



书中所有程序的源代码
可通过扫描此二维码
免费下载。

MATLAB语音信号 分析与合成

(第2版)

宋知用 编著

非
外
借



北京航空航天大学出版社
BEIHANG UNIVERSITY PRESS

MATLAB®
examples

MATLAB® & Simulink® 开发实例系列丛书

MATLAB 语音信号分析与合成

(第2版)

宋知用 编著



配套资料(源程序+数据)

北京航空航天大学出版社

内 容 简 介

语音信号处理是数字信号处理的一个重要分支。本书含有许多数字信号处理的方法和 MATLAB 函数。全书共 10 章。第 1~4 章介绍语音信号处理的一些基本分析方法和手段,以及相应的 MATLAB 函数;第 5~9 章介绍语音信号预处理和特征的提取,包括消除趋势项和基本的减噪方法,以及端点检测、基音的提取和共振峰的提取,并利用语音信号处理的基本方法,给出了多种提取方法和相应的 MATLAB 程序;第 10 章结合各种参数的检测介绍了语音信号的合成、语音信号的变速和变调处理,还介绍了时域基音同步叠加(TD-PSOLA)的语音合成,并给出了相应的 MATLAB 程序。附录 A 中给出了调试复杂程序的方法和思路。

本书可作为从事语音信号处理的本科高年级学生、研究生或科研工程技术人员的辅助读物,也可作为从事信号处理研究与应用的科研工程技术人员的参考用书。

图书在版编目(CIP)数据

MATLAB 语音信号分析与合成 / 宋知用编著. --2 版

. -- 北京:北京航空航天大学出版社,2017.10

ISBN 978-7-5124-2575-0

I. ①M… II. ①宋… III. ①Matlab 软件—应用—语音信号处理 IV. ①TN912.3

中国版本图书馆 CIP 数据核字(2017)第 256066 号

版权所有,侵权必究。

MATLAB 语音信号分析与合成(第 2 版)

宋知用 编著

责任编辑 陈守平

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(邮编 100191) <http://www.buaapress.com.cn>

发行部电话:(010)82317024 传真:(010)82328026

读者信箱:goodtextbook@126.com 邮购电话:(010)82316936

北京宏伟双华印刷有限公司印装 各地书店经销

*

开本:787×1092 1/16 印张:25 字数:656 千字

2018 年 1 月第 2 版 2018 年 1 月第 1 次印刷 印数:4 000 册

ISBN 978-7-5124-2575-0 定价:58.00 元

他、论坛和书^{*}

(代序)

我希望将来退休的时候:每天早上去公园里走走,吃个惬意的早餐,回来的路上顺便去菜市场买菜;回到家里是上午9点左右,然后打开电脑,开始在MATLAB中文论坛里解答学生、研究人员提出的信号处理相关的问题;下午一杯茶,读一些书,调试一些MATLAB程序;晚上的时候,写写书稿,把自己的毕生所学传递给他人。

——这正是宋知用老师现在的生活。

很多读者可能不知道宋知用老师是谁。宋老师从20世纪70年代开始(我们大多数还没有出生!)就在中科院从事信号处理的相关研究工作。后来一直在工业界从事信号、语音方面的研究,可以说在中国,没有人再比宋老师和他的同事们更早从事信号处理(尤其是数字信号处理)方面的工作了。宋老师退休以后,常年“潜伏”在MATLAB中文论坛,为数以千计的学生、研究人员在信号处理方面排忧解难。“潜伏”总是有“危险”的,论坛里的两项数据把宋老师“出卖”了:在线时间+回帖数。到我写这个序言为止,宋老师在线1万多小时,回帖数是1679(当您拿到这本书的时候,相信回帖数已经发生了很大的变化)。请注意:宋老师的1679个回帖全部都是解答问题贴!你不信?到MATLAB中文论坛搜索一下宋老师的回帖吧:<http://www.ilovematlab.cn/home.php?mod=space&uid=13862&do=thread&view=me&type=reply&from=space>

我不能代表MATLAB中文论坛的70万会员,但是我能代表我自己和MATLAB中文论坛,对宋老师说:您辛苦了,您是我们的楷模!

