

# СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ РАСПАРАЛЛЕЛИВАНИЯ НЕЙРОАЛГОРИТМА РАСПОЗНАВАНИЯ РЕЧИ НА ВЫЧИСЛИТЕЛЬНЫХ АРХИТЕКТУРАХ OPENMP И CUDA

Шатохин Н. А., Бондаренко И. Ю.

Донецкий национальный технический университет

## Введение

С момента появления первых компьютеров одним из наиболее важных вопросов развития компьютерной техники был процесс взаимодействия человека с машиной. Долгое время это было доступно только узким специалистам – технологи «общались» с машиной через посредника-программиста. Такая ситуация просуществовала вплоть до появления диалогового интерфейса, когда пользователь смог лично вводить с клавиатуры адресованную машине команду и получать осмысленный ответ. Дальнейшее появление графического интерфейса, в котором отпала необходимость в знании человеком каких-либо команд, привела к повсеместному распространению персональных компьютеров.

Однако человек всегда стремился к более универсальному и естественному способу взаимодействия. Еще в эпоху перфокарт в научно-фантастических романах человек с компьютером разговаривал, как равным себе. Тогда же были предприняты первые шаги по реализации речевого интерфейса.

Распознавание речи — процесс преобразования речевого сигнала в текстовый поток. Речевое общение является естественным и удобным для человека. Задача распознавания речи состоит в том, что бы убрать посредника в общении человека и компьютера. Управление машиной голосом в реальном времени, а также ввод информации посредством человеческой речи намного упростит жизнь современного человека. Научить машину понимать без посредника тот язык, на котором говорят между собой люди – задачи распознавания речи.

Системы автоматического распознавания речи сегодня находят широкое применение в различных областях жизнедеятельности человека. Их используют инвалиды, неспособные пользоваться традиционными методами ввода, автомобилисты за рулем, телефонные операторы для создания роботов-автоответчиков и др.

## 1. Архитектура системы распознавания речи

Распознавание слитной речи представляет собой многоуровневый процесс. После предварительной обработки речевого сигнала и выделения из него информативных признаков выполняется выделение лексических элементов речи. Это первый уровень распознавания. На втором уровне выделяются слоги и морфемы, на третьем — слова, предложения и сообщения. На каждом уровне сигнал кодируется представителями предыдущих уровней. То есть слоги и морфемы составляются из фонем и аллофонов, слова — из слогов и морфем, предложения и сообщения — из слов.

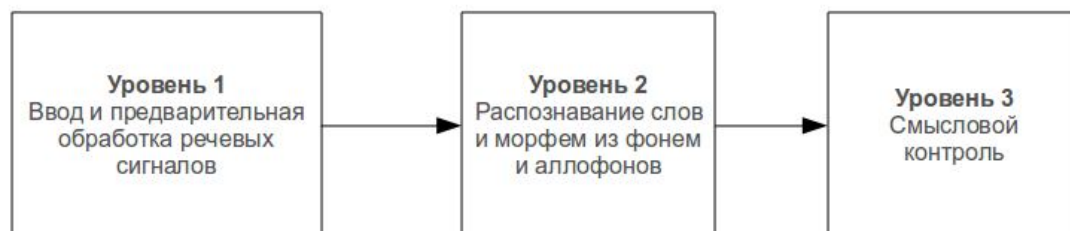


Рис. 1.1. Процесс распознавания речи

В данной работе мы будем использовать искусственные нейронные сети для распознавания

речи. В таком случае, наша система будет выглядеть следующим образом.



Рис. 1.2. Структура системы распознавания речи

Речевой сигнал может поступать как из файла, так и в реальном времени с микрофона. Для того, чтобы звук можно было подать на вход нейросети, необходимо осуществить над ним некоторые преобразования. Очевидно, что представление звука во временной форме неэффективно. Оно не отражает характерных особенностей звукового сигнала. Гораздо более информативно спектральное представление речи. Для получения спектра используется набор полосовых фильтров, настроенных на выделение различных частот, или дискретное преобразование Фурье. Затем полученный спектр подвергается различным преобразованиям, например, логарифмическому изменению масштаба (как в пространстве амплитуд, так и в пространстве частот). Это позволяет учесть некоторые особенности речевого сигнала – понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Как правило, полное описание речевого сигнала только его спектром невозможно. Наряду со спектральной информацией, необходима ещё и информация о динамике речи. Для её получения используются дельта-параметры, представляющие собой производные по времени от основных параметров.

