

ADI 处理器实用丛书

语音信号处理 及 Blackfin DSP 实现

◎ 张雄伟 邹霞 贾冲 李莉 编著



本书特色：

- DSP 定点数值运算与处理方法的系统阐述
- 语音处理基本理论与实际实现的有机结合
- 语音处理算法 Blackfin DSP 实现的实例展示



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

ADI 处理器实用丛书

语音信号处理及 Blackfin DSP 实现

张雄伟 邹霞 贾冲 李莉 编著

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书基于 ADI 公司的高性能、低功耗 Blackfin DSP, 系统地介绍了语音信号处理的基础、原理、方法与 DSP 实现。本书首先对语音处理及 DSP 语音处理系统进行了概述, 介绍了语音信号处理和 DSP 运算的基础知识; 接着介绍了 Blackfin DSP 芯片的硬件结构和指令系统, 并重点介绍了 Visual DSP++集成开发环境; 然后介绍了语音信号的时域分析、变换域分析及线性预测分析, 并基于 Blackfin DSP 实现了有关算法; 最后介绍了一个基于 Blackfin DSP 的语音处理系统的软、硬件设计方法, 并以 ADSP-BF533 为例给出了应用实例。

本书的目的是使读者了解语音信号处理及 Blackfin DSP 芯片的基本原理和方法, 初步掌握语音处理算法用 DSP 实现的软、硬件设计 and 应用系统开发方法, 具备从事语音信号处理算法与应用系统开发的初步能力。

本书内容广泛, 重点突出, 原理阐述深入浅出, 注重理论与实际应用的结合, 实用性强。本书可作为高等院校通信工程、电子工程、信息工程等专业高年级本科生和信号与信息处理、通信与信息系统等相关专业的本科生和大专生的参考教材, 同时也可供语音处理和信息技术研究开发的科研及工程人员参考, 还可作为基于 DSP 芯片进行语音处理系统应用开发人员的培训教材。

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有, 侵权必究。

图书在版编目 (CIP) 数据

语音信号处理及 Blackfin DSP 实现 / 张雄伟等编著. —北京: 电子工业出版社, 2011.10

(ADI 处理器实用丛书)

ISBN 978-7-121-14640-4

I. ①语… II. ①张… III. ①语声信号处理 ②数字信号处理 IV. ①TN912.3 ②TN911.72

中国版本图书馆 CIP 数据核字 (2011) 第 191382 号

策划编辑: 竺南直 徐蔷薇

责任编辑: 谭丽莎

印 刷: 北京市李史山胶印厂
装 订:

出版发行: 电子工业出版社
北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1 092 1/16 印张: 16.75 字数: 397 千字
印 次: 2011 年 10 月第 1 次印刷
印 数: 4 000 册 定价: 36.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

语音是人类相互之间进行交流和通信最方便和最快捷的一种手段,如何高效地实现语音处理、传输、存储或通过语音实现人机交互,是语音信号处理领域中的重要研究课题。近20多年来,语音编码、语音合成、语音识别等语音处理领域的重要分支相继取得了一系列重大进展,研究成果不断涌现;与此同时,数字信号处理(DSP)芯片性能的不不断提高和价格的不断下降,为高性能语音处理算法走向实用提供了可能。

DSP芯片是一类特别适合于进行数字信号处理运算的微处理器,可实时快速实现各种数字信号处理算法。近30年来,DSP芯片发展十分迅速,性能价格比不断提高,已广泛应用于通信、信号处理、自动控制等领域,成为一种十分重要的电子产品的核心芯片。ADI公司的Blackfin系列DSP是一类特别适用于通信、信号处理等领域的DSP芯片。

目前,兼顾介绍语音信号处理与BlackfinDSP实现的参考书籍还不多,基于此,我们编写了本书。本书是作者以近几年来为本科生和研究生讲授语音信号处理及DSP实现相关课程的讲义为基础,结合多年从事语音信号处理及DSP开发的经验,参考相关著作和文献资料编写而成的。

