



# پردازش گفتار

مبانی نظری، الگوریتم‌ها و توسعه محصول  
جلد اول

نویسندگان:

زودانگ هوانگ

آلکس آسرو

شاو-وون هان

(از واحد پژوهش مایکروسافت)

مترجمان:

دکتر هادی ویسی

(هیات علمی دانشگاه تهران)

کبری مفاخری

ناشر کتب مهندسی، کامپیوتر، دانشگاهی، مدیریت

نص

# فهرست

|         |   |
|---------|---|
| ۳۴..... | درک گفتار ۳-۱-۲                           |
| ۴۱..... | آواشناسی و واج شناسی ۲-۲                  |
| ۴۱..... | واج ها ۱-۲-۲                              |
| ۵۳..... | واج گونه: صدا و بافت ۲-۲-۲                |
| ۵۵..... | سرعت گفتار و هم تولیدی ۳-۲-۲              |
| ۵۶..... | هجاءها و واژه ها ۳-۲                      |
| ۵۷..... | هجاءها (Syllables) ۱-۳-۲                  |
| ۵۸..... | واژه ها ۲-۳-۲                             |
| ۶۳..... | نحو (Syntax) و معنی شناسی (Semantics) ۴-۲ |
| ۶۳..... | سازه های نحوی ۱-۴-۲                       |
| ۶۹..... | نقش های معنایی ۲-۴-۲                      |
| ۷۱..... | معنی شناسی واژگانی ۳-۴-۲                  |
| ۷۳..... | صورت منطقی ۴-۴-۲                          |
| ۷۵..... | چشم انداز تاریخی و مطالعه بیشتر ۵-۲       |
| ۷۷..... | منابع                                     |

## فصل ۳ احتمال، آمار و نظریه اطلاعات ۲۹

|         |  |
|---------|--|
| ۸۰..... | نظریه احتمال ۱-۳                             |
| ۸۰..... | احتمال شرطی و قانون بیز ۱-۱-۳                |
| ۸۲..... | متغیرهای تصادفی ۲-۱-۳                        |
| ۸۴..... | میانگین و واریانس ۳-۱-۳                      |
| ۸۷..... | کوواریانس و هم بستگی ۴-۱-۳                   |
| ۸۸..... | بردار های تصادفی و توزیع های چندمتغیره ۵-۱-۳ |
| ۹۰..... | چند توزیع مفید ۶-۱-۳                         |
| ۵۹..... | توزیع گاوسی ۷-۱-۳                            |

|        |           |
|--------|-----------|
| ۵..... | دیباچه    |
| ۷..... | سخن مترجم |

## فصل ۱ مقدمه ۹

|         |   |
|---------|---|
| ۱۰..... | انگیزه ۱-۱                                      |
| ۱۰..... | رابط زبان گفتاری ۱-۱-۱                          |
| ۱۱..... | ترجمه گفتار به گفتار ۲-۱-۱                      |
| ۱۱..... | همکاران دانش ۳-۱-۱                              |
| ۱۲..... | معماری سیستم های زبان گفتاری ۲-۱                |
| ۱۲..... | بازشناسی خودکار گفتار ۱-۲-۱                     |
| ۱۴..... | سنتز گفتار ۲-۲-۱                                |
| ۱۵..... | درک زبان گفتاری ۳-۲-۱                           |
| ۱۶..... | سازمان دهی کتاب ۳-۱                             |
| ۱۷..... | بخش اول- اصول نظری ۱-۳-۱                        |
| ۱۷..... | بخش دوم- پردازش گفتار ۲-۳-۱                     |
| ۱۷..... | بخش سوم- بازشناسی گفتار ۳-۳-۱                   |
| ۱۸..... | بخش چهارم- سیستم های گفتار سازی ۴-۳-۱           |
| ۱۸..... | بخش پنجم- سیستم های زبان گفتاری ۵-۳-۱           |
| ۱۸..... | مخاطبان کتاب ۴-۱                                |
| ۱۹..... | چشم انداز تاریخی و منابعی برای مطالعه بیشتر ۵-۱ |
| ۲۲..... | منابع   |

## فصل ۲ ساختار زبان گفتاری ۲۵

|         |                                 |
|---------|---------------------------------|
| ۲۶..... | سیستم های صدا و گفتار انسان ۱-۲ |
| ۲۶..... | صدا ۱-۱-۲                       |
| ۲۹..... | تولید گفتار ۲-۱-۲               |