Полученные таким образом параметры речевого сигнала считаются его первичными признаками и подаются на вход нейронной сети, на выходе которой будут соответствующие сигналы фонемы. Затем фонемы собираются в слова и предложения.

Наиболее медленным участком этой цепи является нейронная сеть, т.к. для точного распознавания требуется достаточно большое число нейронов (в одном только входном слое для извлечения лишь информативной части спектра, без учета интонации, необходимо около 50-100 нейронов). В то же время, чтобы найти выход одного нейрона необходимо вычислить взвешенную сумму всех его входов и значение пороговой функции. Для вычисления взвешенной суммы необходимо произвести число умножений и сложений в соответствии с числом входных каналов. После этого вычислить значение пороговой функции, а это, в зависимости от сложности самой функции, от одной операции сравнения, до нескольких математических операций.

Но так как все нейроалгоритмы являются высокопараллельными, выходы нейронов одного слоя могут вычисляться одновременно. Следовательно, нейронная сеть может быть эффективно распараллелена, что позволит достичь высокого быстродействия.

## 2. Отображение нейроалгоритма на различные вычислительные архитектуры

В качестве нейросети используется нейронная сеть с обратным распространением ошибки. Входы каждого нейрона такой сети являются выходами всех нейронов предыдущего слоя. Сеть задается числом слоев, числом нейронов в каждом слое, число входов и вектором весовых коэффициентов.