全书共分10章。第1章介绍了语音处理及DSP语音处理系统;第2章介绍了语音处理的基础知识;第3章介绍了BlackfinDSP的数值运算基础;第4章介绍了BlackfinDSP芯片的硬件结构;第5章介绍了BlackfinDSP的指令系统;第6章介绍了BlackfinDSP集成开发调试环境VisualDSP++;第7章介绍了语音信号的时域分析方法并对相关算法进行了BlackfinDSP实时实现;第8章介绍了语音信号的变换域分析方法并对相关算法进行了BlackfinDSP实时实现;第9章介绍了语音信号的线性预测分析方法并对相关算法进行了BlackfinDSP实时实现;第10章介绍了基于BlackfinDSP的语音处理系统的软、硬件设计方法,并以ADSP-BF533为例给出了应用实例,介绍了基于DSP的语音处理系统的设计与调试过程。

本书由张雄伟教授等编著。张雄伟编写了第1、2、3章并对全书进行了统稿,邹霞编写了第7、8章及第9、10章的部分内容,贾冲编写了第4章及第10章的部分内容,李莉编写了第5、6章,陈亮编写了第9章的部分内容,曹铁勇、杨吉斌、吴其前为本书提出了很好的建议,研究生吴海佳、隋璐瑛、赵改华、潘竟峰为本书做了许多具体工作,ADI公司的高威先生和北京亿旗创新科技发展公司提供了有关资料和BlackfinDSP开发板,在此一并向他们表示衷心的感谢。

由于作者水平所限,书中难免存在错漏之处,恳请广大读者给予批评指正。

作 者

2011年5月于解放军理工大学(南京)

目 录

第 1 章 语音处理与 DSP 语音处理系统	1
1.1 引言	1
1.2 语音处理的应用	2
1.3 DSP 语音处理系统	6
1.3.1 DSP 语音处理系统的构成	7
1.3.2 DSP 语音处理系统的特点	7
1.3.3 DSP 语音处理系统的设计	8
1.3.4 DSP 语音处理系统的开发	9
1.4 DSP 芯片	10
1.4.1 DSP 芯片简介	10
1.4.2 ADI 公司的 DSP 芯片	11
习题与思考题	16
第 2 章 语音处理的基础知识	17
2.1 引言	17
2.2 语音的波形及特性	17
2.3 语音的产生	19
2.3.1 发声器官	19
2.3.2 语音的分类	20
2.3.3 基音频率与共振峰	21
2.3.4 语谱图	22
2.4 汉语语音的基本特性	23
2.4.1 声母和韵母	23
2.4.2 元音和辅音	24
2.4.3 汉语的声调	24
2.5 语音信号的产生模型 ^[2, 3]	25
2.6 听觉系统和听觉特性	27
2.6.1 听觉系统	27
2.6.2 听觉特性	28
习题与思考题	30

第3章 Blackfin DSP 的数值运算基础	31
3.1 引言	31
3.2 定点的基本概念	31
3.2.1 数的定标	31
3.2.2 数的转换	33
3.2.3 溢出保护模式	34
3.2.4 符号扩展模式	35
3.2.5 舍入及截尾	35
3.3 DSP 定点运算的模拟实现	36
3.3.1 加法/减法运算的 C 语言定点模拟	36
3.3.2 乘法运算的 C 语言定点模拟	38
3.3.3 除法运算的 C 语言定点模拟	39
3.3.4 程序变量的 Q 值确定	40
3.3.5 浮点至定点变换的 C 程序举例	41
3.4 DSP 定点运算实现的基本原理	43
3.4.1 定点乘法	43
3.4.2 定点加法	44
3.4.3 定点除法	45
3.5 非线性运算的定点实现方法	48
3.5.1 级数展开法	48
3.5.2 查表法	48
3.5.3 混合法	50
习题与思考题	52
第4章 Blackfin DSP 的硬件结构	53
4.1 引言	53
4.2 Blackfin DSP 的资源配置及引脚定义	54
4.2.1 资源配置	54
4.2.2 引脚定义	58
4.3 Blackfin DSP 的内核结构	61
4.3.1 概述	61
4.3.2 寄存器组	61
4.3.3 算术逻辑单元 (ALU)	62
4.3.4 乘法器/累加器	62