得知宋老师的《MATLAB在语音信号分析和合成中的应用》一书即将出版,非常高兴。一位在一线工作近40年的工程师、一位在论坛里解答问题超过1600个的热心人、一位在线时间超过10000小时的会员写的书,我们有什么理由不去喝彩和期待?当我接到要为这本书写几句话这个光荣任务的时候,我对自己说,我必须亲自见到宋老师、必须跟他当面请教以后,我这个晚辈才有资格来写。终于在2013年8月13号,在北航出版社陈守平编辑的安排下,我有幸和宋老师在北京航空航天大学“雕刻时光”咖啡馆里促膝长谈。我们讨论了当今工业界研究方向、学生的教育、信号处理在各行各业的应用等,期间无时无刻不为宋老师的博学和对信号处理的高度驾驭能力所折服。

对于正在从事信号处理,尤其是语音方面研究的学生和学者,你们是幸运的。因为《MATLAB在语音信号分析和合成中的应用》是一本包含了宋老师30多年工作经验的书,书

^{*} 本序言为本书上一版时的序言,由MATLAB中文论坛的创始人math写于2013年8月。截至本书出版,宋老师在论坛的回帖已增至4500左右,最佳答案1000多个。序言中提到的《MATLAB在语音信号分析和合成中的应用》一书是本书上一版的书名。——编辑注

中的每个案例都有详细的解释说明,因为这么多年“驰骋”论坛已经让宋老师深知学生的需求;同时也因为你们有像宋老师这样负责任的作者,一年四季,风雨无阻在论坛里兢兢业业解答大家的提问。

不管什么样的语言在数据面前都是苍白的。我在这里传授给大家一个判断是否值得拥有一本书的小技巧:如果该书作者在 1 600 多个回帖里,没有一个灌水帖、没有一项是敷衍的回答,就像宋老师这样,我想大声对你说:你就从了吧!



我(左)和宋知用老师在北航出版社门口合影(2013.08.13)

张延亮(博士,教授)

2013年8月

(论坛用户名:math, MATLAB 中文论坛(www.iLoveMATLAB.cn)独立创始人)

第 2 版前言

本书第 1 版于 2013 年出版,出版后在广大网友和读者的帮助下我发现第 1 版书中有少量的编写和印刷错误,此次全部予以修正。在此要对所有帮助过我的网友和读者表示感谢!

对张延亮先生(MATLAB 中文论坛的独立创始人,在论坛的 ID 为 math)表示感谢,他在百忙之中为我的书写了序,介绍了本书。也对北航出版社的陈守平编辑表示感谢,多年来一直得到她的帮助和支持。对我家人的长期支持也深表感谢!

同时,我也通过 MATLAB 中文论坛发现过去的 3 年期间好多读者都没有认真阅读本书第 1 版的前言,没有按照要求在 MATLAB 中对本书提供的函数设置好必要的路径,使得经常发生一些低级的错误,例如找不到某些函数,等。在此再次特别强调,要通过 **set path** 设置子目录的路径:

(1) 运行本书程序,请设置 basic_tbx 子目录;

(2) 因为本书中引用的大部分语音信号都在 speech_signal 子目录中,所以请设置 speech_signal 子目录;

(3) 要进行 EMD 处理,请设置 EMD 子目录;

(4) 要运行主体-延伸基音检测,请设置 Pitch_ztlib 子目录;

(5) 要进行时域基音同步叠加语音合成(PSOLA),请设置 Psola_lib 子目录。

上述子目录都在随书赠送的配套资料之中。

宋知用

2017.3



配套资料(源程序+数据)

