

СПОСОБ АНАЛИЗА СЕГМЕНТОВ РЕЧЕВЫХ СИГНАЛОВ

Цупрев Н.И., Литвинцова Ю.Н., Литвинцов Д.А.

Россия, Москва, МГТУ им. Н.Э.Баумана, кафедра ИУ-1.

E-mail: pupkov@iu1.bmstu.ru

Проблема анализа и распознавания речи как часть общей задачи ввода информации в ЭВМ на естественном языке, является одной из приоритетных проблем, связанных с созданием систем с искусственным интеллектом.

В настоящее время наиболее качественное распознавание речи обеспечивают устройства, работающие с ограниченным словарем, ориентированные на одного диктора.

В большинстве систем речевого общения анализ речевых сигналов с целью их описания осуществляется с помощью оценки фактических характеристик сегментов текущей реализации, расчет которых производится либо с помощью интегральных преобразований (Фурье, Лапласа, Гилберта), либо итерационными методами, либо на основе корреляционных функций. Для устранения временной вариативности речевой реализации применяются методы динамического программирования [1].

При этом используются как детерминированные, так и вероятностные характеристики анализируемых сигналов.

Однако ни один из методов анализа речевых сигналов не предлагает способа однозначного и конечного кодового описания таких элементов речи как, например, фонема.

Сигнал, поступающий на вход системы анализа, подвергается различным методам обработки с целью выявления тех или иных его свойств.

До настоящего времени, выделенные характеристики использовались в качестве признаков и сравнивались с заранее сформированным «усредненным» описанием.

Предлагается СПОСОБ АНАЛИЗА сигнала путем сравнения его характеристик до и после некоторого преобразования, например, интегрирования, дифференцирования, усреднения (метод анализа путем оценки реакции характеристик сигнала на воздействие) на переменном интервале.

Рассмотрим предложенную методику на примере слова «САМ».

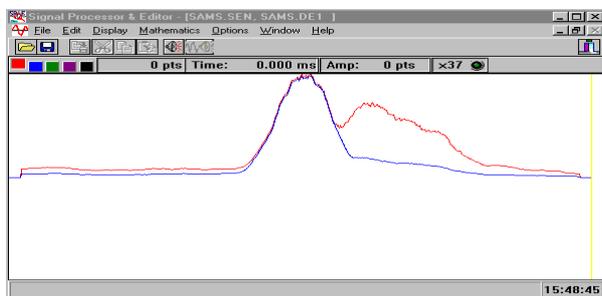


Рис.1. Сравнение энергий исходного сигнала и после дифференцирования.

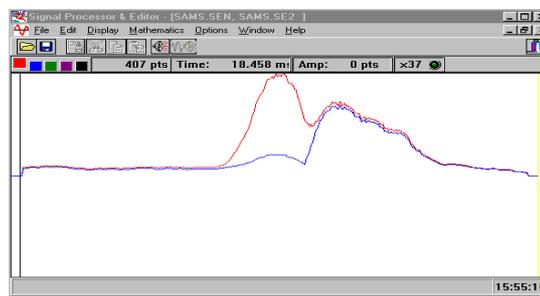


Рис.2. Сравнение энергий исходного сигнала и после усреднения по 2 точкам.

На рис.1 представлено сравнение энергий исходного сигнала и сигнала после дифференцирования, на рис.2– сравнение энергий исходного сигнала и после усреднения по 2 точкам.

Физическое явление при рассмотрении с позиций теории случайных процессов можно описать в любой момент времени путем усреднения величин по множеству выборочных функций (для сигнала – выборочных значений на интервале), представляющих данный случайный процесс (сигнал) [2].

Среднее значение (первый момент распределения) случайного сигнала $X(t)$ на интервале вычисляется по известной формуле:

$$\mu_x(t_j) = \frac{1}{T} \cdot \int_0^T X(t) dt$$

В большинстве случаев, в частности для оценки речевых сигналов, используют момент второго порядка – среднее значение квадрата (осредненную энергию) [2]:

$$\Psi_X^2 = \int_0^T X^2(t) dt.$$

Для стационарных случайных процессов (сигналов) эти моменты постоянны и не изменяются во времени.

Существует большой класс физических явлений, в том числе и устная речь, относящихся к разряду нестационарных процессов, где эти принципы нарушаются.

Естественно предположить, что анализ нестационарных сигналов необходимо проводить на интервале переменной длительности, т.е. размер выборки должен варьироваться.

Вычисленные характеристики на разных интервалах сравниваются между собой и по характеру их соотношения определяются стационарные и переходные участки.

На рис.3 представлен случайный сигнал и его энергетические характеристики на разных интервалах анализа.

Точки пересечения энергетических кривых определяют моменты перестройки источника сигнала – переходные процессы, соотношение энергий в интервалах между пересечениями – характер сгенерированного процесса, длительность участка – скорость протекания.