4.3.5	桶形移位器	62
4.4	Blackfin DSP 的存储空间	62
4.4.1	片内存储器	64
4.4.2	片外存储器	64
4.4.3	I/O 存储器	65
4.5	Blackfin DSP 的事件处理	65
4.5.1	概述	65
4.5.2	系统中断过程	67
4.5.3	事件控制寄存器	69
4.5.4	事件处理寄存器	70
4.6	Blackfin DSP 的 DMA 传输	71
4.6.1	概述	71
4.6.2	基于描述符的 DMA 传输	71
4.6.3	基于寄存器的 DMA 传输	72
4.6.4	存储器 DMA (MDMA)	73
4.7	Blackfin DSP 的典型片内外设	73
4.7.1	串行外设接口 (SPI)	74
4.7.2	同步串行接口 (SPORT)	74
4.7.3	异步串行接口 (UART)	75
4.7.4	并行外部接口 (PPI)	76
4.7.5	通用可编程 I/O 引脚 PF	77
4.7.6	实时时钟 (RTC)	77
4.7.7	定时器	78
4.8	Blackfin DSP 的时钟	78
4.9	Blackfin DSP 的电源管理	80
4.9.1	动态电源管理控制器	80
4.9.2	运行方式	80
4.9.3	外设时钟	81
4.9.4	动态电源电压控制	82
4.10	Blackfin DSP 的内核模式	82
4.10.1	概述	82
4.10.2	用户模式	83
4.10.3	监控模式	83
4.10.4	仿真模式	83

4.10.5 空闲状态	84
4.10.6 复位状态	84
4.11 Blackfin DSP 的安全特性	84
4.12 Blackfin DSP 的引导模式	85
习题与思考题	86
第 5 章 Blackfin DSP 的指令系统	87
5.1 引言	87
5.2 数据寻址方式	87
5.3 Blackfin DSP 的指令约定	88
5.4 Blackfin DSP 指令集	92
5.4.1 程序流程控制指令	92
5.4.2 加载/存储 (LOAD/STORE) 指令	95
5.4.3 数据转移指令	102
5.4.4 栈控制指令	104
5.4.5 控制编码 (CC) 位管理指令	107
5.4.6 逻辑运算指令	110
5.4.7 位操作指令	112
5.4.8 移位/循环指令	115
5.4.9 算术运算指令	119
5.4.10 外部事件操作 (External Event Management) 指令	129
5.4.11 缓存 (Cache) 控制指令	133
5.4.12 视频像素操作指令	134
5.4.13 向量操作指令	139
5.4.14 并行指令	144
习题与思考题	147
第 6 章 Blackfin DSP 集成开发调试环境 VisualDSP++	148
6.1 引言	148
6.2 VisualDSP++ 软件的安装	148
6.3 建立仿真环境	150
6.4 C/C++ 编译器	153
6.4.1 C/C++ 编译器简介	153
6.4.2 数据类型	153
6.4.3 在 VisualDSP++ 中配置编译器	154

6.5	汇编器	154
6.5.1	汇编程序结构和语法	154
6.5.2	预处理器	155
6.6	链接器	156
6.6.1	链接过程概述	156
6.6.2	链接文件	156
6.6.3	在 VisualDSP++ 中配置链接器	157
6.7	加载器	158
6.8	调试实例与调试工具使用	158
6.8.1	打开一个工程	159
6.8.2	编译链接工程文件	160
6.8.3	加载可执行文件	161
6.8.4	调试 (Debug) 工程	161
6.8.5	设置断点 (Breakpoint)	162
6.8.6	VisualDSP++ 调试工具的使用	163
	习题与思考题	169
第 7 章	语音信号的时域分析及 Blackfin DSP 实现	170
7.1	引言	170
7.2	语音短时分析技术	170
7.3	语音时域参数提取	172
7.3.1	短时能量和平均幅度	172
7.3.2	短时平均过零率	175
7.3.3	短时自相关分析	177
7.4	语音时域参数的 Blackfin DSP 实现	183
7.4.1	短时平均过零率的 Blackfin DSP 实现	183
7.4.2	短时能量的 Blackfin DSP 实现	183
7.4.3	短时自相关分析的 Blackfin DSP 实现	184
	习题与思考题	186
第 8 章	语音信号的变换域分析及 Blackfin DSP 实现	187
8.1	引言	187
8.2	FFT 算法的基本原理	187
8.2.1	DFT 的基本原理	187
8.2.2	FFT 算法的导出	188