本书所有配套资料均可通过 QQ 浏览器扫描二维码免费下载。读者也可以通过以下网址下载全部资料: <http://www.buaapress.com.cn/upload/download/20171013mlyyxh.ZIP> 或者 <https://pan.baidu.com/s/1o78SNJs>。若有与配套资料下载或本书相关的其他问题,请咨询北京航空航天大学出版社理工图书分社,电话(010)82317036,邮箱:goodtextbook@126.com。

前言

我与计算机打交道已有 40 多年。在 20 世纪 70 年代初开始接触计算机,最早使用的是晶体管的 108 乙机,用 5 孔电报纸带上机,用机器语言编程,调试一个小程序可能需要数周甚至数月。随着技术的发展,我曾使用过 Nova1200、z80、Intel8086 / 80286 / 80386 等,而编程使用的计算机语言也从汇编到 BASIC、FORTRAN 等。数字信号处理最基础的是快速傅里叶变换和数字滤波器,但在不同的机种用不同的编程语言都需要编制相应的基础性处理程序,这些都极其耗费时间和精力。

21 世纪初我偶然接触到 MATLAB 语言,立刻被它的功能所吸引。MATLAB 除了编程简单外,还有强大的工具箱(全世界的精英都为工具箱做贡献)。一些基础性的处理程序都已包含在工具箱中,不需要用户去从事这方面的开发工作,这样节省了大量的重复性的工作。对于用户来说,只须考虑怎样利用工具箱来实现自己的想法和算法。MATLAB 的程序大部分不需要编译、链接等一套烦琐的过程,输入程序后就能运行。MATLAB 是一种方便、实用、高效的计算机语言。

互联网的发展形成了很多以 MATLAB 为基础的科技讨论社区。在这些论坛社区里集中了来自社会各界和高校各学科各专业的 MATLAB 使用者、爱好者。我从 2002 年开始在国内一些成立较早的论坛社区如研学论坛和振动论坛等的信号处理版块中参与各类问题的解答。MATLAB 中文论坛成立后,同样在信号处理版块经常与各种程度的 MATLAB 使用者打交道,解答使用者通常会遇到的一些问题。我发现在对语音信号进行处理的过程中,有好多用户对于怎么把 MATLAB 应用于语音处理不甚熟悉,这些因素促使我萌生了编写本书的想法。

运用 MATLAB 处理语音信号至少需要掌握三方面的知识:语音信号处理的基础理论、数字信号处理的基础理论和 MATLAB 的编程技术。本书的目的是帮助本科高年级学生和硕士生尽快掌握怎么把 MATLAB 应用于语音信号的分析和合成中去,因此书中介绍了语音信号处理的基础知识,介绍了语音分析和合成的基本方法,以及相应的 MATLAB 函数和程序,读者以这些方法、函数和程序为基础,进一步去解决自己的问题,可少走一些弯路。书中介绍的各种方法都还有继续改善和拓展的空间,使它们更加完善,取得更好的效果。本书也适合于从事数字信号处理的广大高校师生和科研工作人员作为参考用书。但阅读本书的读者应掌握数字信号处理的基本知识,以及 MATLAB 编程的基本技能。

本书介绍语音信号分析和合成处理的基础、原理、方法和应用。全书共 10 章,具体内容如下:

第 1 章介绍语音信号处理的基础知识,如发音器官与听觉器官、语音信号的数学模型和语音信号感知特性等。

第 2 章和第 3 章介绍语音信号特征分析的基本处理技术,包括时域分析、频域分析、同态分析、DCT 和 MFCC 分析、小波和小波包变换的分析以及 EMD 分析等方法。

