

УДК 519.651

## Алгоритм оценки параметров авторегрессионной модели элементарных речевых единиц

Губочкин И.В.

*Нижегородский государственный лингвистический университет им. Н.А. Добролюбова  
603155, Россия, г. Нижний Новгород, ул. Минина, д. 31а*

*e-mail: jhng@yandex.ru*

*получена 21 апреля 2012*

**Ключевые слова:** автоматическое распознавание речи, элементарные речевые единицы, информационное рассогласование, фонема

Рассмотрена задача оценки параметров авторегрессионной модели элементарных речевых единиц типа фонем. Предложен итерационный алгоритм поиска авторегрессионной модели фонемы, заданной множеством ее реализаций, в основе которого лежит метод Ньютона, предназначенный для численной минимизации функций. Для этого были получены аналитические выражения для градиента и гессиана величины информационного рассогласования Кульбака–Лейблера между моделями авторегрессии. В ходе экспериментальных исследований на наборе фонем английского языка показано, что разработанный алгоритм, по сравнению с алгоритмом выбора эталона фонемы на основе критерия минимума суммы информационных рассогласований, требует меньших вычислительных затрат на больших объемах данных, а число необходимых итераций слабо зависит от объема входных данных. Кроме того, предложенный алгоритм позволяет находить такие модели фонем, которые обеспечивают более высокую вероятность правильного распознавания.

**Введение.** Практически все современные методы автоматического распознавания речи (АРР) используют предварительное разбиение анализируемого сигнала на короткие сегменты [1]. Таким образом, указанная задача сводится к поэтапному распознаванию речи. Сначала распознаются элементарные речевые единицы (ЭРЕ) типа отдельных фонем. На втором этапе распознаются слова и отдельные фразы, представленные как последовательности различных ЭРЕ с учетом каких-либо правил (например, автоматных или статистических грамматик). Из этого следует, что задача распознавания ЭРЕ является одной из основных при АРР. В рамках байесовского подхода такая задача обычно формулируется как задача статистической классификации. Одним из перспективных инструментов для ее решения является принцип минимума информационного рассогласования [2] в метрике Кульбака–Лейблера. Его эффективность была подтверждена на целом ряде примеров из практики [3] – [6]. В частности в работе [5] был предложен алгоритм распознавания множества ЭРЕ, представленных их авторегрессионными (АР) моделями. Одной

из особенностей данного алгоритма является необходимость выбора некоторой эталонной реализации каждой ЭРЕ. В качестве эталонов в [5] использовались наиболее типичные на слух и вид реализации.

Другая работа [4] посвящена решению задачи автоматического выбора эталонной реализации при распознавании изолированных слов. Для этого был предложен критерий минимума суммы информационных рассогласований (МСИР). С его помощью среди множества реализаций заданного слова производился выбор эталона, по которому в дальнейшем вычислялись параметры его АР-модели. Из полученных АР-моделей затем формировался рабочий словарь, который использовался для решения задачи распознавания изолированных слов. При этом отмечалось существенное снижение числа ошибок распознавания по сравнению с ручным выбором эталонов.

В дальнейшем критерий МСИР был использован в рамках информационной теории восприятия речи [7] для определения ЭРЕ как информационного центра-эталона (ИЦЭ) в пределах множества ее реализаций. Однако следует отметить, что подобный критерий не является строго оптимальным в теоретико-информационном смысле, поскольку поиск ведется на конечном множестве реализаций. Кроме того, для выбора оптимальной по критерию МСИР реализации требуется полный перебор всех возможных альтернатив. На практике число реализаций различных ЭРЕ в современных речевых базах (например, ТИМІТ, NTИМІТ, АURORA), предназначенных для обучения систем АРР, может достигать нескольких тысяч. В этом случае поиск эталона будет требовать значительных временных и вычислительных затрат.

Для решения указанных выше проблем был разработан новый алгоритм оценки параметров АР-модели ЭРЕ. Его особенностью является то, что он позволяет находить строго оптимальную АР-модель с заданной точностью. В качестве критерия оптимальности используется минимум величины среднего информационного рассогласования между АР-моделью ЭРЕ и представляющими ее реализациями. В своей основе алгоритм использует итерационную минимизацию указанной величины с применением известного метода Ньютона [8], имеющего высокую скорость сходимости. Таким образом, не требуется проводить полный перебор всех возможных альтернатив для поиска оптимальной модели заданной ЭРЕ.