## 2.1. Отображение нейроалгоритма на однопоточную архитектуру

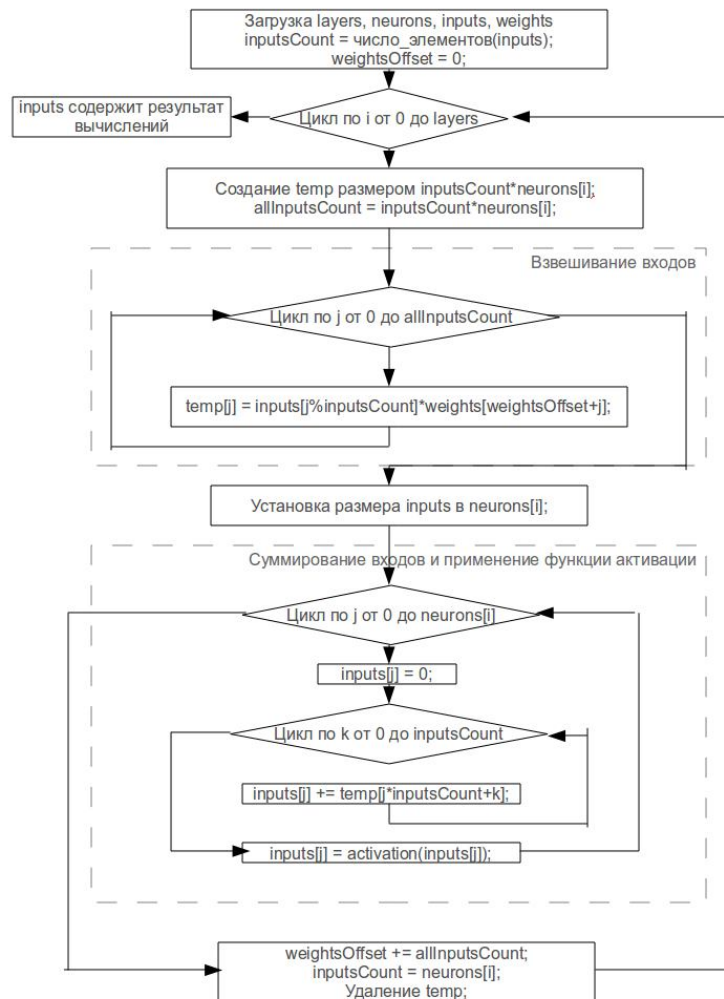


Рис. 2.1. Отображение нейроалгоритма на однопоточную архитектуру

## 2.2. Отображение нейроалгоритма на архитектуру OpenMP

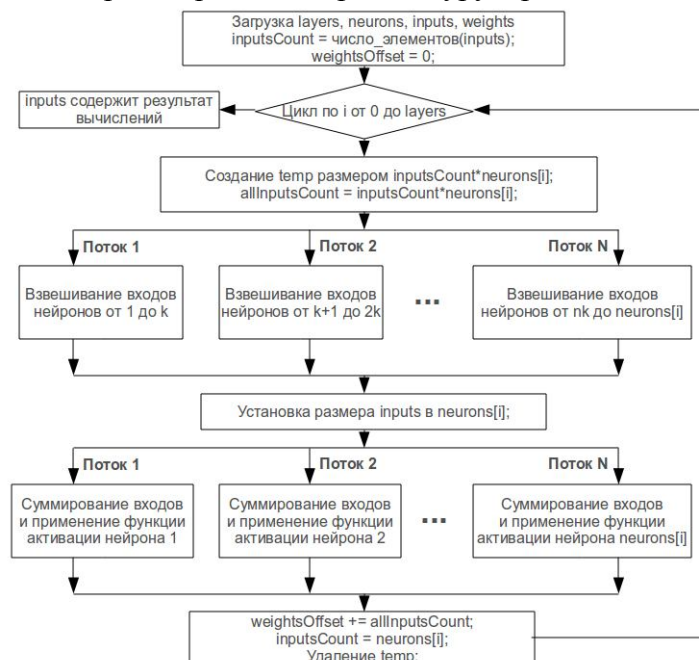


Рис. 2.2. Отображение нейроалгоритма на архитектуру OpenMP

## 2.3. Отображение нейроалгоритма на архитектуру CUDA

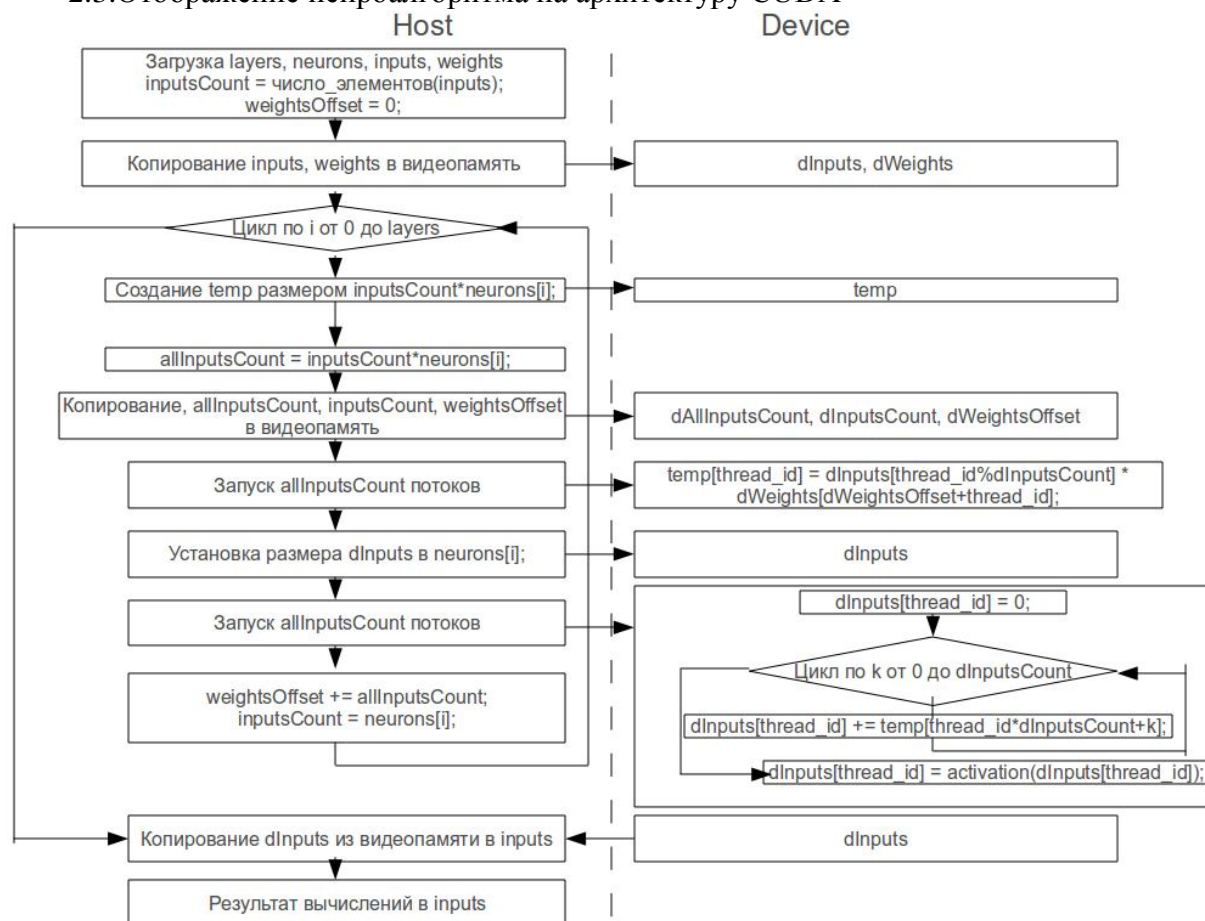


Рис. 2.3. Отображение нейроалгоритма на архитектуру CUDA

## 3. Описание экспериментов

### 3.1. Описание исходных данных для экспериментов и методики экспериментов

На вход приложений подается два бинарных файла. Первый содержит данные о структуре нейронной сети (см. табл. 1), второй — набор входных данных.

