УДК 651.926:681.3
ББК 32.96
Ш 96

К.С. Шурыгин

Разработка алгоритма фонетического анализа речи на основе информационной теории восприятия речи

(Рецензирована)

Аннотация

Рассмотрена задача фонетического анализа речи на основе информационной теории восприятия речи. Предложен новый алгоритм классификации. Приведены результаты практического исследования оптимальных параметров.

***Ключевые слова**: автоматическая обработка речи, информационная теория восприятия речи, фонема, критерий минимума информационного рассогласования.*

K.S. Shurygin

Development of an algorithm of speech phonetic analysis on the basis of the information theory of speech perception

Abstract

The paper examines the problem of speech phonetic analysis on the basis of the information theory of speech perception. The author suggests a new algorithm of classification and gives results of practical research of optimum parameters.

***Key words**: automatic speech processing, informational theory of speech perception, phoneme, criterion for a minimal information mismatch.*

Введение

В области автоматической обработки речи перспективной является информационная теория восприятия речи (ИТВР). В статье [1] даются ее базовые понятия. В соответствии с ИТВР все множество элементарных речевых единиц (ЭРЕ) в сознании человека разбивается на конечное число подмножеств - кластеров. Каждое такое подмножество имеет четко очерченные границы вокруг некоторого центра – эталона, который определяется по аналогии с центром масс, но в метрике Кульбака-Лейблера [2]. Такой кластер является фонемой. Причем, чем больше элементов включает в себя фонема, тем устойчивее и, следовательно, точнее определяется ее центр-эталон. Таким образом, для обработки слитной речи необходимо сформировать классифицированный набор реализаций ЭРЕ. Разработке алгоритма фонетического анализа на основе ИТВР посвящена данная работа.

Критерий минимума информационного рассогласования

Задача распознавания образов при статистическом подходе сводится к проверке R гипотез о законе распределения выборки [3]:

$$W_r: \quad \mathbf{P}_X = \mathbf{P}_r, \qquad r = \overline{1,R}. \tag{1}$$

Здесь $\mathbf{P}_r$ – r-я альтернатива распределения, предполагаемая точно заданной; при этом множество альтернатив $\{\mathbf{P}_r\}$ исчерпывает собой все допустимое многообразие законов распределения выборки X.

В работе [4] было показано, что выражение для оптимальной решающей статистики при применении строгого критерия минимума информационного рассогласования (МИР) и при гауссовом распределении сигнала $\mathbf{P}(X_r)$ сводится к виду

$$\rho_{x,r} \stackrel{\Delta}{=} \frac{1}{F}\sum_{f=1}^{F}\left(\frac{G_x(f)}{G_r(f)} + \ln\frac{G_r(f)}{G_x(f)}\right) - 1 \to \min_{r=1,R}. \quad (2)$$

Здесь: $G_x(f)$ – выборочная оценка спектральной плотности мощности (СПМ) анализируемого сигнала X в функции дискретной частоты f; $G_r(f)$ – СПМ r-го сигнала из словаря эталонов; F – верхняя граница частотного диапазона сигнала или используемого канала связи; R – размер или объем рабочего словаря. Если выполнить нормировку коэффициентов линейного предсказания сигнала по дисперсии их порождающего шума, то получим из выражения (1), стандартную формулировку выборочной оценки величины информационного рассогласования (ВИР) между сигналом X на входе и r-м сигналом из словаря в частотной области [5]:

$$\rho_{x,r} = \frac{1}{F}\sum_{f=1}^{F}\frac{\left|1+\sum_{m=1}^{p} a_r(m)\exp\left(-\frac{j\pi mf}{F}\right)\right|^2}{\left|1+\sum_{m=1}^{p} a_x(m)\exp\left(-\frac{j\pi mf}{F}\right)\right|^2} - 1, \quad (3)$$

где $T = 1/(2F)$ – период дискретизации речевого сигнала.

Синтез алгоритма

Все множество альтернативных распределений $\{\mathbf{P}_r\}$ разобьем на R^2 всевозможных пар $(\mathbf{P}_i, \mathbf{P}_j), i, j \le R$. Затем вычислим для каждой такой пары удельную величину их взаимного информационного рассогласования (ИР) [6]:

$$\rho(\mathbf{P}_i / \mathbf{P}_j) = \rho_j(X_i) = \sigma_j^2(X_i)/\sigma_j^2 + \ln(\sigma_j^2/\sigma_j^2(X_i)) - 1, \quad (4)$$

где X_i - n-выборка из i-ой генеральной совокупности $\mathbf{P}_i$.