Данная статья посвящена выводу и экспериментальному исследованию предлагаемого алгоритма.

**Постановка задачи.** В [7] в качестве ИЦЭ некоторой ЭРЕ выбирается такая реализация речевого сигнала  $\mathbf{x}_r^* \in \{\mathbf{x}\}_r$ ,  $r = \overline{1, R}$ , которая характеризуется минимальной суммой информационных рассогласований относительно всех других реализаций данной ЭРЕ,

$$\mathbf{x}_r^* = \arg \min_k \sum_{l=1}^{L_r} \rho_{l,k}, k = \overline{1, L_r}, \quad (1)$$

где  $L_r$  – число реализаций  $r$ -й ЭРЕ,  $\rho_{l,k}$  – величина информационного рассогласования между звуками-реализациями  $\mathbf{x}_{r,l}$  и  $\mathbf{x}_{r,k}$ .

Информационное рассогласование (ИР) между неизвестным сигналом  $\mathbf{x}$  и эталоном  $\mathbf{r}$ , представленными их АР-моделями, при условии нормировки по дисперсии

порождающего шума, определяется в спектральной области как [6]:

$$\rho_{\mathbf{x},\mathbf{r}} = \frac{1}{F} \sum_{f=1}^F \frac{\left| 1 + \sum_{k=1}^P a_{\mathbf{r}}(k) e^{-j\pi k f / F} \right|^2}{\left| 1 + \sum_{k=1}^P a_{\mathbf{x}}(k) e^{-j\pi k f / F} \right|^2} - 1. \quad (2)$$

Здесь  $P$  – порядок АР-модели,  $a_{\mathbf{x}}(k)$  и  $a_{\mathbf{r}}(k)$  – элементы векторов авторегрессии сигналов  $\mathbf{x}$  и  $\mathbf{r}$  соответственно,  $F$  – верхняя граница частотного диапазона сигнала. На основе выражения (2) нетрудно определить информационное рассогласование между множеством сигналов  $\{\mathbf{x}\}$  и эталоном  $\mathbf{r}$  как среднее выборочное значение:

$$\rho_{\{\mathbf{x}\},\mathbf{r}} = \frac{1}{L} \sum_{l=1}^L \frac{1}{F} \sum_{f=1}^F \frac{\left| 1 + \sum_{k=1}^P a_{\mathbf{r}}(k) e^{-j\pi k f / F} \right|^2}{\left| 1 + \sum_{k=1}^P a_{\mathbf{x}}^l(k) e^{-j\pi k f / F} \right|^2} - 1, \quad (3)$$

где  $L$  – число элементов множества. Задача оценки параметров АР-модели ЭРЕ состоит в нахождении такого вектора  $\mathbf{a}_{\mathbf{r}}^*$ , что

$$\rho_{\{\mathbf{x}\},\mathbf{r}} \Big|_{\mathbf{a}_{\mathbf{r}}^*} \rightarrow \min. \quad (4)$$

Выражение (4) эквивалентно критерию МСИР (1) за исключением того, что выбор элементов вектора  $\mathbf{a}_{\mathbf{r}}^*$  производится из множества действительных чисел.

Для решения поставленной задачи могут быть использованы итерационные методы численной оптимизации функций. Одним из наиболее широко используемых видов таких методов является градиентная оптимизация, суть которой состоит в поиске минимума или максимума заданной функции, двигаясь вдоль направления, заданного ее градиентом.

Одним из известных методов градиентной оптимизации является метод Ньютона. Его особенностью является то, что он обладает высокой скоростью сходимости и подходит для минимизации квадратичных функций, к которым относится и величина ИР, определенная в виде (2) или (3) при фиксированных значениях элементов векторов  $\mathbf{a}_{\mathbf{x}}^l$ . Рассмотрим более подробно построение алгоритма численной минимизации  $\rho_{\{\mathbf{x}\},\mathbf{r}}$  с помощью указанного выше метода.

