованы; 2) звонкие и шумные согласные всегда вокализованы; 3) глухие согласные, кроме аффрикат ц и ч, всегда невокализованы; 4) аффрикаты могут встречаться при любом допустимом (в смысле уравнения (5)) наборе значений признаков  $y_1$  и  $y_2$ , они в речи голосом могут быть как вокализованы ( $\bar{ц}, \bar{ч}$ ), так и невокализованы.

Произнося гласные, сонорные или й шепотом, мы обнаруживаем, что слушатель воспринимает их как те же самые фонемы, произнесенные голосом. Таким образом, эти фонемы устойчивы к

изменению режима речи. Если же мы попытаемся произнести шепотом звонкие шумные согласные, то увидим, что они становятся другими фонемами: б переходит в п, в — в ф и т. д. Правда, лицам с достаточно тонким слухом иногда удается в речи шепотом различить б от п, в от ф и т. д., основываясь на тончайших отличиях в произношении этих звуков. Этот факт мы, однако, в уравнении (6) не учитываем, полагая, что преждевременно ставить задачу о создании автомата, способного выполнять столь тонкую работу по распознаванию звуков речи. Пытаясь произнести голосом глухие согласные, мы всегда переводим их в другие фонемы: к превращается в г, х — в г̄, п — в б и т. д. Если же озвончатся аффрикаты, то ц переходит в ц̄, а ч — в ч̄. Таким образом, в этом случае мы остаемся в пределах той же самой фонемы.

**Список литературы:** 1. *Эффект сглаживания в слухе* / О. М. Абрамов, А. Я. Дрюченко, С. А. Усенко, Ю. П. Шабанов-Кушнарченко. — Проблемы бионики, 1977, вып. 19, с. 31—37. 2. *Шабанов-Кушнарченко Ю. П.* Об алгебре конечных предикатов. — АСУ и приборы автоматики, 1979, вып. 50, с. 14—20. 3. *Аванесов Р. И., Сидоров В. П.* Очерк грамматики русского литературного языка. — М.: Учпедгиз, 1945. — 236 с. 4. *Бондаренко Л. В.* Осциллографический анализ речи. — Л.: Высшая школа, 1965. — 46 с.

*Поступила в редколлегию 29.03.82.*