Таблица 1

Описание структуры нейронной сети

Название поля	Размер	Назначение
layers	4 байта	Количество слоев нейросети
inputsCount	4 байта	Количество входов нейросети
neurons	4*layers байт	Массив — число нейронов в каждом слое
weights	4*(inputsCount*neurons[0] + ... +neurons[layers- 1]*neurons[layers] байт	Массив весовых коэффициентов каждого нейрона

Второй файл содержит список речевых сигналов (4 байтовых чисел с плавающей запятой типа float) из базы ТИМТ, предварительно обработанных спектральным анализатором. Приложение считывает из файла группу сигналов, количество которых равно inputsCount, и подает их на вход нейросети. После вычисления выходных сигналов, приложение берет следующую группу и вновь подает на вход нейросети и т. д. до тех пор, пока не будет обработан весь файл. Время, затраченное на обработку всего файла, является результатом работы приложения.

Тестирование проводилось на компьютере со следующими характеристиками:

**Процессор:** Intel Core2Duo T7500 2,2GHz

**Оперативная память:** 2048Mb DDR2

**Видеокарта:** Nvidia GeForce 8600M GS

**Видеопамять:** 128Mb DDR2

Для тестирования была создана четырехслойная сеть с 400 входами и 454, 300, 150, 41 нейронами в каждом слое соответственно. Были обработаны 5 файлов, содержащих 1 пакет речевых сигналов, 10, 100, 1000 и 10000 соответственно.

### 3.2. Сравнительный анализ результатов экспериментов на различных архитектурах

Результаты экспериментов приведены в табл. 2.

Таблица 2

Результаты экспериментов

Число пакетов	Время вычисления (нс)		
	CPU	OpenMP	CUDA
1	7189740	5530569	1892036
10	43053301	28702200	10250785
100	332213405	201341457	70683703
1000	3293173884	1780093991	621353563
10000	32732682709	16615574979	5455447118

### Выводы

Как мы можем увидеть из результатов, распараллеливание действительно ускоряет работу приложения. Причем, чем больше объем данных, тем выше прирост производительности. Это объясняется тем, что на порождение потоков (в случае OpenMP) и на копирование данных в видеопамять (в случае CUDA) также затрачивается время. Следовательно, распараллеливание может быть эффективно только в тех задачах, в которых есть необходимость обрабатывать большие объемы данных. Автоматическое распознавание устной речи является одной из таких задач.

### Список литературы

1. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения. [Электронный ресурс] — 2003. - Режим доступа: <http://www.frolov-lib.ru/books/hi/index.html>
2. Алексеев В. Услышь меня, машина. / В. Алексеев // Компьютерра, - 1997. - №49.
3. Ф. Уоссермен «Нейрокомпьютерная техника: Теория и практика». Перевод на русский язык Ю. А. Зуев, В. А. Точенов, 1992.
4. Винцюк Т.К. «Анализ, распознавание и интерпретация речевых сигналов.» -Киев: Наук. Думка, - 1987. -262 с.

## Заявка участника

**Название доклада** Сравнительный анализ эффективности распараллеливания нейроалгоритма распознавания речи на вычислительных архитектурах OpenMP и CUDA

**Секция выступления** 3 - Искусственный интеллект, нейросетевые и эволюционные алгоритмы, экспертные системы.

**ВУЗ** Донецкий национальный технический университет

**Автор** Шатохин Николай Андреевич

**Адрес автора** г. Донецк, ул. Щетинина, д. 40, кв. 120

**E-mail автора** [n.shatokhin@gmail.com](mailto:n.shatokhin@gmail.com)

**Телефон автора** +380979562244

**Соавторы** \_\_\_\_\_

**Руководитель** Бондаренко Иван Юрьевич

**Звание** \_\_\_\_\_

**Научная степень** \_\_\_\_\_

**Должность** ассистент кафедры ПМИ ДонНТУ

**Адрес руководителя** Донецк, 83003, пр.Ильича, 42, кв. 173

**E-mail руководителя** [bond005@yandex.ru](mailto:bond005@yandex.ru)

**Телефон руководителя** 301-07-57 (раб)