

УДК 519.76:681.3

*М. Ф. БОНДАРЕНКО, канд. техн. наук,
В. Л. НИКОЛАЕНКО*

МЕТОДИКА АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ ПО ИХ «ДИНАМИЧЕСКИМ ПОРТРЕТАМ»

Предметом внимания исследователей, занимающихся автоматической обработкой речевых сигналов, являются не отдельные звуки речи, а слоги, слова, фразы, т. е. отрезки речевого сигнала. Их длина в конечном итоге определяется выбранной частотой дискретизации аналогового сигнала и физическими возможностями операционных систем ЭВМ, которые связаны с ограничениями на размеры массивов, организуемых в оперативной памяти ЭВМ. Нами, с учетом психофизического эффекта сглаживания в слухе [1], выбрана частота дискретизации речевого сигнала, равная 38 кГц. А поскольку отладка алгоритмов обработки речевых сигналов осуществлялась в ДОС ЕС ЭВМ (что давало возможность организовать в оперативной памяти массив размерностью в 32000 элементов), мы имели возможность автоматически обрабатывать отрезки речевого сигнала длительностью 0,8 С. Таким образом, нам были доступны для анализа не только отдельные звуки и слоги, но и слова. Заметим, что по данным работы [2] средняя длительность произнесения слова для русской речи составляет 0,8 С, а средняя длина слова — 2,9 слога.

Известно, что открытые слоги в русском языке занимают доминирующее положение [3]. Наиболее распространенный тип слога — согласный-гласный (СГ). Как отмечено в [4], речь есть объединение в непрерывную последовательность открытых слогов, каждый из которых может содержать различное количество согласных. По данным работы [5], открытые слоги типа СГ и ССГ занимают соответственно 52, 34 и 27, 43% текста объемом 100550 слогов. Это обстоятельство определило наш выбор типа слога.

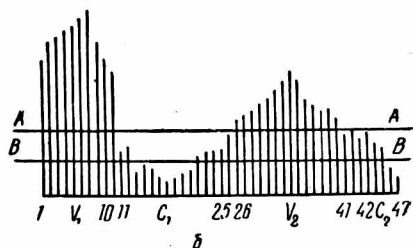
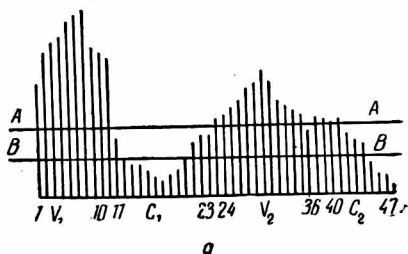
Были подобраны слова, содержащие согласные в положении между гласными с ударением на гласном, следующим за согласным. По мнению ряда исследователей, для опознавания согласного бóльшую роль играет переходной участок С—Г, чем Г—С. Поэтому в списке слов, предназначенном для анализа согласных, представлены всевозможные сочетания согласного с последующим гласным. В случае, когда сочетание согласного с гласным принципиально возможно в русском языке, но слова с таким сочетанием не зафиксированы, были выбраны псевдослова. Например, в русском языке не зафиксировано слово с сочетанием $[x'a]$, хотя оно в принципе возможно. В этом случае в список слов было введено слово «ахяна».

В качестве аудиторов к исследованиям привлекали сотрудников кафедры вычислительной техники и вычислительного центра Харьковского института радиоэлектроники: пять мужчин и две женщины, эпизодически десять мужчин и три женщины в возрасте от 20 до 40 лет, естественно пользующихся русской речью и не являющихся профессиональными дикторами. Стилль произношения — естественная разговорная речь.

Для записи речевых сигналов в оперативную память или на внешние запоминающие устройства ЭВМ (пакеты дисков ёмкостью 7,29 Мбайт) использовали микрофон МД-66А. Уровень шумов составлял 80 децибел.

Предварительный анализ речевых сигналов осуществляли с помощью так называемых «динамических портретов», отражающих динамику интенсивности речевых сигналов.

Опишем процедуру получения динамических портретов речевых сигналов. Все значения отсчетов параметрического кода речевого сигнала, полученного на выходе аналого-цифрового преобразователя устройства ввода речевых сигналов в ЕС ЭВМ, на отрезке определенной длительности заменим значением максимального отсчета. Полученный массив максимальных значений пронормируем по амплитуде с фиксированным репером, для каждого максимального значения получим нормирующий множитель, которым воспользуемся для построения динамического



«Динамический портрет» слова «один»: а — фиксированная длина временного интервала; б — длина временного интервала равна длине периода основного тона голоса

портрета отрезка речевого сигнала на АЦПУ ЭВМ (рисунок)