Элементы, для которых выполняется правило

$$\rho_j(X_i) \le \rho_0 \quad (5)$$

при $j=1$, образуют первый кластер, ρ_0 - некоторый пороговый уровень (сверху) для минимальной решающей статистики из выражения (2). Если второй элемент не вошел в первый кластер, то строим второй кластер по правилу (5) для $j=2$. Если же вошел, то второй кластер строим по условию (5) для $j=3$. Группируем таким образом элементы множества X. При возникновении спорной ситуации, когда правило (5) выполняется для нескольких элементов, т.е. он попадает сразу в несколько кластеров, предпочтение отдается тому из них, для которого значение решающей статистики меньше. Таким образом, получаем набор речевых образов $X_r = \{\mathbf{x}_{r,j}\}$.

После кластеризации возникает следующий вопрос – что же брать в качестве образа каждой фонемы. В соответствии с ИТВР решать эту проблему можно с помощью метода минимума суммы информационных рассогласований [7]. Этот подход заключается в нахождении информационного центра по множеству различных реализаций одной фонемы.

Пусть каждый речевой образ $X_r = \{\mathbf{x}_{r,j}\}$ представлен конечным множеством объема V_r. Пользуясь выражением (4) можно получить матрицу ИР $V_r x V_r$ между элементами внутри фонемы. Находим сумму минимума информационного рассогласования для каждого элемента фонемы по правилу [1]

$$M_r = \sum_{i \ne j}^{V_r} \rho_j(X_i) \to \min_{i,j}. \quad (6)$$

После этого реализацию с минимальным значением суммарного информационного рассогласования $M_r^{\min}$ считаем эталонной, а ее параметры становятся образом всего кластера. Таким образом, находим центры-эталоны для всех фонем.

Пример. Для экспериментальных исследований синтезированного алгоритма был взят текст первой главы романа А.С. Пушкина "Капитанская дочка", проговорен и записан в память персонального компьютера в виде звукового файла. Для этого применялась специальные программные и аппаратные средства: динамический микрофон AKG D77 S и ламповый микрофонный предусилитель ART TUBE MP Project Series USB. Частота дискретизации встроенного аналого-цифрового преобразователя была установлена равной 8 кГц – общепринятая частота при обработке речи. Продолжительность записи составила около полутора минут. Далее по алгоритму (7)-(9) [8] были выделено множество ЭРЕ. По алгоритму (3)-(6) был проведен анализ данного множества при разных значениях ρ_0. На рисунке 1 показана зависимость количества выделенных фонем от порога ρ_0.

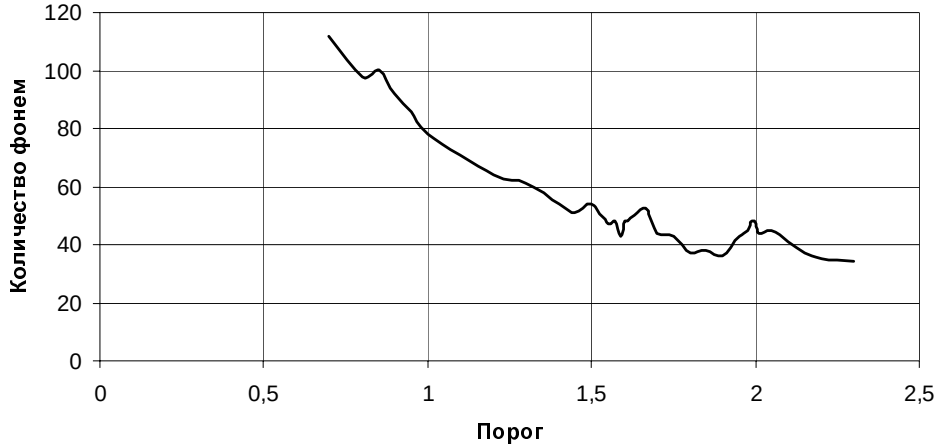

Рис. 1. График зависимости количества фонем от порога ρ_0

Проблему выбора оптимального значения можно решить по принципу относительной стабилизации фонетического состава речевого сигнала. С одной стороны, при малых значениях порогов мы получаем чрезмерно большое количество фонем, с небольшим различиями между собой, в теоретико-информационном смысле. С другой стороны, при слишком больших значениях порогов в один кластер, возможно, попадут реализации заведомо разных фонем. А это безусловная ошибка фонетического анализа. Следовательно, значения порога ρ_0 следует выбирать в точках на графике, где количество классифицированных фонем достаточно представительно. Это соответствует промежутку ρ_0=1,2..2. Для более точного выделения оптимальных значений порога построим график зависимости величины среднего информационного рассогласования между эталонами фонем от порога ρ_0. Данная зависимость представлена на рисунке 2.