**Синтез алгоритма.** Сначала найдем решение задачи (4), когда множество  $\{\mathbf{x}\}$  состоит из одного элемента. В данном случае нам необходимо найти минимум выражения (2). При этом определим частную производную информационного рассогласования сигналов  $\mathbf{x}$  и  $\mathbf{r}$  как  $g_m = \frac{\partial \rho_{\mathbf{x},\mathbf{r}}}{\partial a_{\mathbf{r}}(m)}$ . Для ее нахождения введем следующие две функции:

$$\Lambda(f) = \left| 1 + \sum_{k=1}^P a_{\mathbf{r}}(k) e^{-j\pi k f / F} \right|^2, \quad (5)$$

$$D(f) = \left| 1 + \sum_{k=1}^P a_{\mathbf{x}}(k) e^{-j\pi k f / F} \right|^2. \quad (6)$$

Тогда формулу (2) можно переписать следующим образом:

$$\rho_{\mathbf{x},\mathbf{r}} = \frac{1}{F} \sum_{f=1}^F \frac{\Lambda(f)}{D(f)} - 1. \quad (7)$$

Найдем частную производную от  $\Lambda(f)$ . С учетом того, что  $e^{i\phi} = \cos \phi + i \sin \phi$  получаем:

$$\frac{\partial \Lambda(f)}{\partial a_{\mathbf{r}}(m)} = \frac{\partial}{\partial a_{\mathbf{r}}(m)} \left[ \left( \underbrace{1 + \sum_{k=1}^P a_{\mathbf{r}}(k) \cos(-\pi k f / F)}_A \right)^2 + \left( \underbrace{\sum_{k=1}^P a_{\mathbf{r}}(k) \sin(-\pi k f / F)}_B \right)^2 \right]. \quad (8)$$

Теперь возведем в квадрат часть  $A$  полученного выражения:

$$A^2 = 1 + \sum_{k=1}^P a_{\mathbf{r}}(k)^2 \cos^2(-\pi k f / F) + 2 \sum_{k=1}^P a_{\mathbf{r}}(k) \cos(-\pi k f / F) + 2 \sum_{\substack{k=1, P-1 \\ n=k+1, P}} a_{\mathbf{r}}(k) a_{\mathbf{r}}(n) \cos(-\pi k f / F) \cos(-\pi n f / F).$$

Аналогичное выражение можно получить и для части  $B$ :

$$B^2 = \sum_{k=1}^P a_{\mathbf{r}}(k)^2 \sin^2(-\pi k f / F) + 2 \sum_{\substack{k=1, P-1 \\ n=k+1, P}} a_{\mathbf{r}}(k) a_{\mathbf{r}}(n) \sin(-\pi k f / F) \sin(-\pi n f / F).$$

Отсюда нетрудно найти производные для квадратов частей  $A$  и  $B$  выражения (8):

$$\frac{\partial A^2}{\partial a_{\mathbf{r}}(m)} = 2 \cos(\pi m f / F) \left[ 1 + \sum_{k=1}^P a_{\mathbf{r}}(k) \cos(\pi k f / F) \right],$$

$$\frac{\partial B^2}{\partial a_{\mathbf{r}}(m)} = 2 \sin(\pi m f / F) \left[ \sum_{k=1}^P a_{\mathbf{r}}(k) \sin(\pi k f / F) \right].$$

Тогда выражение для частной производной  $g_m$  с учетом (7) приобретает следующий вид:

$$g_m = \frac{1}{F} \sum_{f=1}^F \frac{1}{D(f)} \left[ \frac{\partial A^2}{\partial a_{\mathbf{r}}(m)} + \frac{\partial B^2}{\partial a_{\mathbf{r}}(m)} \right].$$

Также найдем гессиан  $\mathbf{G}$  выражения (2). Для этого введем следующее обозначение для второй частной производной от  $\Lambda(f)$ :

$$c_{m,n}(f) = \frac{\partial^2 \Lambda(f)}{\partial a_{\mathbf{r}}(m) \partial a_{\mathbf{r}}(n)} = 2 [\cos(\pi m f / F) \cos(\pi n f / F) + \sin(\pi m f / F) \sin(\pi n f / F)].$$

В результате получаем:

$$\mathbf{G} = \frac{1}{F} \sum_{f=1}^F \frac{1}{D(f)} \begin{bmatrix} 2 & c_{1,2}(f) & \cdots & c_{1,P}(f) \\ c_{2,1}(f) & 2 & \cdots & c_{2,P}(f) \\ \vdots & \vdots & \ddots & \vdots \\ c_{P,1}(f) & c_{P,2}(f) & \cdots & 2 \end{bmatrix}. \quad (9)$$