8.3 语音信号的频域分析	193
8.3.1 短时傅里叶变换	193
8.3.2 短时傅里叶反变换	199
8.3.3 语谱图	202
8.3.4 频域分析应用——频域基音检测	203
8.4 FFT 算法的 VisualDSP++ 编程实现	206
习题与思考题	208
第 9 章 语音信号的线性预测分析及 Blackfin DSP 实现	209
9.1 引言	209
9.2 LP 分析的基本原理	209
9.3 LP 正则方程的自相关解法和自协方差解法	212
9.3.1 自相关解法	212
9.3.2 自协方差解法	213
9.4 自相关方程的杜宾递推算法	214
9.5 模型增益 G 的确定	218
9.6 LP 分析的 Blackfin DSP 实现	221
习题与思考题	230
第 10 章 语音处理系统的 Blackfin DSP 实现	231
10.1 引言	231
10.2 语音处理系统的硬件设计	231
10.2.1 DSP 硬件系统设计的一般步骤	231
10.2.2 语音处理系统的原理图	232
10.2.3 DSP 基本硬件设计	237
10.2.4 DSP PCB 布板基础	247
10.3 语音处理系统的软件设计	248
10.3.1 系统初始化	248
10.3.2 中断服务程序编写	251
10.3.3 主处理程序编写	252
10.3.4 存储单元分配	254
10.4 独立系统形成	256
习题与思考题	257
参考文献	258

第 1 章 语音处理与 DSP 语音处理系统

1.1 引言

语音是人类相互之间进行交流时使用最多、最自然、最基本也是最重要的信息载体。在高度信息化的今天，语音处理的一系列技术及其应用已经成为信息社会不可或缺的重要组成部分。

语音的产生是一个复杂的过程，包括心理和生理等方面的一系列动作。当人需要通过语音表达某种信息时，这种信息首先以某种抽象的形式表现在说话人的大脑里，然后转换为一组神经信号，这些神经信号作用于发声器官进而产生携带信息的语音信号。

语音信号处理的研究起源于对发声器官的模拟。1939 年，美国人 H.Dudley 展出了一个简单的发声过程模拟系统，以后发展为声道的数字模型。利用该模型可以对语音信号进行各种频谱及参数的分析，同时也可根据分析获得的频谱特征或参数变化规律，合成语音信号，实现机器的语音合成。

目前，对语音信号进行研究一般都基于语音信号的数字表示，因此，语音信号的数字表示是进行语音信号数字处理的基础。语音信号数字化的理论依据是我们熟知的采样定理，即只要采样频率足够高，就可以用时域上周期抽取的样点来表示一个带限信号。语音信号的离散表示基本上可以分为两大类：波形表示和参数表示，如图 1-1 所示。波形表示仅仅是通过采样和量化的过程保存模拟语音信号的“波形”，而参数表示则是把语音信号表示成某种语音产生模型的输出。为了得到参数表示，首先必须对语音进行采样和量化，然后再进一步处理以得到语音产生模型的参数。语音产生模型的参数一般可分为两大类：一类是激励参数；另一类是声道参数。

人们历来重视对语音信号和语音通信的研究。社会的进步对语音通信提出了更高的要求，需要更高的语音质量和更低的数码率，从而推动了语音编码技术的发展。而自动控制和计算机科学的发展又要求用语音沟通人与机器的信息交流，要求机器能听懂人说话和对人说话，甚至还要能辨别出是谁在说话，这又推动了语音识别、说话人识别和语音合成技术的研究，从而使语音处理技术得到迅速的发展。语音编码、语音识别、说话人识别、语音合成等技术的基础都是对语音信号特征的认识，都要利用数字信号处理的一些基本技术来分析和处理语音信号，而更深层次的发展涉及人的发音和听觉机理，与生理学、语言学甚至心理学有关。