۱۷۶ ..... ۲-۵-۴ معیار تفکیک کننده  
 ۱۷۹ ..... ۳-۵-۴ رشد دادن درخت  
 ۱۷۹ ..... ۴-۵-۴ مقادیر گم شده و حل تعارض  
 ۱۸۰ ..... ۵-۵-۴ سؤالات پیچیده  
 ۱۸۲ ..... ۶-۵-۴ اندازه درست برای درخت  
 ۱۸۷ ..... ۶-۴ چشم انداز تاریخی و مطالعه بیشتر  
 ۱۸۹ ..... منابع

### فصل ۵ پردازش سیگنال دیجیتال ... ۱۹۵

۱۹۶ ..... ۱-۵ سیگنال ها و سیستم های دیجیتال  
 ۱۹۷ ..... ۱-۱-۵ سیگنال های سینوسی  
 ۱۹۹ ..... ۲-۱-۵ سایر سیگنال های دیجیتال  
 ۱۹۹ ..... ۳-۱-۵ سیستم های دیجیتال  
 ۲۰۲ ..... ۲-۵ تبدیل های بسامد-پیوسته  
 ۲۰۲ ..... ۱-۲-۵ تبدیل فوریه  
 ۲۰۴ ..... ۲-۲-۵ تبدیل Z  
 ۲۰۵ ..... ۳-۲-۵ تبدیل های Z تابع های مقدماتی  
 ۲۰۷ ..... ۴-۲-۵ ویژگی های تبدیل Z و تبدیل فوریه  
 ۲۰۹ ..... ۳-۵ تبدیل های بسامد گسسته  
 ۲۱۰ ..... ۱-۳-۵ تبدیل فوریه گسسته (DFT)  
 ۲۱۱ ..... ۲-۳-۵ تبدیل فوریه سیگنال های دوره ای  
 ۲۱۴ ..... ۳-۳-۵ تبدیل فوریه سریع (FFT)  
 ۲۱۴ ..... ۱-۳-۳-۵ تبدیل فوریه سریع پایه ۲  
 ۲۱۸ ..... ۴-۳-۵ پیچش دایره ای  
 ۲۱۹ ..... ۵-۳-۵ تبدیل کسینوس گسسته (DCT)  
 ۲۲۰ ..... ۴-۵ فیلترها دیجیتال و پنجره ها  
 ۲۲۰ ..... ۱-۴-۵ فیلتر پایین گذر ایده آل  
 ۲۲۱ ..... ۲-۴-۵ تابع های پنجره  
 ۲۲۳ ..... ۳-۴-۵ فیلترهای FIR  
 ۲۲۷ ..... ۴-۴-۵ فیلترهای IIR  
 ۲۳۰ ..... ۵-۵ پردازش دیجیتال سیگنال های آنالوگ  
 ۲۳۱ ..... ۱-۵-۵ تبدیل فوریه سیگنال های آنالوگ  
 ۲۳۱ ..... ۲-۵-۵ قضیه نمونه برداری  
 ۲۳۳ ..... ۳-۵-۵ تبدیل آنالوگ به دیجیتال  
 ۲۳۴ ..... ۴-۵-۵ تبدیل دیجیتال به آنالوگ

۲-۳ نظریه تخمین ..... ۱۰۱  
 ۱-۲-۳ تخمین کمینه میانگین مربعات خطا ..... ۱۰۲  
 ۲-۲-۳ تخمین درست نمایی بیشینه ..... ۱۰۶  
 ۳-۲-۳ تخمین بیز و تخمین بیشینه احتمال پسین ..... ۱۱۰  
 ۳-۳ آزمون معنی داری ..... ۱۱۵  
 ۱-۳-۳ سطح معنی داری ..... ۱۱۶  
 ۲-۳-۳ آزمون نرمال (آزمون Z) ..... ۱۱۷  
 ۳-۳-۳ آزمون نیکویی برازش  $\chi^2$  ..... ۱۱۸  
 ۴-۳-۳ آزمون جفت های سازگار ..... ۱۲۱  
 ۴-۳ نظریه اطلاعات ..... ۱۲۲  
 ۱-۴-۳ آنتروپی ..... ۱۲۳  
 ۲-۴-۳ آنتروپی شرطی ..... ۱۲۵  
 ۳-۴-۳ قضیه رمزگذاری منبع ..... ۱۲۶  
 ۴-۴-۳ اطلاعات متقابل و رمزگذاری کانال ..... ۱۲۸  
 ۵-۳ چشم انداز تاریخی و مطالعه بیشتر ..... ۱۳۱  
 ۱۳۳ ..... منابع