Рассмотрим теперь алгоритм минимизации выражения (2) с помощью метода Ньютона. Для этого зададим начальное приближение АР-модели эталона  $\mathbf{a}_{\mathbf{r},0}$  и введем следующие обозначения:  $\mathbf{H} = \mathbf{G}^{-1}$ ,  $\mathbf{g} = [g_1, g_2, \dots, g_P]^T$ . Тогда

$$\mathbf{a}_{\mathbf{r},k+1} = \mathbf{a}_{\mathbf{r},k} + \eta_k^* \mathbf{s}_k \quad (10)$$

где  $\eta_k^*$  – это оптимальный размер шага  $k$  вдоль направления  $\mathbf{s}_k$ , которое задается следующей формулой:

$$\mathbf{s}_k = \frac{\mathbf{H} \mathbf{g}_k}{\|\mathbf{H} \mathbf{g}_k\|}. \quad (11)$$

Размер шага  $\eta_k^*$  может быть найден различными способами, например, с помощью линейного поиска. В нашем случае, с учетом специфики решаемой задачи, хорошим приближением для выбора величины шага может служить использование собственных чисел  $\lambda$  матрицы  $\mathbf{H}$ . При этом оптимальное значение выбирается как

$$\begin{aligned} \bar{\mathbf{a}}(\eta) &= \mathbf{a}_{\mathbf{r},k} + \eta \mathbf{s}_k, \\ \eta_k^* &= \arg \min_{\lambda} \rho(\mathbf{x}, \mathbf{r} | \bar{\mathbf{a}}(\lambda)), \end{aligned} \quad (12)$$

где  $\rho(\mathbf{x}, \mathbf{r} | \bar{\mathbf{a}}(\lambda))$  – это информационное рассогласование между сигналом  $\mathbf{x}$  и эталоном  $\mathbf{r}$ , который представлен оценкой его АР-модели  $\bar{\mathbf{a}}(\bullet)$ . Остановка алгоритма происходит в случае, когда изменение величины ИР становится достаточно малым, т.е.  $|\rho_{\mathbf{x},\mathbf{r}}^{k+1} - \rho_{\mathbf{x},\mathbf{r}}^k| < \epsilon$ , где  $\epsilon$  – небольшое положительное число, либо число шагов  $k$  превысит некоторое максимально допустимое значение.

С учетом (3) нетрудно обобщить алгоритм (10) – (12) на случай множества анализируемых сигналов  $\{\mathbf{x}\}$ . Для этого введем следующие обозначения:

$$\begin{aligned} \mathcal{G} &= \frac{1}{L} \sum_{l=1}^L \mathbf{G}^l, \\ \mathcal{H} &= \mathcal{G}^{-1}, \\ \mathcal{g} &= \frac{1}{L} \sum_{l=1}^L \mathbf{g}^l. \end{aligned}$$

Тогда для решения задачи (4) выражение (11) можно переписать следующим образом:

$$\mathbf{s}_k = \frac{\mathcal{H}\mathbf{g}_k}{\|\mathcal{H}\mathbf{g}_k\|}. \quad (13)$$

Предложенный алгоритм (10) – (13) был использован для оценки параметров оптимальных по критерию (4) АР-моделей нескольких элементарных речевых единиц типа фонем. Результаты проведенных экспериментальных исследований приведены далее.

**Результаты экспериментальных исследований.** Для экспериментальных исследований алгоритма оценки параметров в формулировке (10) – (13) были выбраны десять фонем английского языка из широко известной речевой базы ТИМТ [9]:  $[uh]$ ,  $[ih]$ ,  $[w]$ ,  $[hh]$ ,  $[sh]$ ,  $[p]$ ,  $[b]$ ,  $[nx]$ ,  $[ch]$ ,  $[s]$ . Эта база содержит записи речи 630 дикторов, говорящих на восьми основных диалектах американского английского. Каждая запись находится в цифровом формате (частота дискретизации 16 кГц, 16 бит) в отдельном файле и имеет орфографическую, фонетическую и словарную транскрипцию.

