

Рис. 2.

Л.К. Бабенко, Д.С. Субботин, В.М. Федоров, П.Ю. Юрков
Россия, г. Таганрог, ТРТУ

ОПРЕДЕЛЕНИЕ ГРАНИЦ МЕЖДУ ФОНЕМАМИ НЕЙРОСЕТЕВЫМ МЕТОДОМ

Разработка методов автоматического распознавания речи в настоящее время осуществляется посредством анализа частотного представления речевого сигнала. Основу процедур распознавания при этом составляет критериальное сравнение образа некоторого отрезка указанного представления речевого сигнала с эталонами в выбранном пространстве признаков. Как следует из [1], в качестве эталонов выбирают, как правило, одну или несколько фонологических единиц, множество которых составляют аллофоны, фонемы, дифоны, слоги, слова. В дальнейшем на основе распознавания указанных фонологических единиц осуществляется их интерпретация буквами языка, на котором происходит передача речевых сообщений, согласующаяся со словарно-грамматическими правилами, введенными в систему. Результатом распознавания являются графические эквиваленты фонологически определенных фрагментов исходного речевого сигнала как единиц языка.

Одну из ключевых задач при создании систем автоматического распознавания речи составляет разработка методов сегментации исходного речевого сигнала на отрезки, которые представляют образы тех фонологических единиц, эталоны которых сформированы в выбранном пространстве признаков. Указанную задачу принято называть проблемой сегментации. Как следует из вышеизложенного, в качестве решения представленной задачи предусматривается получение адекватной сегментации исходного речевого сигнала. То есть полученные в результате сегментации отрезки должны соответствовать концептам выбранных фонологических единиц, а, следовательно, быть денотатно-завершенными.

Отличие множества денотатов, задающих множество ключевых объектов **B**, от множества денотатов, связанных с фонологическими единицами, не составляет принципиальной разницы соответствующих задач. Действительно, на основе изучения фонетики языка, на котором происходит передача информации речевым сообщением, всегда возможно денотаты из **B** определить как фонологические объекты. Таким образом, понятия адекватности сегментации исходного речевого сигнала в рамках рассматриваемых задач будем считать идентичными. Доминирующей идеологией решения проблемы сегментации была и остается попытка разделения звуковых волн на отрезки, соответствующие определенным лингвистическим фонологическим единицам. Поставленные во главу угла фонологические единицы предопределили тенденцию описания их множествами признаков с обязательным установлением числовых пределов изменения последних, полученных при проведении субъективно-статистических исследований. В соответствии со

сказанным ранее, алгоритмизация расчетов указанных групп признаков на протяжении речевого сигнала должна привести к выявлению таких отрезков, которые представляют образы фонологических единиц, а следовательно, и к решению задачи автоматической сегментации речевых сообщений. Необходимо отметить, что большинство исследователей считает задачу сегментации невыполнимой [2], так как фонемы не имеют явных физических границ.

В данной работе проблема сегментации речи решается путем обучения нейронной сети на выделенные признаки границ фонологических единиц, а именно фонем, с последующим их распознаванием в речевом сигнале. Для выделения признаков границ фонем речевой сигнал представляется в частотной области с помощью Фурье [3] и дискретного вейвлет-преобразования [4].

Исходный речевой сигнал x_i , $i=1:256$ разбивается на перекрывающиеся фрагменты с шагом $S=128$, т. е. формируется множество векторов $v = \{v_k\}_{k=0}^M$, где $M = \frac{N-256}{S}$.

Для каждого вектора вычисляются спектральные коэффициенты дискретного преобразования Фурье, которые могут быть определены по следующей формуле:

$$X_{km} = \frac{1}{256} \sum_{l=0}^{255} x_{kS+l} e^{-j \frac{2\pi lm}{256}}.$$

Затем формируется множество векторов, представляющих собой мгновенные логарифмические спектральные срезы исходного речевого сообщения:

$$\mathcal{G} = \{\mathcal{G}_k\}_{k=0}^M, \text{ где } \mathcal{G}_k = (\log |X_{k0}|, \dots, \log |X_{k255}|).$$

Далее в полученном множестве векторов выделяются смежные группы, состоящие из D векторов, для каждой из которых определяется среднее значение логарифмического спектра мощности в соответствии с формулой

$$\tilde{\mathcal{G}}_h = \frac{1}{D} \sum_{q=0}^{D-1} \mathcal{G}_{h+q} \text{ где } h=0:M-D-1.$$