### فصل ۴ بازشناسی الگو ..... ۱۳۵

۱-۴ نظریه تصمیم گیری بیز ..... ۱۳۶  
 ۱-۱-۴ قوانین تصمیم گیری با کمترین میزان خطا ... ۱۳۷  
 ۲-۱-۴ تابع های تمایز ..... ۱۴۰  
 ۲-۴ چگونگی ساختن دسته بندی کننده ها ..... ۱۴۱  
 ۱-۲-۴ دسته بندی کننده های گاوسی ..... ۱۴۳  
 ۲-۲-۴ افزایش تعداد ویژگی ها ..... ۱۴۳  
 ۳-۲-۴ تخمین زدن نرخ خطا ..... ۱۴۷  
 ۴-۲-۴ مقایسه دسته بندی کننده ها ..... ۱۴۹  
 ۳-۴ آموزش تمایزی ..... ۱۵۰  
 ۱-۳-۴ تخمین اطلاعات متقابل بیشینه ..... ۱۵۱  
 ۲-۳-۴ تخمین نرخ خطای کمینه ..... ۱۵۶  
 ۳-۳-۴ شبکه های عصبی ..... ۱۵۸  
 ۴-۴ روش های تخمین بی ناظر ..... ۱۶۳  
 ۱-۴-۴ چندی سازی برداری ..... ۱۶۳  
 ۲-۴-۴ الگوریتم EM ..... ۱۶۹  
 ۳-۴-۴ تخمین چگالی تلفیق گاوسی چندمتغیری ..... ۱۷۱  
 ۵-۴ درخت های دسته بندی و رگرسیون ..... ۱۷۴  
 ۱-۵-۴ انتخاب مجموعه سؤالات ..... ۱۷۶

۲۹۰..... ۳-۴-۶ کپستروم سیگنال های دوره ای

۲۹۱..... ۴-۴-۶ کپستروم سیگنال های گفتار

۲۹۲..... ۵-۴-۶ جداسازی منبع-فیلتر از طریق کپستروم

۲۹۳..... ۵-۶ نمایش های مبتنی بر ادراک

۲۹۳..... ۱-۵-۶ تبدیل دوخطی

۲۹۴..... ۲-۵-۶ کپستروم مل-بسامد

۲۹۶..... ۳-۵-۶ پیش بینی خطی ادراکی (PLP)

۲۹۷..... ۶-۶ بسامدهای فورمنت

۲۹۸..... ۱-۶-۶ ردیابی آماری فورمنت

۳۰۱..... ۷-۶ نقش زیرویمی

۳۰۲..... ۱-۷-۶ روش خودهمبستگی

۳۰۴..... ۲-۷-۶ روش هم بستگی متقاطع نرمال شده

۳۰۷..... ۳-۷-۶ مشروط سازی سیگنال

۳۰۷..... ۴-۷-۶ ردیابی زیرویمی

۳۰۹..... ۷-۷ چشم انداز تاریخی و مطالعه بیشتر

۳۱۱..... منبع

## فصل ۷ رمزگذاری و فشرده سازی گفتار ۳۱۵

۳۱۵..... ۱-۷ ویژگی های رمزگذارهای گفتار

۳۱۷..... ۲-۷ رمزگذارهای شکل موج عددی

۳۱۸..... ۱-۲-۷ مدوله سازی رمز پالس (PCM) خطی

۳۲۰..... ۲-۲-۷ PCM قانون  $M$  و قانون  $A$

۳۲۱..... ۳-۲-۷ PCM تطبیقی

۳۲۳..... ۴-۲-۷ چندی سازی تفاضلی

۳۲۳..... ۳-۷ رمزگذارهای حوزه ی بسامد عددی

۳۲۶..... ۱-۳-۷ مزایای پوشاندن

۳۲۸..... ۲-۳-۷ رمزگذارهای تبدیل

۳۲۸..... ۳-۳-۷ صدا برای مصرف کننده ها

۳۲۹..... ۴-۳-۷ پخش صدای دیجیتال (DAB)