Реализации ЭРЕ были получены путем извлечения из звуковых файлов базы ТИМТ необходимых фрагментов данных в соответствии с фонетической разметкой. Из этих фрагментов в дальнейшем была сформирована экспериментальная фонетическая база.

По каждому сигналу ЭРЕ из фонетической базы были вычислены векторы АР-коэффициентов  $\mathbf{a}_x$  фиксированного порядка  $P=12$ . Для этого применялась рекуррентная процедура Берга–Левинсона [10]. Также были заданы начальные приближения АР-моделей эталонов  $\mathbf{a}_{r,0}$  и условие останова  $\epsilon = 10^{-4}$ .

На рис. 1 представлены графики спектральной плотности мощности (СПМ), вычисленные с использованием (5) и (6). Здесь штриховой линией показана СПМ начального приближения АР-модели, тонкими сплошными линиями показаны СПМ для двух реализаций выбранной фонемы, а полужирная линия соответствует СПМ, найденной с помощью алгоритма (10) – (13) оптимальной АР-модели.

Из рисунка видно, что найденная модель в определенной степени учитывает особенности обоих сигналов. Данный факт является подтверждением того, что разработанный алгоритм позволяет находить адекватные АР-модели сигналов, оптимальные по критерию (4).

На следующем этапе экспериментальных исследований было проведено сравнение величины среднего информационного рассогласования (3) для АР-модели эталона, выбранного по критерию МСИР (1) и АР-модели, найденной с помощью разработанного алгоритма. Полученные результаты для разных объемов выборки реализаций фонем  $L$  представлены в таблице 1.

Из таблицы видно, что величина информационного рассогласования (3) для АР-модели, найденной с помощью разработанного алгоритма, меньше, чем у модели эталона, выбранного с помощью алгоритма, основанного на использовании критерия МСИР. Следует отметить, что для вокализованных фонем (таких, как  $[uh]$ ,  $[ih]$ ,  $[w]$ ,  $[nx]$ ), наблюдается большее различие в величинах ИР между полученными АР-моделями (порядка 0,3 – 0,6). Вместе с тем, для невокализованных фонем это различие меньше (0,1 – 0,2). Данный факт можно объяснить тем, что вокализованные фонемы имеют более сложную структуру и их спектры сильнее зависят