Затем вводится функция однородности речевого сообщения, определяемая соотношением

$$\mathfrak{R}_g = e^{-\|\tilde{\mathcal{G}}_g - \tilde{\mathcal{G}}_{g+1}\|}, \text{ где } g=0:M-D-2.$$

При разложении речевого сигнала с помощью дискретного вейвлет-преобразования используется вейвлет Дебеша. Разложение является многоуровневым, причем на каждом уровне вычисляется энтропия Шенона H^k уровня k :

$$H^k = - \sum_{i=0}^{255} (s_i^k)^2 \log(s_i^k)^2.$$

Здесь s_i^k - значения речевого сигнала на уровне k дискретного разложения исходного сигнала, $k=1 \dots N$. В данной работе $N=6$.

Сформированный таким образом вектор использовался для обучения искусственной нейронной сети, в качестве которой применялся трехслойный персептрон с десятью нейронами в скрытом слое и одним выходным нейроном.

После обучения на сеть подавался речевой сигнал другого диктора с произвольным словом. На рис.1 приведены результаты распознавания границ между фонемами слова «один».

Как видно из рисунка, сеть хорошо определила переход от фонемы «а» к фонеме «д», от фонемы «д» к «и» и от «и» к «н». К недостаткам надо отнести пропуск границ фонем в некоторых словах. Так, в слове «два» наблюдается пропуск границы между фонемами «д» и «в». Это связано с коротким участком переход-

ных процессов между фонемами для данного диктора, в случае других дикторов граница перехода определяется достаточно четко. В процессе дальнейшего распознавания фонем и восстановления слова объединенные фонемы будут восстановлены методом поиска по словарю.

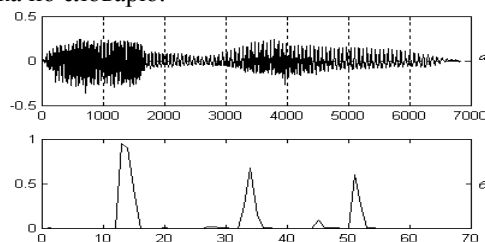


Рис1.

a- график речевой сигнал слова «один», *б*- распознанные границы фонем.

Библиографический список

1. Методы автоматического распознавания речи /Под. ред. У.Ли. М.: Мир, 1983. Т.1. 326 с.
2. Косарев Ю.А. Естественная форма диалога с ЭВМ. М.: Машиностроение, Ленингр. отд., 1989.

А.М. Косарев

Россия, г. Москва, 16 ЦНИИ МО РФ

РАСЧЕТ РАЗБОРЧИВОСТИ РЕЧИ В СОСТАВНЫХ ТРАКТАХ СВЯЗИ С НЕСКОЛЬКИМИ ТИПАМИ РЕЧЕПРЕОБРАЗУЮЩИХ УСТРОЙСТВ ПРИ ВОЗДЕЙСТВИИ ШУМОВ И ПОМЕХ

Предложена методика расчета разборчивости речи в составном тракте связи при низкочастотном сопряжении нескольких речепреобразующих устройств различных типов и воздействии канальных помех и акустических шумов.

The procedure of the speech intelligibility design for the low frequency interface of some different speech forming devices and the effect of the channel noise and acoustic noise is proposed.

В практике организации связи часто возникает потребность сопряжения нескольких телефонных сетей, использующих оборудование на основе различных методов аналого-цифрового преобразования речи. В основе расчета разборчивости речи по каналам связи лежат известные методы артикуляционных испытаний и расчетов разборчивости формантным способом. В предлагаемой методике расчет разборчивости проводится с помощью контроля по участкам приемного тракта соотношений сигнал/помеха. Методика позволяет оценить расчётным путем величину разборчивости речи в тракте телефонной связи при различных типах речепреобразующих устройств (РПУ), в том числе с учётом уровней и спектров акустических шумов в месте передачи и приёма, а также числа ошибок в цифровом канале связи.

Структурная схема тракта рассматриваемого типа состоит из n участков:

первый участок включает в себя абонентское телефонное устройство (АТУ), содержащее звукоприемник и ветвь передачи телефонного аппарата, соединительную линию, причем речевой сигнал в месте передачи подвергается воздействию акустического шума с уровнем интенсивности B_n ;

на втором и до $n-1$ участке тракта включены РПУ любого типа и цифровой канал связи, причём на линейный сигнал воздействует помеха, характеризующая вероятностью ошибки $p_{owl, \dots, n-1}$, которая зависит от соотношения сигнал/помеха;

последний участок тракта содержит соединительную линию, приемную ветвь