

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

# АВТОМАТИЧЕСКОЕ ВЫДЕЛЕНИЕ ЩЕЛЕВЫХ СОГЛАСНЫХ В РЕЧЕВОМ СИГНАЛЕ

Данная работа является продолжением исследований речевых сигналов с использованием их «динамических портретов», начатых в [1] с целью построения алгоритмов автоматического выделения щелевых согласных в потоке речи.

Согласно [2] к щелевым согласным русского языка отнесем следующие фонемы: [ф], [ф'], [в], [в'], [с], [с'], [з], [з'], [ш], [ш':], [ж], [ж':], [х], [х']. То обстоятельство, что фонема является щелевой согласной, на языке алгебры конечных предикатов [3] может быть записано в виде следующего индивидуального предиката:

$$\begin{aligned} \text{Щел}(x) = & x^{\text{ф}} \vee x^{\text{ф}'} \vee x^{\text{в}} \vee x^{\text{в}'} \vee x^{\text{с}} \vee x^{\text{с}'} \vee x^{\text{з}} \vee x^{\text{з}'} \vee \\ & \vee x^{\text{ш}} \vee x^{\text{ш}'} \vee x^{\text{ж}} \vee x^{\text{ж}'} \vee x^{\text{х}} \vee x^{\text{х}'} \end{aligned}$$

Характерной особенностью щелевых согласных является наличие шума при их артикуляции [4], а кривые, характеризующие акустическое колебание, отличаются большим по сравнению со взрывными числом переходов через ноль и сравнимым в случае скользящих согласных [т'], [д'], [ц], [ч]. Кроме того, анализ элементарных сегментов гласных и сонорных звуков русского языка дает основание говорить о том, что элементарные сегменты, соответствующие фазе выдержки при артикуляции вокализованных звуков, отличаются достаточно высокой стабильностью. Этого нельзя сказать об элементарных сегментах, соответствующих таким фазам артикуляции как экскурсия и рекурсия. При этом кривая, характеризующая акустическое колебание, в пределах элементарного сегмента начинает деформироваться за счет конечной его части, примыкающей к следующему элементарному сегменту (участок АВ на рис. 1). Для конечной части элементарного сегмента характерны небольшие амплитудные отклонения. Поэтому при деформации конечной части сегмента часто возникает ситуация, когда переход сигнала из положительной области в отрицательную или наоборот осуществляется не через один ноль, а через целую их «пачку». Аналогичная ситуация наблюдается при анализе параметрического кода речевого сигнала, соответ-

ствующего фазе выдержки смычных согласных, т. е. паузы, где также встречаются «пачки» нолей. Для щелевых согласных характерно отсутствие «пачек» нолей, и переход сигнала из положительной области в отрицательную или наоборот преимущественно происходит «чисто», т. е. «без ноля».

Поэтому было предложено для характеристики щелевых согласных ввести в рассмотрение кривую, характеризующую число переходов через ноль на интервале определенной фиксированной длины. При подсчете числа переходов через ноль учитывались только «чистые» переходы, или переходы «без ноля». В дальнейшем описанную выше кривую будем называть кривой числа переходов через ноль.

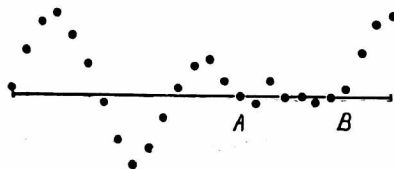


Рис. 1. Графическое представление элементарного сегмента акустической волны

На рис. 2, а — д показаны динамические портреты отрезков речевых сигналов, соответствующих словам «ужасный», «шкафу», «завял», «ахэмз», «рахит». Кривая числа переходов через ноль отмечена на рисунке символом «о».

Всего было проанализировано 280 слов трех дикторов, содержащих твердые и мягкие варианты щелевых согласных: по одному набору слов для двух дикторов и по два набора — для одного. При этом наборы слов были составлены так, что щелевой согласный находился в сочетании с любым последующим гласным, т. е. рассматривались слоги типа гсА, гсО, гсУ, гсЭ, гсЫ, гсЯ, гсЕ, гсЮ, гсЕ, гсИ.