۳۳۰..... ۴-۷ پیش بینی خطی تحریکی رمز (CELP)

۳۳۰..... ۱-۴-۷ سیستم رمزکننده صدای LPC

۳۳۱..... ۲-۴-۷ تحلیل به وسیله ترکیب

۳۳۴..... ۳-۴-۷ پیش بینی زیرویمی: کتاب رمز تطبیقی

۳۳۴..... ۴-۴-۷ وزن دهی ادراکی و فیلتر کردن پسین

۳۳۵..... ۵-۴-۷ چندی سازی پارامتر

۲۳۵..... ۶-۵ پردازش سیگنال چندنرخ

۲۳۶..... ۱-۶-۵ کاهش

۲۳۷..... ۲-۶-۵ درون یابی

۲۳۷..... ۳-۶-۵ نمونه برداری مجدد

۲۳۸..... ۷-۵ بانک های فیلتر

۲۳۸..... ۱-۷-۵ فیلترهای تریبلی مزدوج دو بانندی

۲۴۱..... ۲-۷-۵ بانک فیلترهای چندوضوحی

۲۴۲..... ۳-۷-۵ FFT به عنوان بانک فیلتر

۲۴۴..... ۴-۷-۵ تبدیل های هم پوشان مدوله شده

۲۴۶..... ۸-۵ فرآیندهای تصادفی

۲۴۷..... ۱-۸-۵ آمار فرآیندهای تصادفی

۲۴۹..... ۲-۸-۵ فرآیندهای ایستا

۲۵۲..... ۳-۸-۵ سیستم های LTI با ورودی های تصادفی

۲۵۲..... ۴-۸-۵ چگالی توان طیف

۲۵۴..... ۵-۸-۵ نويز

۲۵۵..... ۹-۵ چشم انداز تاریخی و مطالعه بیشتر

۲۵۷..... منبع

## فصل ۶ نمایش های سیگنال گفتار ۲۵۹

۲۶۰..... ۱-۶ تحلیل فوریه کوتاه مدت

۲۶۵..... ۱-۱-۶ طیف نگارها

۲۶۶..... ۲-۱-۶ تحلیل زیرویمی هم زمان

۲۶۶..... ۲-۶ مدل آکوستیکی تولید گفتار

۲۶۷..... ۱-۲-۶ تحریک چاکنایی

۲۶۷..... ۲-۲-۶ الحاق لوله بی اتلاف

۲۷۱..... ۳-۲-۶ مدل های منبع-فیلتر تولید گفتار

۲۷۳..... ۳-۶ رمزگذاری پیش بینی کننده خطی (LPC)