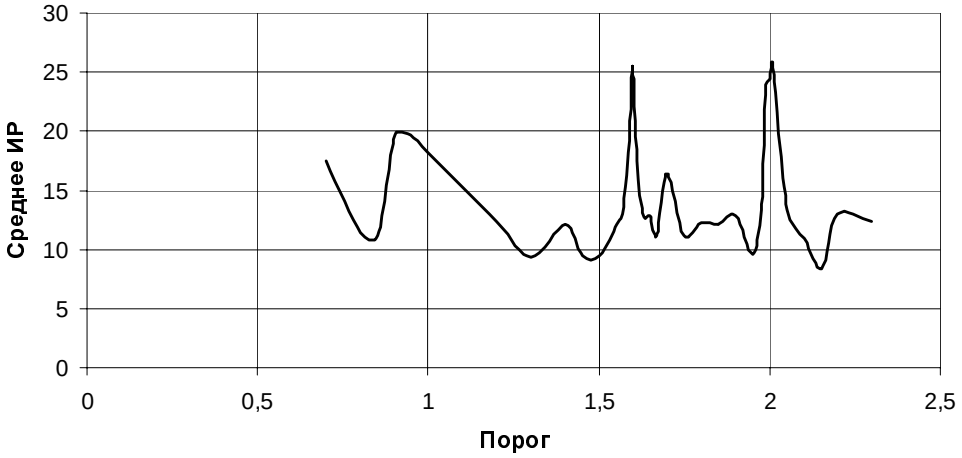

Рис. 2. График зависимости средней величины ИР от порога ρ_0

Из рисунков видно, что оптимальным значением порога ρ_0 являются значения равные 1,6 и 2,01. При этих значениях порога фонетическая база получается наиболее полная в теоретико-информационном смысле, как по количеству фонем, так и по наполнению базы.

Заключение

Таким образом, благодаря проведенному исследованию предложен новый алгоритм классификации в задаче формирования фонетической базы данных и проведено его экспериментальное исследование. Ключевым моментом алгоритма является нахождение информационного центра-эталона речевого образа, идея которого была предложена в [1]. В результате экспериментального исследования были определены значения оптимального порога.

Примечания:

1. Савченко В.В. Информационная теория восприятия речи // Известия вузов России. Радиоэлектроника. 2007. Вып. 6. С.3-9.
2. Кульбак С. Теория информации и статистика. М., 1967.
3. Савченко В.В. Автоматическая обработка речи по критерию минимума информационного рассогласования на основе метода обеляющего фильтра // Радиотехника и электроника. 2005. Т. 50, № 3. С. 309-315.
4. Савченко В.В. Различение случайных сигналов в частотной области // Радиотехника и электроника. 1997. Т.42, № 4. С. 426-431.
5. Савченко В.В., Акатьев Д.Ю. Автоматическое распознавание речи по критерию минимального информационного рассогласования с переспросом // Известия вузов России. Радиоэлектроника. 2006. Вып. 1. С. 20-29.
6. Савченко В.В. Автоматическое распознавание речи методом дерева на основе информационного $(R+1)$ -элемента // Известия вузов России. Радиоэлектроника. 2006. Вып. 4. С. 13-22.
7. Савченко В.В., Акатьев Д.Ю., Шерстнев С.Н. Метод оптимального обучающего словаря в задаче распознавания речевых сигналов по критерию минимального информационного рассогласования // Известия вузов. Радиоэлектроника. 2006. Вып. 5. С. 10-14.
8. Савченко, В.В., Акатьев, Д.Ю., Карпов, Н.В. Анализ фонетического состава речевых сигналов методом переопределенного дерева // Системы управления и информационные технологии. 2008. № 2 (32). С. 297-303.

References:

1. Savchenko V.V. Information theory of speech perception // News of Russian higher schools. Radio electronics. 2007. Issue 6. P 3-9.
2. Kulbak S. The theory of the information and statistics. M., 1967.
3. Savchenko V.V. Automatic speech processing by criterion for a minimum of an information mismatch on the basis of a method of the whitening filter // Radio Engineering and Electronics. 2005. V. 50, No. 3. P. 309-315.
4. Savchenko V.V. Distinction of casual signals in frequency area // Radio Engineering and Electronics. 1997. V. 42. No. 4. P. 426-431.
5. Savchenko V.V., Akatiev D.Yu. Automatic speech recognition by criterion for the minimal information mismatch with re-questioning // News of Russian higher schools. Radio Electronics. 2006. Issue 1. P. 20-29.
6. Savchenko V.V. Automatic speech recognition by a method of a tree on the basis of an information $(R+1)$ element // News of Russian higher schools. Radio Electronics. 2006. Issue 4. P. 13-22.
7. Savchenko V.V., Akatiev D.Yu., Sherstnev S.N. The method of the optimum training dictionary in a problem of speech signals recognition by the criterion for the minimal information mismatch // News of Russian higher schools. Radio Electronics. 2006. Issue 5. P. 10-14.
8. Savchenko V.V., Akatiev D.Yu., Karpov N.V. The analysis of phonetic structure of speech signals by a method of the redefined tree // Control systems and information technologies. 2008. No. 2 (32). P. 297-303.