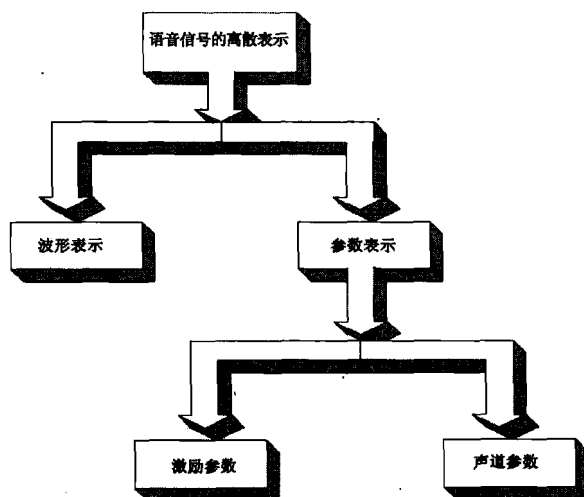


图 1-1 语音信号的表示方法

语音信号数字处理是一门涉及诸多学科的交叉学科，它以生理学、心理学、语言学及声学等学科为基础，以信息论、控制论、系统论的理论为指导，是通过应用信号处理、统计分析、模式识别等现代技术手段发展形成的一门综合性学科。20 世纪 80 年代以前，线性预测编码技术（LPC）是语音信号数字处理研究领域最重要的研究成果。20 世纪 80 年代以后，分析合成技术、矢量量化技术、隐马尔可夫模型（HMM）等极大地推动了语音编码、语音识别技术的发展。20 世纪 90 年代以后，神经网络、小波分析、分形及混沌等新技术在语音处理领域的应用将语音处理研究提高到了一个新的水平。

1.2 语音处理的应用

语音处理的应用非常广泛，最基本的应用就是语音的数字传输，即对语音进行数字化后在数字通信系统中进行传输，以实现数字语音通信。图 1-2 中列出了语音处理的一些典型应用。

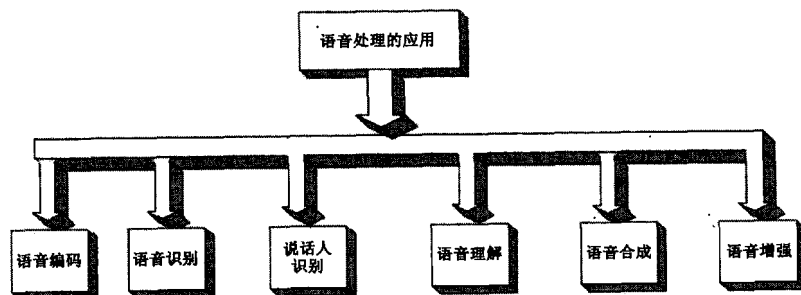


图 1-2 语音处理的一些典型应用

下面简要介绍一下这些应用。

1. 语音编码

语音编码是语音数字处理最重要的一种应用。语音编码的目的是用尽可能低的数码率来获得尽可能高的合成语音质量。实现语音编码（特别是中低速率）的设备通常称为声码器。虽然光和微波通信等系统可以提供很宽的频带，但在很多情况下仍然需要压缩编码速率以节省频带。一方面，编码后可以在有限带宽的信道上传输多路语音，提高信道的利用率；另一方面，可以在窄带的模拟信道（如短波）上传输数字语音。特别是在军事通信系统等需要复杂加密的应用场合，声码器具有不可替代的作用。此外，语音的数字存储、语音应答等也是语音编码的重要应用。在语音编码技术中，线性预测、矢量量化、码本激励等是最重要的几种实现技术。

根据语音编码的采样率，可以分为窄带（4kHz）语音编码、宽带（7kHz）语音编码和20kHz的音乐带宽编码。窄带语音编码的采样率通常为8kHz，一般应用于语音通信中；宽带语音编码的采样率通常为16kHz，一般用于要求更高音质的应用中，如会议电视；而20kHz的音乐宽带编码主要适用于音乐数字化，其采样频率高达44.1kHz。