۲۷۳..... ۱-۳-۶ اصل تعامد

۲۷۵..... ۲-۳-۶ حل معادلات LPC

۲۸۱..... ۳-۳-۶ تحلیل طیفی از طریق LPC

۲۸۲..... ۴-۳-۶ خطای پیش بینی

۲۸۴..... ۵-۳-۶ نمایش های معادل

۲۸۷..... ۴-۶ پردازش کپستروال

۲۸۸..... ۱-۴-۶ کپستروم حقیقی و مختلط

۲۸۸..... ۲-۴-۶ کپستروم فیلترهای قطب-صفر



۳۸۸ ..... تنوع گوینده ۳-۱-۹

۳۸۸ ..... تنوع محیط ۴-۱-۹

۳۸۹ ..... چگونگی اندازه گیری خطاهای بازشناسی گفتار ۲-۹

۳۹۱ ..... پردازش سیگنال: استخراج ویژگی ها ۳-۹

۳۹۱ ..... دریافت سیگنال ۱-۳-۹

۳۹۲ ..... شناسایی پایان سیگنال گفتار ۲-۳-۹

۳۹۴ ..... MFCC و ویژگی های پویای آن ۳-۳-۹

۳۹۶ ..... تغییر ویژگی ۴-۳-۹

۳۹۷ ..... مدل سازی آوایی: انتخاب واحدهای مناسب ۴-۹

۳۹۸ ..... مقایسه واحدهای مختلف ۱-۴-۹

۳۹۹ ..... وابستگی به بافت ۲-۴-۹

۴۰۱ ..... واحدهای صوتی-آوایی خوشه بندی شده ۳-۴-۹

۴۰۶ ..... صورت های پایه واژگانی ۴-۴-۹

۴۰۹ ..... مدل سازی صوتی: امتیازدهی ویژگی های ۵-۹

۴۰۹ ..... انتخاب توزیع های خروجی HMM ۱-۵-۹

۴۱۱ ..... آموزش گفتار مجزا در مقابل پیوسته ۲-۵-۹

۴۱۳ ..... روش های تطبیق: کمینه کردن ناسازگاری ها ۶-۹

۴۱۴ ..... بیشینه احتمال پسین (MAP) ۱-۶-۹

۴۱۷ ..... رگرسیون خطی درست نمایی بیشینه ۲-۶-۹

۴۲۰ ..... مقایسه MLLR و MAP ۳-۶-۹

۴۲۱ ..... مدل های خوشه بندی شده ۴-۶-۹

۴۲۲ ..... معیار اطمینان: اندازه گیری قابلیت اعتماد ۷-۹

۴۲۲ ..... مدل های مکمل ۱-۷-۹

۴۲۳ ..... مدل های تبدیل ۲-۷-۹

۴۲۵ ..... مدل های ترکیبی ۳-۷-۹

۴۲۶ ..... سایر روش ها ۸-۹

۴۲۶ ..... شبکه های عصبی ۱-۸-۹

۴۲۸ ..... مدل های قطعه ۲-۸-۹

۴۳۳ ..... مطالعه موردی: سیستم بازشناسی گفتار و ۹-۹

۴۳۴ ..... چشم انداز تاریخی و مطالعه بیشتر ۱۰-۹

۴۳۷ ..... منابع

۳۳۶ ..... استانداردهای CELP ۶-۴-۷

۳۳۹ ..... رمزگذارهای گفتار با نرخ بیت پایین ۵-۷

۳۴۰ ..... رمزکننده صدای LPC تحریک مخلوط ۱-۵-۷

۳۴۱ ..... رمزگذاری هارمونیک ۲-۵-۷

۳۴۵ ..... درون یابی شکل موج ۳-۵-۷

۳۴۸ ..... چشم انداز تاریخی و مطالعه بیشتر ۶-۷

۳۴۹ ..... منابع

### فصل ۸ مدل مخفی مارکوف ..... ۳۵۳

۳۵۳ ..... زنجیره مارکوف ۱-۸

۳۵۵ ..... تعریف مدل مخفی مارکوف ۲-۸

۳۵۸ ..... برنامه نویسی پویا و پیچش زمانی پویا ۱-۲-۸

۳۶۰ ..... چگونه یک HMM را ارزیابی کنیم؟ ۲-۲-۸

۳۶۲ ..... چگونگی رمزگشایی HMM ۳-۲-۸

۳۶۴ ..... چگونگی تخمین زدن پارامترهای HMM ۴-۲-۸

۳۶۷ ..... HMM های پیوسته و نیمه پیوسته ۳-۸

۳۶۳ ..... HMM های با مخلوط چگالی پیوسته ۱-۳-۸

۳۶۹ ..... HMM های نیمه پیوسته ۲-۳-۸

۳۷۱ ..... مسائل عملی در استفاده از HMM ها ۴-۸

۳۷۱ ..... برآوردهای اولیه ۱-۴-۸

۳۷۲ ..... توپولوژی (ساختار) مدل ۲-۴-۸

۳۷۳ ..... معیارهای آموزش ۳-۴-۸

۳۷۴ ..... درون یابی حذف شده ۴-۴-۸

۳۷۵ ..... هموارسازی پارامتر ۵-۴-۸

۳۷۶ ..... نمایش مقادیر احتمال ۶-۴-۸

۳۷۸ ..... محدودیت های HMM ۵-۸

۳۷۸ ..... مدل سازی دیرش ۱-۵-۸

۳۸۰ ..... فرض مرتبه اول ۲-۵-۸

۳۸۰ ..... فرض استقلال شرطی ۳-۵-۸

۳۸۱ ..... چشم انداز تاریخی و مطالعه بیشتر ۶-۸

۳۸۳ ..... منابع

### فصل ۹ مدل سازی صوتی ..... ۳۸۵

۳۸۶ ..... تنوع در سیگنال گفتار ۱-۹

۳۸۶ ..... تنوع در بافت ۱-۱-۹

۳۸۷ ..... تنوع سبک ۲-۱-۹