Представленные слова содержат интересные нас щелевые согласные: губно-зубные — глухой [ф] и звонкий [в'], однофокусные — глухой [с] и звонкий [з], двухфокусные — глухой [ш] и звонкий [ж], заднеязычный [х]. Нетрудно заметить, что кривая числа переходов через ноль достаточно рельефно обозначает на динамических портретах слов сегменты щелевых согласных. Исключение составляет звонкий губно-зубной [в'] (рис. 2, в). Поэтому вопрос об автоматическом выделении в речевом сигнале сегмента, соответствующего губно-зубному звонкому щелевому с использованием кривой числа переходов через ноль не может быть решен — необходимо привлечение дополнительной акустической информации, почерпнутой из акустического анализа сигнала.

Следует упомянуть также о том, что поскольку шумовая составляющая у глухой губно-зубной [ф] по своей частоте распределена на достаточно широком спектральном диапазоне порядка 4 кГц и более [4], то «колпак» PQR (рис. 2, а) может быть высоким — сегмент, соответствующий [з] (рис. 2, г), средним — сегмент глухого [ф] (рис. 2, б) и низким, как в случае звонкого [в] (рис. 2, в). Так, из двадцати динамических

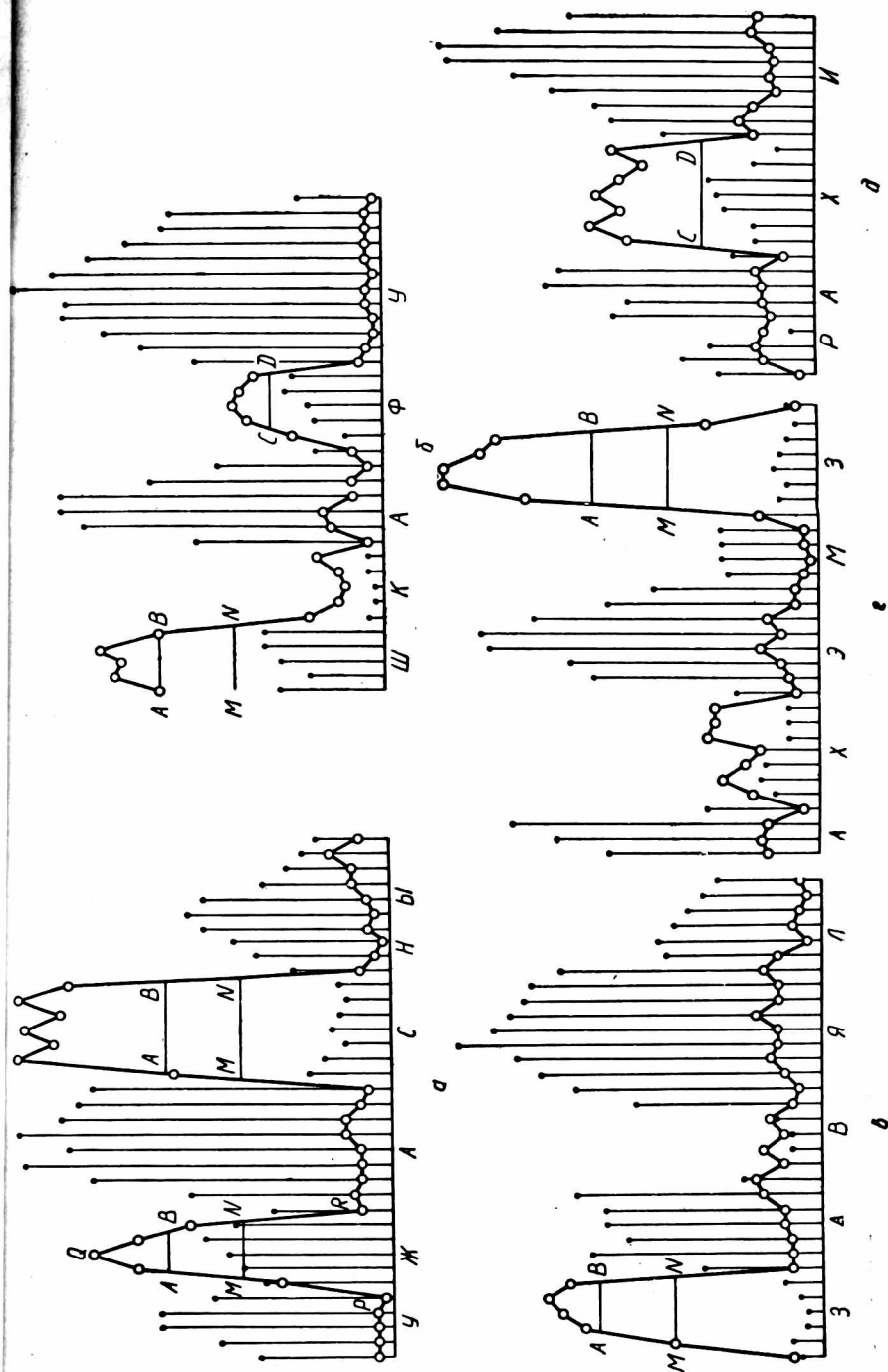


Рис. 2. Динамические портреты слов:  
а — «ужасный»; б — «шкафу»; в — «завял»; г — «ахэмз»; д — «рахит»