近几十年来，语音编码技术发展非常迅速。以窄带语音编码为例，自20世纪70年代国际上推出64kbit/s PCM语音编码标准（G.711）以来，已相继有32kbit/s ADPCM（G.726）、16kbit/s LD-CELP（G.728）、8kbit/s CS-ACELP（G.729）等国际标准推出。而地区性或行业性的标准也有不少，如第二代移动通信系统中的语音编码（13kbit/s GSM），美国国防部的2.4kbit/s 保密电话标准（LPC10/10e、MELP/MELPe）等。

目前，在2.4kbit/s以上的编码速率，质量已得到人们的认可，并已广泛应用。未来的研究重点是突破2.4kbit/s以下极低速率的语音编码技术和算法。

2. 语音识别

语音识别的作用是将语音转换成等价的书面信息，也就是让计算机听懂人说话。目前语音识别已经成为语音信号数字处理研究领域中的重点和难点技术。语音识别可以有許多分类方法，例如，根据语音识别对象来划分，可以分为孤立词识别、连续语音识别等；根据词汇量来划分，可以分为小词汇量（50个词以下）语音识别、中词汇量（50~500个词）语音识别、大词汇量（500个词以上）和超大限词汇量（几万个词）语音识别等；根据对说话人的要求来划分，可以分为特定说话人（speaker dependent）语音识别、多说话人语音识别和非特定说话人（speaker independent）语音识别等。

语音识别虽然从原理上看实现并不困难，但在实际实现时遇到的困难很多。例如，发音的多变性、不同人发同一个音、同一个人不同的条件下发同一个音等都会有不同的发音参数；发音的模糊性，即在实际的连续语音流中语音声学变量与音素变量之间不存在一对对应关系；语音流中变化多端的音变现象，这些音变对人类的听觉系统来说很容易辨认，

但机器识别却很不容易。

语音识别的应用很广，如声控打字、用声音控制计算机等。如果将语音识别与语音合成结合起来，就可以实现极低比特率的语音通信系统。

目前，一些中小词汇表的孤立词或连续语音识别系统已进入市场。目前研究的重点是实现大词汇表、非特定人的连续语音识别系统，它可用于人机直接对话、语音打字机及两种语音之间的直接通信等一系列重要场合。

3. 说话人识别

说话人识别的作用是根据语音辨别说话人，广义的语音识别也包括说话人识别。但说话人识别并不注意语音信号中的语义内容，而是希望从语音信号中提取出人的特征，即根据语音判别是谁讲的话。语音信号既载有说话人的语言信息，同时也载有说话人本身的特征信息。每个人的发音器官都有自己的特征，说话时也都有自己的特殊语言习惯。在分析语音信号时，可以提取说话人的个人特征，从而有可能识别说话人是谁。在研究语音识别时，要消除说话人的个人特征，以免影响识别的准确率；而在研究说话人识别时，则要专门研究人的特征，从众多的语音信号中分析和提取这些特征，去除不含个人特征的语音信息。

说话人识别有两类：说话人确认；说话人辨认。前者是指确认说话人的身份，说话人说一句或几句测试语句，经处理后获取的特征参数与储存的特定人语音的参数比较，做出“是与否”的判决。后者是要辨认待识别的语音来自若干人中的哪一位，要将待识别语音与每一位的语音进行比较，找出距离最近的语音所对应的说话人。从语音信号处理的角度来看，二者基本上是相同的，都需要确定选用的参数和计算距离的准则。前者需确定“是与否”的门限，后者需与待识别语音比较它们各自的距离。比较的方法大致与识别语音的方法相同，也要进行端点检测和时间规整。参数的选择原则，一是要能反映说话人的个性，二是要兼顾识别率和复杂程度。比较简单的特征参数是基音和能量，也可以用 LPC 参数与共振峰，但其计算量稍大。也有用语谱图来识别的，称为“声纹”。

然而语音是动态的，它和说话人所处的环境、情绪和身体状况关系很大。一个人在不同时间、不同情况下讲同一句话，差异也不一定比不同人小，不像“指纹”是静态的、绝对的。在现阶段还需结合识别人员的经验以提高识别的准确率，这方面的研究还在继续。说话人识别错误有两种情况：一是没有认出正确的说话人；二是认错了人，后者的效果往往更坏。