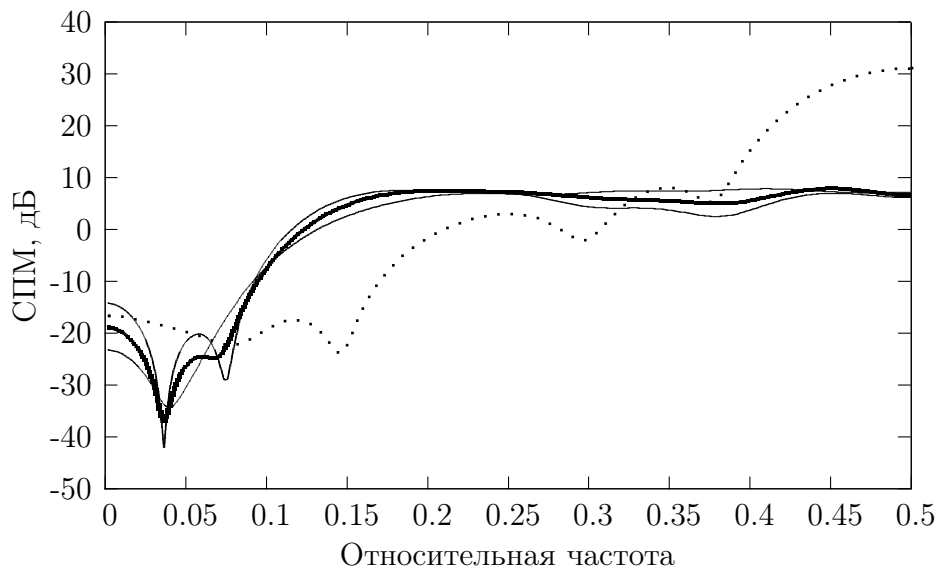


Рис. 1. Результат работы алгоритма оценки параметров АР-модели

Таблица 1. Величина среднего информационного рассогласования

Алгоритм		Фонема									
		[uh]	[ih]	[w]	[hh]	[sh]	[p]	[b]	[nx]	[ch]	[s]
алг. (10) – (13)	L=50	1,21	0,69	0,96	1,02	0,55	0,47	0,6	1,3	0,69	0,32
	L=100	1,26	0,98	0,98	1,14	0,74	0,52	0,82	1,4	0,7	0,44
	L=200	1,23	1,1	1,16	1,17	0,73	0,6	0,86	1,46	0,71	0,54
	L=300	1,24	1,19	1,09	1,13	0,71	0,59	0,91	1,52	0,71	0,74
МСИР	L=50	1,71	1,03	1,23	1,5	0,72	0,5	0,71	1,79	0,83	0,38
	L=100	1,63	1,28	1,24	1,51	0,95	0,58	0,89	1,8	0,86	0,5
	L=200	1,54	1,39	1,46	1,54	0,81	0,67	0,92	1,87	0,82	0,59
	L=300	1,53	1,49	1,37	1,39	0,79	0,65	0,97	1,96	0,8	0,89

от параметров речевого тракта конкретного диктора. Предложенный же алгоритм находит АР-модель с такими параметрами, которые в максимальной степени учитывают особенности речевого тракта всех дикторов.

В таблице 2 приведено количество шагов алгоритма (10) – (13), необходимое для достижения заданной точности при различном числе реализаций ЭРЕ. Видно, что алгоритм сходится после небольшого (от 4 до 15) числа итераций, которое слабо зависит от объема входных данных.

В ходе экспериментальных исследований также проводилось сравнение вычислительной эффективности рассматриваемых алгоритмов. В качестве иллюстрации на рис. 2 представлена зависимость времени работы алгоритма (10) – (13) (обозначена символами '\*)') и алгоритма на основе критерия (1) (обозначена символами '+') от количества реализаций фонемы [s]. Линиями обозначены соответствующие аппроксимации. Алгоритмы были реализованы в виде компьютерной программы на языке matlab и исполнялись в среде Matlab 2006 на персональном компьютере со следующими характеристиками: ОС Windows XP, ЦП AMD Athlon II X4 3 GHz, 1024 МВ оперативной памяти. Видно, что зависимость временных затрат для алгоритма с использованием МСИР близка к квадратичной, что соответствует характеру

Таблица 2. Скорость сходимости алгоритма оценки параметров AP-модели

Число реализаций	Фонема									
	[uh]	[ih]	[w]	[hh]	[sh]	[p]	[b]	[nx]	[ch]	[s]
$L=50$	5	5	6	7	10	6	6	5	13	11
$L=100$	5	5	6	8	12	8	7	5	12	10
$L=200$	6	5	5	7	13	7	5	6	15	11
$L=300$	5	4	4	7	12	6	7	6	12	11

производимых вычислений. По сравнению с этим время работы алгоритма (10) – (13) практически линейно зависит от числа реализаций. Аналогичные результаты были получены и для других фонем из анализируемого набора. На основе приведенных данных можно сделать вывод, что разработанный алгоритм более эффективен по скорости вычислений при оценке параметров AP-модели ЭРЕ на множествах реализаций большого объема (от 1000 и выше).