портретов слов (по десять слов двух дикторов), содержащих глухой щелевой согласный, на двух «колпак» оказался низким и на четырех — высоким. Это усложняет процедуру выделения в речевом сигнале сегмента, соответствующего глухим [ф], [ф'], но, как видим из приведенных цифр, не делает ее безнадёжной. Тем более, что наличие «высокого колпака» в сочетании с другими признаками делает глухой щелевой уникальным.

Твердый вариант заднеязычного щелевого согласного [х] (рис. 2, з) незначительно отличается поведением кривой числа переходов через ноль на анализируемом сегменте от губно-зубного звонкого [в]. В обоих случаях «колпак», накрывающий согласный, либо отсутствует, либо является низким. А поскольку согласные [в], [х] характеризуются еще и слабой интенсивностью акустических колебаний, то, согласно принятой методике анализа речевых сигналов [1], они попадут в одну группу вместе со взрывными согласными. Поэтому для отделения губно-зубных [в], [в'] от взрывных согласных используется такой признак, как длительность согласного (точнее ширина области атенсивности, соответствующая анализируемому согласному). Для отделения заднеязычного твердого [х] от взрывных согласных используется такой признак, как среднее значение интенсивности сигнала на анализируемом сегменте. Для заднеязычного она оказывается выше, чем для взрывных, фаза выдержки которых соответствует акустической паузе.

Мягкий вариант заднеязычного щелевого [х'] (рис. 2, д) характеризуется наличием «среднего колпака», накрывающего сегмент согласного, т. е. мягкий [х'] по числу переходов через ноль сравним с глухим губно-зубным твердым [ф] и мягким [ф'].

Попутно заметим, что на динамических спектрограммах открытых слогов мягкий [х'] в слове ихИ «обладает столь мощным фрикативным высокочастотным шумом, что этот слог почти невозможно отличить от слога исИ» [4], т. е. с позиций спектрального анализа трудно отделить [х'] от [с'] и от [с]. В нашем случае (сравним сегменты интересующих согласных на рис. 2, а и д) удастся провести надежную границу, отделяющую [х'] от [с].

Чтобы выделить в речевом сигнале сегменты, соответствующие однофокусным и двухфокусным щелевым [с], [з], [ш], [ж], воспользуемся методом сечений [1]. Проведем срез кривой числа переходов через ноль на уровне 60 % от нормализованного по амплитуде сигнала (отрезки АВ на рис. 2). Поступая так, мы обязательно получим срез «высоких колпаков» и не заденем «колпаки» средней высоты. Для выделения в речевом сигнале сегментов, соответствующих глухим губно-зубным [ф], [ф'] и мягкому заднеязычному [х'], проведем срез кривой числа переходов

через ноль на уровне 30 % от нормализованного по амплитуде речевого сигнала. В этом случае мы обязательно сделаем срез «колпаков» средней высоты (отрезки  $CD$  на рис. 2, б и д). Конечно, при этом будут сделаны срезы и «высоких колпаков», если не сделать эти срезы последовательно и не исключить из рассмотрения после среза «высоких колпаков» выявленные сегменты.

Уровень срезов кривой числа переходов через ноль в 60 и 30 % выбран из тех соображений, что на имеющемся статистическом материале событие «есть срез» и «есть срез вниз и нет среза вверх» имеют вероятность, близкую к единице.

Для уточнения границ сегментов поступим следующим образом. Получив срезы  $AB$ ,  $CD$ , будем «спускаться» по левой и правой сторонам «колпака» до тех пор, пока не встретим строку динамического портрета. Левая и правая точки встречи дадут нам уточненные границы сегментов. Эта процедура может быть легко реализована программно, если вспомнить, что и динамический портрет, и кривая числа переходов через ноль — это массивы в памяти ЭВМ.