还有一些识别难度更大，但更有实际价值的领域，如下所示。

(1) 用通过电话信道的语音进行“说话人识别”，由于电话频带窄、有失真、噪声大，不同信道条件各异，所以该识别十分困难。但由于大量待识别语音均取自于电话信道，所以解决这问题的意义很大。

(2) 在“辨认”说话人时，语句往往不能规定，在没有指定语句条件下的识别较困难，必须有更多的样本用做训练和测试，以降低误识率，这类无指定测试语句的识别称为“与

文本无关”的说话人识别，而在有指定语句条件下进行的识别称为“与文本有关”的说话人识别。

4. 语音理解

语音理解是指利用知识表达和组织等人工智能技术进行语句的自动识别和语意的理解。它与语音识别的主要不同是对语法和语义知识的充分利用程度。由于人们对语音具有广泛的知识，可以对要说的话具有一定的预见性，所以人对语音具有感知分析的能力。依靠人对语言和谈论的内容具有的广泛知识，以及利用知识提高计算机理解语言的能力，是语音理解研究的核心。

利用理解能力，不仅可以排除噪声的影响，理解上下文的意思并用来纠正错误，澄清不确定的语义，而且能够处理不合语法或不完整的语句。一个语音理解系统除了包括原语音识别所要求的部分外，还必须增加知识处理部分。知识处理包括知识的自动收集、知识库的形成、知识的推理与检验等，当然还希望能具备自动进行知识修正的能力。因此，语音理解可以认为是信号处理与知识处理的产物。语音知识包括音位知识、音变知识、韵律知识、词法知识、句法知识、语义知识及语用知识。这些知识涉及语音学、汉语语法、自然语言理解及知识搜索等许多交叉学科。

实现完善的语音理解系统是非常困难的，然而面向特定任务的语音理解系统是可以实现的，如飞机票预售系统、银行业务、旅馆业务的登记及询问系统等。

5. 语音合成

语音合成的目的就是让计算机说话。

最简单的语音合成应当是语音响应系统，其实现技术非常简单。在计算机内建立一个语音库，将可能用到的单字、词组或一些句子的声音信号编码后存入计算机，当输入所要的字、词组或句子代码时，就能调出对应的数码信号并将其转换成声音。

按规则的文字—语音合成系统是将文字转换成语音，让计算机模仿人来朗读文本。该系统具有以下作用：有一个存储基本语音单元的音库；当用各种方式输入文字信息时，计算机能将文字内容按照语言规则，转换成由基本音元组成的序列；按说话时音元连接的规则控制音元序列，输出连续自然的声音。这种系统也称“文—语转换系统”(Text To Speech, TTS)。

建立音库时，语音单元的选择是一个很重要的问题。因为一种语言的音素通常只有几十个，采用音素作为音元可以降低存储容量，但用音素合成语音非常复杂，而且自然度较差，所以一般认为汉语中采用音节作为音元比较合适，这是因为汉语中的一个音节就是一个字的音，汉语中只有412个无调音节，形成音库比较适中。也可以用单字和词组作为音元，但一个字不能只存一种发音，这是因为汉语是多音字，字的发音与上下文有关，只有存储与上下文关联的几种发音，使用时按上下文关系调用，合成的语音才能比较自然，这

就要求系统有很大的存储容量。

该系统中的“规则”有两层含义：一是文字变语言，如“。”要置换成“句号”；二是要按照复杂的语音规则和上下文的关系决定音调、语气、重音、音长、停顿、过渡等，组成发音控制参数序列。

要使文—语转换系统合成出高质量的语音，不仅要掌握语音信号的数字处理技术，而且要有语言学知识的支持。

更高层次的合成是“按概念或意向到语音的合成”，即将“概念、意向”组成语言并变成声音，就如大脑形成说话内容并控制发声器官产生声音一样。

6. 语音增强

在实际的应用环境中，语音都会不同程度地受到环境噪声的干扰。语音增强就是对带噪语音进行处理，降低噪声的影响，改善听觉的效果。有些语音编码和语音识别系统在无噪声或噪声很小的环境中性能很好，但当环境噪声增大时，性能却急剧下降。因此，最大程度地去除噪声，改善听觉效果，也是语音编码和语音识别等系统必须解决的问题。