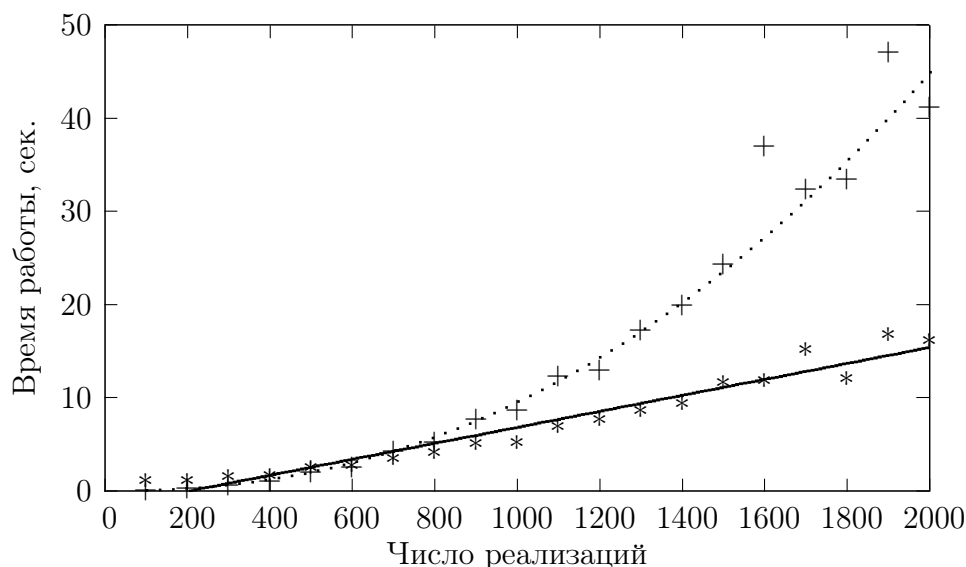


Рис. 2. Время работы алгоритмов в зависимости от числа реализаций

В заключительной части исследований было проведено сравнение вероятностей правильного распознавания ЭРЕ с использованием AP-моделей, найденных алгоритмом на основе критерия (1) и разработанным алгоритмом. Данная вероятность для  $r$ -й ЭРЕ вычислялась как

$$P_r = 1 - \frac{w_{\text{ош},r}}{L_r},$$

где  $w_{\text{ош},r}$  – число ошибочно распознанных реализаций  $r$ -й ЭРЕ. Полученные результаты для объема выборки  $L=300$  реализаций по каждой фонеме приведены в таблице 3.

Из таблицы видно, что разработанный алгоритм в большинстве случаев позволяет находить такую AP-модель заданной ЭРЕ, которая обеспечивает более высокую точность распознавания по сравнению с моделью эталона, выбранного с помощью



Таблица 3. Вероятность правильного распознавания фонем

Алгоритм	Фонема										Итого
	[uh]	[ih]	[w]	[hh]	[sh]	[p]	[b]	[nx]	[ch]	[s]	
алг. (10) – (13)	0,36	<b>0,63</b>	<b>0,77</b>	<b>0,34</b>	0,5	0,59	0,23	<b>0,43</b>	<b>0,52</b>	<b>0,89</b>	<b>0,53</b>
МСИР	0,37	0,56	0,67	0,33	0,68	0,7	0,25	0,34	0,32	0,84	<b>0,51</b>

критерия МСИР (1). Для фонем [uh] и [b] найденная АР-модель обеспечивала несущественное ухудшение точности распознавания. Вероятность правильного распознавания фонем [sh] и [p] ухудшилась на 36% и 19% соответственно, что, однако, не помешало получить в целом более высокий результат.

**Выводы.** В статье был предложен итерационный алгоритм оценки параметров оптимальной по критерию минимума среднего информационного рассогласования АР-модели ЭРЕ типа фонем. Алгоритм быстро сходится к оптимальному решению, причем число необходимых итераций слабо зависит от объема входных данных. Сравнение с алгоритмом выбора эталона из множества реализаций по критерию МСИР показало, что предложенный алгоритм показывает большую вычислительную эффективность на выборках, состоящих из 1000 и более реализаций. Кроме того, он позволяет находить такие АР-модели ЭРЕ, которые обеспечивают более высокую вероятность правильного распознавания.

**Дальнейшие направления исследований.** Предложенный алгоритм оценки параметров АР-модели может быть обобщен и на случай смеси АР-моделей. В этом случае логично рассчитывать на повышение эффективности распознавания за счет одновременного использования сразу нескольких моделей, которые лучше представляют ту или иную ЭРЕ. Для реализации подобного подхода можно использовать как градиентные методы поиска, так и ЕМ-алгоритм или его модификации. Однако способ их применения и оценка эффективности требует дополнительных исследований.