Замена значений отсчетов параметрического кода речевого сигнала на значение максимального отсчета, по нашему мнению, позволяет сохранить форму огибающей интенсивности речевого сигнала. Длина временного отрезка, на котором осуществляется такая замена, определяется целями тех задач, для решения которых привлекается данный динамический портрет. Нами была выбрана фиксированная длина временного отрезка, равная 10 мС, что гарантирует наличие хотя бы одного элементарного сегмента на выбранном отрезке речевого сигнала [6]. Длина временного отрезка может быть и переменной и равняться, в частности, длине периода основного тона голоса. Как показано на рисунке (позиция *а*), замена значений параметрического кода на значение максимального отсчета осуществлялась на фиксированном отрезке, равном 10 мС. Во втором случае (позиция *б*) — на отрезке, длина которого совпадала с длиной периода основного тона голоса. Как видно, отличия между динамическими портретами не существенны. С другой стороны, время, затрачиваемое ЭВМ на выделение периодов основного тона голоса, значительно больше, чем в первом случае, когда замена значений параметрического кода речевого сигнала осуществлялась на отрезке времени фиксированной длины.

Устанавливая определенные значения величины среза $A - A$ (пороговые значения), можно выделить в динамическом портрете характерные части: V_1 , V_2 , C_1 , C_2 (рисунок, позиция *а*).

Части V_1 и V_2 имеют то общее, что линия среза $A - A$ пересекает строки динамического портрета. Различаются же эти части шириной среза. Для V_1 она равна 10 (число строк динамического портрета, через которые проходит линия среза $A - A$), а для V_2 — 17. Акустическим эквивалентом частей V_1 и V_2 в слове «один», изображенном на рисунке, являются энергетические вокализованные участки, соответствующие звукам «о» и «и».

Аналогично, рассматривая части C_1 и C_2 , мы замечаем, что общим для них является то, что строки динамического портрета остаются ниже линии среза $A - A$. Различаются части C_1 и C_2 тоже шириной среза. Для C_1 она равна 13, а для C_2 — 7. Акустическим эквивалентом этих частей в слове «один» являются участки слова, соответствующие согласным звукам «д» и «н».

Таким образом, осуществляя на определенном уровне срез в динамическом портрете слова, можно выделить его характерные части, т. е. осуществить сегментацию речевого потока на акустически однородные области.

Кроме того, части V_1 и V_2 несут информацию о том, существенно или несущественно возвышается над линией среза часть динамического портрета. Эта информация может быть использована при характеристике ударного слога в слове.

Части C_1 и C_2 позволяют, в свою очередь, получить информацию о том, соответствует ли часть динамического портрета слова под линией среза вокализованному или невокализованному звуку, если оценить, «сильно» или «слабо» заполнена строками часть динамического портрета под линией среза.

Выделение характерных частей в речевом сигнале по его динамическому портрету путем проведения среза осуществляется с использованием принципа преобладания. Срез $A-A$ (рисунок, позиция a) оставляет строку № 36 динамического портрета ниже среза, что привело бы к появлению области V_3 (строки № 36—40). Однако, удовлетворяя принцип по преобладанию, мы имеем одну область V_2 , что согласуется с фонетическим составом слова «один».

Полезным оказывается проведение дополнительных срезов в динамическом портрете слова. Например, срез $B-B$ (рисунок, позиция a) позволяет выявить отрезок речевого сигнала, соответствующий собственно смычке взрывного согласного «д». Это строки № 13—19. Переходные области гласных звуков: № 11, 12, 20—23, 42—43.

Таким образом, анализ динамических портретов отрезков речевых сигналов (слов) методом сечений позволяет, во-первых, выявлять акустически однородные области в речевом сигнале, во-вторых, путем проведения дополнительных срезов, выделять переходные области гласных звуков.

Список литературы: 1. *Эффект сглаживания в слухе* / О. М. Абрамов, А. Я. Дрюченко, С. А. Усенко, Ю. П. Шабанов-Кушнаренко. — Пробл. бионики, 1977, вып. 19, с. 31—37. 2. *Сапожков М. А., Михайлов В. Г.* Вокодерная связь. — М.: Радио и связь, 1983. — 248 с. 3. *Русская грамматика* / Под ред. Н. Ю. Шведовой. — М.: Наука, 1982. — 783 с. 4. *Матусевич М. И.* Современный русский язык. Фонетика. — М.: Просвещение, 1976. — 288 с. 5. *Елкина В. Н., Юдина Л. С.* Статистика открытых слогов русской речи. — Вычисл. системы, 1964, вып. 14, с. 55—91. 6. *Дрюченко А. Я.* Некоторые вопросы автоматического распознавания речи. — Пробл. бионики, 1981, вып. 26, с. 54—60.

Поступила в редколлегию 06.04.84.