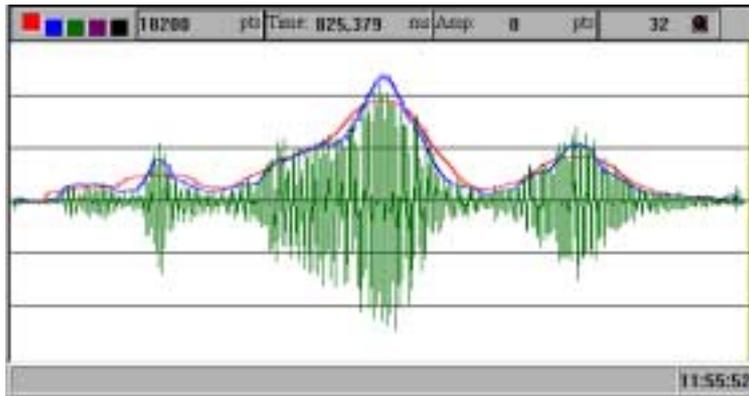


Рис.3.

Анализ представленных результатов по оценке реакции на аналитическое воздействие (дифференцирование и усреднение по изменяемому интервалу) позволяет сделать следующие выводы:

- 1) щелевые (рис.1) не реагируют на дифференцирование при частоте дискретизации 22050 Гц.
- 2) на усреднение (рис.2) по минимальному интервалу (два отсчета) – не реагируют гласные и сонанты.

Таким образом, в зависимости от характера генерируемый сигнал реагирует на преобразования строго определенным образом.

Обработка сигналов на разных интервалах анализа позволяет решить несколько проблем:

- 1) сегментация сигнала на интервалы анализа;
- 2) для многих процессов, где в качестве случайного параметра выступает время реализации – временная нормализации;
- 3) выявление быстро протекающих процессов.

ЛИТЕРАТУРА

1. Г.Корн, Т.Корн. Справочник по математике. – М.: Наука. 1984. 834 с. (С. 696).
2. Дж.Бендад, А.Пирсол. Измерение и анализ случайных процессов. – М.: Мир, 1971. – 408с.



WAY OF THE ANALYSIS OF SEGMENTS OF SPEECH SIGNALS

Tsuprev N.I., Litvintsova Ju.N., Litvintsov D.A.

Russia, Moscow, Bauman States Technical University, faculty IU-1.

E-mail: pupkov@iu1.bmstu.ru

The problem of the analysis and recognition of speech as a part of the common task of input of the information in the COMPUTER on the natural language, is one from priority problems connected to creation of systems to an artificial intelligence.

The now most qualitative recognition of speech is ensured with devices working with the limited dictionary, oriented on one speaker.

In majority of systems of speech dialogue the analysis of speech signals with the purpose of their description is carried out with the help of evaluations of the actual characteristics of segments of current implementation, which account is made or with the help of of integrated conversions (Fourier, Laplace, Gilbert), or iterative methods, or because of of correlation functions. The methods of dynamic programming are applied to elimination temporary variable of speech implementation [1].

For want of it are used both determined, and probability characteristics of analyzable signals.

However any from methods of the analysis of speech signals does not offer a way of the unequivocal (unambiguous) and final code description of such units of speech as, for example, phoneme.

Signal, acting (arriving) on an input (entrance) of the system of the analysis, is subjected to various methods of processing with the purpose of revealing those or others it of properties.

Till now, the selected (allocated) characteristics were used character and were compared to the beforehand generated "«average" description.

the WAY of the ANALYSIS of a signal is offered by matching it of the characteristics before some conversion, for example, integration, derivation, averaging (method of the analysis by an evaluation of response of the characteristics of a signal on effect) on a variable interval.

The physical phenomenon for want of consideration from positions of the theory of casual processes can be described at any moment of time by average of values on set of sample functions (for a signal - sample values on an interval), representing the given casual process (signal) [2].

Average value (first moment of allocation) casual signal $X(t)$ on an interval is calculated under the known formula:

$$\mu_x(t_f) = \frac{1}{T} \cdot \int_0^T X(t) dt$$

In most cases, in particular (personally) for an evaluation of speech signals, I use a moment of the second order - average value of quadrate (average energy) [2]:

$$\Psi_X^2 = \int_0^T X^2(t) dt.$$

For stationary casual processes (signals) these moment are fixed and do not vary in time.

There is a large class of the physical phenomena, including oral speech relating bit of non-stationary processes, where these principles are infringed.

It is natural to assume, that the analysis of non-stationary signals is necessary for conducting on an interval of variable duration, i.e. the size of sample should vary.

The calculated characteristics on different intervals are compared among themselves and on a character of their parity (ratio) the stationary and transition sites are determined.

The analysis of represented outcomes on an evaluation of response on analytical effect (derivation and averaging on a changed interval) allows to make the following outputs:

- 1) slot-hole (slotted) do not react to derivation for want of to sampling rate of 22050 Hz.
- 2) two counting) - do not react to averaging on a minimum interval (publics and sonant).

Thus, depending on a character the generated signal reacts to conversions strictly definitely.

The processing of signals on different intervals of the analysis allows to decide some problems:

- 1) segmentation of a signal on intervals of the analysis;
- 2) for many processes, where as the casual parameter the time of implementation - temporary normalization appears;
- 3) revealing of fast flowing past processes.