## Список литературы

1. Picone J. Signal Modeling Techniques in Speech Recognition // Proceedings of the IEEE. Vol. 81, No. 9. P. 1215 – 1246.
2. 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 – 314 (Savchenko V.V. Avtomaticheskaya obrabotka rechi po kriteriyu minimuma informatsionnogo rassoglasovaniya na osnove metoda obelyayushchego filtra // Radiotekhnika i elektronika. 2005. T. 50, 3. S. 309 – 314 [in Russian]).
3. Savchenko V.V. Autoregressive Model for Recognition of Speech Signals Based on Theoretical Information Approach // VI International Congress on Mathematical Modeling. N. Novgorod: University of N. Novgorod, 2004. P. 347.
4. Савченко В.В., Акатьев Д.Ю., Шерстнев С.Н. Метод оптимального обучающего словаря в задаче распознавания речевых сигналов по критерию минимального информационного рассогласования // Известия высших учебных заведений России. Радиоэлектроника. 2006. Вып. 5. С. 10 – 14 (Savchenko V.V., Akatjev D.Yu., Sherstnev S.N. Metod

- optimalnogo obuchayushchego slovarya v zadache raspoznavaniya rechevykh signalov po kriteriyu minimalnogo informatsionnogo rassoglasovaniya // *Izvestiya vysshikh uchebnykh zavedeny Rossii. Radioelektronika*. 2006. Вып. 5. S. 10 – 14 [in Russian]).
5. Савченко В.В., Акатьев Д.Ю., Карпов Н.В. Автоматическое распознавание элементарных речевых единиц методом обеляющего фильтра // *Известия высших учебных заведений России. Радиоэлектроника*. 2007. Вып. 4. С. 11 – 19 (Savchenko V.V., Akatev D.Yu., Karpov N.V. Avtomaticheskoe raspoznavanie elementarnykh rechevykh edinit metodom obelyayushchego filtra // *Izvestiya vysshikh uchebnykh zavedeny Rossii. Radioelektronika*. 2007. Вып. 4. S. 11 – 19 [in Russian]).
  6. Савченко В.В., Пономарев Д.А. Оптимизация фонетической базы данных по группе дикторов на основе информационной теории восприятия речи // *Информационные технологии*. 2009. №12. С. 7 – 12 (Savchenko V.V., Ponomarev D.A. Optimizatsiya foneticheskoy bazy dannykh po gruppe diktorov na osnove informatsionnoy teorii vospriyatiya rechi // *Informatsionnye tekhnologii*. 2009. 12. S. 7 – 12 [in Russian]).
  7. Савченко В.В. Информационная теория восприятия речи // *Известия высших учебных заведений России. Радиоэлектроника*. 2007. Вып. 6. С. 3 – 9. (Savchenko V.V. Informatsionnaya teoriya vospriyatiya rechi // *Izvestiya vysshikh uchebnykh zavedeny Rossii. Radioelektronika*. 2007. Вып. 6. S. 3 – 9 [in Russian])
  8. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. М.: Наука, 1975 (Pshenichny B. N., Danilin Yu. M. Numerical Methods in Extremal Problems. MIR Publishers, Moscow, Russia, 1978.).
  9. Fisher William M., Doddington George R. and Goudie-Marshall Kathleen M. The DARPA Speech Recognition Research Database: Specifications and Status // *Proceedings of DARPA Workshop on Speech Recognition*. 1986. P. 93 – 99.
  10. Марпл С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 584 с. (Marple Jr S. L. Digital spectral analysis with applications. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1987. 512 p.)

## An Algorithm for Parameters Estimation of Autoregressive Model of Basic Speech Units

Gubochkin I.V.

*Linguistics University of Nizhny Novgorod, Minin st., 31a, Nizhny Novgorod, 603155, Russia*

**Keywords:** automatic speech recognition, basic speech units, information divergence, phoneme

The article considers the problem of estimating autoregressive model parameters of elementary speech units such as phonemes. It is suggested an iterative algorithm based on the Newton numerical minimization technique to search an autoregressive model of phonemes specified its multiple samples. For this purpose the analytical expressions of the gradient and the Hessian of Kullback–Leibler information divergence between autoregressive models were computed. Experimental studies on a set of English phonemes showed that the developed algorithm requires less computational effort for large amounts of data, and iterations count depends little on the amount of input data as opposed to reference phoneme selection algorithm based on the criterion of a minimum sum of information divergence. Moreover, the proposed algorithm allows finding models of phonemes, which provide a higher probability of correct recognition.

### Сведения об авторе:

**Губочкин Иван Вадимович,**

Нижегородский государственный лингвистический университет

им. Н.А. Добролюбова,

канд. техн. наук, старший преподаватель.

Автор более 20 публикаций. Научные интересы – автоматическая обработка и распознавание речевых сигналов, обработка изображений.