

НАУЧНАЯ МЫСЛЬ



А.Г. Шишкин

**МЕТОДЫ
ЦИФРОВОЙ ОБРАБОТКИ
И РАСПОЗНАВАНИЯ РЕЧИ**



Данная книга доступна
в цветном исполнении
в электронно-библиотечной
системе Znanium



НАУЧНАЯ МЫСЛЬ

СЕРИЯ ОСНОВАНА В 2008 ГОДУ

А.Г. ШИШКИН

**МЕТОДЫ ЦИФРОВОЙ
ОБРАБОТКИ
И РАСПОЗНАВАНИЯ РЕЧИ**

МОНОГРАФИЯ

znanium.com

электронно-библиотечная система

Москва
ИНФРА-М
2024

УДК 004.934(075.4)

ББК 81.111

Ш65

Автор:

Шишкин А.Г., доктор физико-математических наук, профессор кафедры автоматизации научных исследований факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова

Шишкин А.Г.

Ш65 Методы цифровой обработки и распознавания речи : монография / А.Г. Шишкин. — Москва : ИНФРА-М, 2024. — 347 с. — (Научная мысль). — DOI 10.12737/1904325.

ISBN 978-5-16-018017-5 (print)

ISBN 978-5-16-111023-2 (online)

В монографии рассмотрены теория, алгоритмы и практические методы реализации цифровой обработки и распознавания речевых сигналов. Представлены основы математического анализа цифровых сигналов, необходимые для обработки речи. Кратко изложена акустическая теория речеобразования с построением общей дискретной модели. Рассмотрены основные характерные признаки речевых сигналов, а также методы их выделения. Подробно описаны скрытые марковские модели и архитектура традиционных систем распознавания на их основе. Рассмотрены взвешенные конечные преобразователи, использующиеся для повышения эффективности и ускорения процесса декодирования акустических сигналов. Представлены основные архитектуры искусственных нейронных сетей и примеры основанных на них интегральных (end-to-end) систем распознавания речи.

Предназначена для студентов, аспирантов, научных работников и специалистов, занимающихся вопросами обработки речевых сигналов, распознавания образов и искусственного интеллекта.

УДК 004.934(075.4)

ББК 81.111



Данная книга доступна в цветном исполнении в электронно-библиотечной системе Znanium

ISBN 978-5-16-018017-5 (print)

ISBN 978-5-16-111023-2 (online)

© Шишкин А.Г., 2023

Оглавление

Введение	3
Глава 1. Краткая история систем цифровой обработки речи	7
1.1. Что такое распознавание речи	7
1.2. История систем автоматического распознавания речи	13
1.3. Классификация систем распознавания речи	19
Глава 2. Основные понятия и методы цифровой обработки сигналов	21
2.1. Сигналы и системы	21
2.2. Импульсная характеристика систем	26
2.3. Свертка	28
2.4. Корреляция	31
2.5. Преобразование Фурье	34
2.5.1. Ряды Фурье	34
2.5.2. Феномен Гиббса	37
2.5.3. Свойства рядов Фурье	39
2.5.4. Интеграл Фурье	41
2.5.5. Дискретные преобразования Фурье	43
2.5.6. Некоторые особенности применения ДПФ	52
2.6. Теорема Котельникова и наложение спектров	60
2.7. Z-преобразование	65
2.8. Дискретное косинусное преобразование	74
2.9. Цифровые фильтры	75
Глава 3. Акустическая теория речеобразования	89
3.1. Структура речеобразующей системы человека	89
3.2. Акустические свойства звуков русского языка	92
3.3. Передаточная функция голосового тракта	103
3.4. Модель акустической трубы с учетом потерь энергии	106
3.5. Моделирование голосового тракта как соединение труб без потерь	110
3.6. Дискретная модель речевого сигнала	114
Глава 4. Кратковременный анализ речевых сигналов	118
4.1. Предварительная обработка сигнала	119
4.2. Кратковременная функция нулевых пересечений	122
4.3. Кратковременная энергия сигнала	124
4.4. Автокорреляционная функция	126
4.5. Коэффициенты линейного предсказания	129
4.6. Кепстральные коэффициенты	139
4.7. Мел-частотные кепстральные коэффициенты	146
4.8. Частота основного тона	152
4.9. Определение формант	161
4.10. Преобразование параметров	167

Глава 5. Скрытые модели Маркова	170
5.1. Дискретные цепи Маркова	170
5.2. Скрытые модели Маркова	173
5.3. Основные задачи СММ и их решения	179
5.3.1. Оценка вероятности последовательности наблюдений для данной модели	179
5.3.2. Выбор «оптимальной» последовательности состояний	186
5.3.3. Выбор параметров модели	195
5.4. Масштабирование СММ	201
Глава 6. Архитектура распознающей системы на основе скрытых моделей Маркова и взвешенных конечных преобразователей	206
6.1. Системы распознавания, основанные на СММ	206
6.1.1. Акустическая модель	210
6.1.2. Модель языка. N-грамм модели	220
6.1.3. Словарь произношений	230
6.1.4. Декодирование	231
6.2. Взвешенные преобразователи с конечным числом состояний	237
6.3. Декодирование на основе ВКП	251
Глава 7. Искусственные нейронные сети и интегральные (end-to-end) системы распознавания речи	263
7.1. Искусственный нейрон	264
7.2. Геометрический смысл функционирования нейрона	268
7.3. Нейросеть прямого распространения	269
7.3.1. Прямое распространение сигнала	270
7.3.2. Обучение нейросети	271
7.4. Нейросеть с радиальными базисными функциями	274
7.5. Рекуррентные нейросети	277
7.5.1. Архитектура простейшей рекуррентной нейросети	277
7.5.2. Метод обратного распространения ошибки во времени	279
7.5.3. Долгая краткосрочная память (LSTM)	285
7.5.4. Управляемый рекуррентный блок (GRU)	288
7.5.5. Двухнаправленные рекуррентные нейросети	289
7.5.6. Глубокие рекуррентные нейросети	290
7.6. Сверточные нейросети	291
7.7. Коннекционная временная классификация	296
7.7.1. Алгоритм коннекционной временной классификации	296
7.7.2. Модель Deep Speech	309
7.7.3. Deep Speech 2	311
7.8. Рекуррентные трансдюсеры (RNN-T)	313
7.9. Модель sequence-to-sequence	317
7.10. Модели на основе механизма внимания	320
7.11. Трансформеры	326
Список использованной литературы	339