Для того чтобы в классе звуков  $[c]$ ,  $[c']$ ,  $[z]$ ,  $[z']$ ,  $[ш]$ ,  $[ш']$ ,  $[ж]$ ,  $[ж']$  однофокусные  $[c]$ ,  $[c']$ ,  $[z]$ ,  $[z']$  отделить от двухфокусных  $[ш]$ ,  $[ш']$ ,  $[ж]$ ,  $[ж']$ , нужно оценить среднее значение интенсивности сигнала на сегменте согласного. Двухфокусные являются более энергоемкими согласными по сравнению с однофокусными. Это можно наблюдать и в степени затемнения динамических спектрограмм [4]. Поэтому среднее значение интенсивности сигнала на анализируемом сегменте будет больше для двухфокусных щелевых  $[ш]$ ,  $[ш']$ ,  $[ж]$ ,  $[ж']$ .

В качестве признака, отделяющего звонкие от глухих как однофокусные щелевые, так и двухфокусные, примем длину отрезка  $MN$ .  $MN$  — это срез «высокого колпака» на уровне 40 % от нормализованного по амплитуде речевого сигнала. Для глухих щелевых длина отрезка  $MN$  больше, чем для звонких. Уровень в 40 % выбран потому, что для всех рассмотренных выборок отделение звонких от глухих наиболее надежно.

Особое положение занимает мягкий глухой двухфокусный  $[ш']$ . Для него длина отрезка  $MN$  оказывается существенно больше, чем для твердого  $[ш]$ . Это обстоятельство позволяет отделить мягкий глухой двухфокусный от твердого.

Итак, такие признаки как наличие или отсутствие срезов  $AB$ ,  $CD$ , среднее значение интенсивности сигнала на анализируемом сегменте, длина отрезка  $MN$  позволяют получить следующие классы для щелевых согласных: глухие однофокусные  $[c]$ ,  $[c']$ ; звонкие однофокусные  $[z]$ ,  $[z']$ ; глухой твердый двухфокусный  $[ш]$ ; глухой мягкий двухфокусный  $[ш']$ ; звонкие двухфокусные  $[ж]$ ,  $[ж']$ ; заднеязычный твердый  $[x]$ ; губно-зубные звонкие  $[в]$ ,  $[в']$ ; губно-зубные глухие  $[ф]$ ,  $[ф']$  и заднеязычный мягкий  $[х']$ .

В заключение приведем основные этапы алгоритма автоматического выделения щелевых согласных в потоке речи.

1. Согласно методике, изложенной в [1], получаем динамический портрет речевого сигнала. 2. По динамическому портрету выделяем в речевом сигнале сегменты, характеризующиеся большой интенсивностью, средней и малой. 3. Определяем длину сегмента с малой интенсивностью и отделяем [в], [в'] от взрывных согласных. 4. Определяем среднее значение интенсивности сигнала на сегменте с малой интенсивностью и отделяем [х] от взрывных согласных. 5. Подсчитываем число переходов через ноль на интервалах 10 мС. 6. Проводим срез АВ кривой числа переходов через ноль и выделяем в речевом сигнале сегменты, соответствующие согласным [с], [с'], [з], [з'], [ш], [ш':], [ж], [ж':]. 7. Уточняем границы сегментов. 8. Для полученных сегментов подсчитываем среднее значение интенсивности сигнала и отделяем [с], [с'], [з], [з'] от [ш], [ш':], [ж], [ж':]. 9. Для полученных сегментов проводим срез MN кривой числа переходов через ноль и отделяем [с], [с'] от [з], [з']; [ш], [ш':] от [ж], [ж':] и [ш':] от [ш].

**Список литературы:** 1. Бондаренко М. Ф., Николаенко В. Л. Методика анализа речевых сигналов по их динамическим портретам.— Пробл. бионики, 1985, вып. 35, с. 12—21. 2. Черкасов Л. Н. О классификации согласных современного русского литературного языка по способу образования.— Вopr. языкознания, 1980, № 3, с. 36—45. 3. Шабанов-Кушнарeнко Ю. П. Теория интеллекта: Математические средства.— Х.: Вища шк.. Изд-во при Харьк. ун-те, 1984.— 144 с. 4. Динамические спектры речевых сигналов / М. Ф. Деркач, Р. Я. Гумецкий, Б. М. Гура, М. Е. Чабан.— Львов: Вища шк. Изд-во при Львов ун-те, 1983.— 168 с.

*Поступила в редколлегию 28.11.84.*