实际语音可能遇到的干扰可以分为以下几类：

(1) 周期性噪声，如电气干扰，发动机旋转引起的干扰等，这类干扰在频域上表现为一些离散的窄峰；

(2) 冲激噪声，如电火花、放电产生的噪声干扰，这类干扰在时域上表现为突然出现的窄脉冲；

(3) 宽带噪声，这是指高斯噪声或白噪声一类的噪声，其特点是频带宽，几乎覆盖整个语音频带；

(4) 语音噪声，如话筒中同时进入多个人的声音，或者在传输时遇到串音引起的语音噪声。

对于上述各种不同类型的噪声，语音增强的方法也是不同的。例如，周期性噪声可以用滤波的方法滤除；冲激噪声可以通过相邻的样本值，采取内插方法滤除，或者利用非线性滤波器滤除；宽带噪声是一种难以滤除的干扰，因为它与语音具有相同的频带，在消除噪声的同时将不可避免地影响语音的质量，现在常用的方法有谱减法、自相关相减法、最大似然估计法、自适应抵消法等；语音噪声也是很难消除的，一般可以采用自适应技术跟踪某个说话人的特征的方法来消除。

1.3 DSP 语音处理系统

通常，语音处理应用可基于通用的可编程 DSP 芯片来实时实现，为便于说明，本书将“基于可编程 DSP 芯片的语音处理系统”简称为“DSP 语音处理系统”。

1.3.1 DSP 语音处理系统的构成

如图 1-3 所示为一个典型的 DSP 语音处理系统。

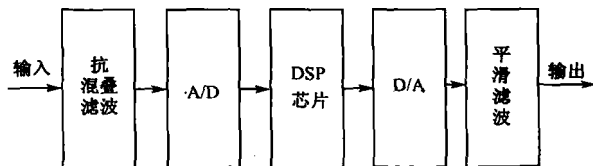


图 1-3 典型的 DSP 语音处理系统

一般地，输入的模拟语音信号首先进行抗混叠滤波，然后进行模/数（A/D）变换将模拟语音信号变换成数字比特流。根据奈奎斯特（Nyquist）抽样定理，对于模拟语音信号，为保持信息的不丢失，抽样频率至少必须是输入语音信号最高频率的 2 倍。

DSP 芯片的输入是 A/D 变换后得到的以抽样形式表示的数字语音信号，根据系统要求，DSP 芯片对输入语音信号按照特定的算法进行处理，这是语音处理系统的关键。最后，经过处理后的数字样值再经数/模（D/A）变换转换为模拟样值，之后再进行平滑滤波就可得到连续的模拟波形。

当然，图 1-3 给出的 DSP 语音处理系统只是一个典型的模型系统，并不是所有的系统都必须包括该模型系统中的各个部分。例如，语音识别系统的输出并不是语音，而是识别结果，如文字、数字、标点符号等。

1.3.2 DSP 语音处理系统的特点

DSP 语音处理系统以数字信号处理为基础，具有数字处理系统的以下全部优点。

（1）接口方便。DSP 语音处理系统与其他以现代数字技术为基础的系统或设备都是相互兼容的，接口方便、容易。

（2）编程方便。DSP 语音处理系统中的可编程 DSP 芯片可使设计人员在开发过程中灵活方便地对软件进行修改和升级。

（3）稳定性好。DSP 语音处理系统以数字处理为基础，受环境温度及噪声的影响较小，可靠性高、稳定性好。

（4）精度高。DSP 语音处理系统采用最常用的 16 位定点 DSP 芯片，系统可以达到 10^{-5} 的精度（ $1/2^{15}$ ）。

（5）可重复性好。模拟系统的性能受元器件参数性能变化的影响比较大，而数字系统基本不受影响，因此便于调试、测试和大规模生产。

（6）集成方便。DSP 语音处理系统中的数字部件有高度的规范性，便于大规模集成。

DSP 语音处理系统的突出优点已经使其在语音编码、语音识别、说话人识别、语音增