第 4 章介绍线性预测分析方法。介绍了线性预测的模型、方程的建立、线性预测的自相关和自协方差解、线性预测的格型法解、由线性预测导出的其他参数和线谱对分析法等。

第 5 章介绍了带噪语音信号及预处理。介绍了信噪比的概念和带噪语音信号的产生,以及最小二乘法消除趋势项和数字滤波等。

第6章介绍语音端点的检测方法。首先从能量和过零率导出通用的双门限单参数和双参数的端点检测,接着介绍相关法、方差法、谱距离法、谱熵法、能零比和能熵比法、小波变换和EMD分解法等端点检测,最后给出在低信噪比下端点检测的方法。

第7章介绍语音信号的减噪。介绍了利用自适应滤波器减噪、基本谱减法和改进谱减法的减噪,以及维纳滤波减噪等方法。

第8章介绍语音信号的基音检测技术。首先介绍了在基音检测中的端点检测和预滤波,接着分别介绍了倒谱法、自相关法、平均幅度差函数法和线性预测等方法的基音检测。这些方法中虽在基音检测后都进行了平滑处理,但在基音周期中都会有野点发生。这里提出了主体-延伸的基音检测方法,并详细说明了该方法的原理和处理步骤,改善了基音检测的结果。本章的最后介绍了带噪语音的基音检测技术。

第9章介绍语音的共振峰检测。首先介绍了在共振峰检测中的预加重和端点检测,接着介绍了倒谱法和线性预测法的共振峰检测,并在线性预测的基础上进一步介绍了简单LPC和改进的LPC对连续语音的共振峰检测。本章最后介绍了HHT法的共振峰检测。

第10章介绍语音信号的合成算法。首先介绍了数据接叠的三种方法,在此基础上介绍了频谱参数的语音合成、线性预测系数和预测误差的语音合成、线性预测系数和基音参数的语音合成、基音和共振峰的语音合成。通过合成技术又介绍了语音信号的变速和变调算法,最后介绍了波形拼接合成技术以及时域基音同步叠加(TD-PSOLA)的合成方法。

在附录A中以主体-延伸基音检测法为例,说明了程序的调试和修改方法,以帮助读者调试和修改程序使之成为适合自己的应用函数和程序。

本书中除第1章外的各章均附有函数和程序。书中经常会调用的一些函数(自编函数或取自其他应用工具箱中的函数),已集中在basic_tbx工具箱中,在运行本书的程序前请把该工具箱设置(用set path设置)在工作路径下。当要运行EMD处理时,要把emd工具箱设置在工作路径下;当要运行主体-延伸基音检测时,要把Pitch_ztlib工具箱设置在工作路径下;当要进行时域基音同步叠加语音合成时,要把psola_lib工具箱设置在工作路径下;当要应用本书提供的语音数据时,最好把speech_signal设置在工作路径下。本书的所有函数和程序都在MATLAB R2009a版本下调试通过。

本书的读者-作者在线交流平台为<http://www.ilovematlab.cn/forum-173-1.html>。本书作者将通过该平台与广大读者交流,解决大家在阅读本书过程中遇到的问题,分享彼此的学习经验,从而达到共同进步的目的。

在编写本书的过程中,作者得到了MATLAB中文论坛创始人张延亮(math)博士的指导和帮助,也得到了北京航空航天大学出版社陈守平编辑的支持与鼓励,中国语音学会会长、中国社会科学院人类学与民族学研究所鲍怀翘教授对本书的编写提出了不少宝贵意见,在此向他们表示衷心的感谢。

由于编写时间仓促,加之作者学识所限,书中如有错误和疏漏之处,恳请广大读者和各位专家批评指正。

目 录

第 1 章 语音的产生和感知	1
1.1 发声器官	1
1.2 语音信号的数字模型	2
1.2.1 激励模型	3
1.2.2 声道模型	4
1.2.3 辐射模型	7
1.3 语音的感知	7
1.3.1 人耳的构造	7
1.3.2 听觉感受性	8
1.3.3 掩蔽效应	8
1.3.4 响 度	10
1.3.5 音 高	11
参考文献	11
第 2 章 语音信号的时域、频域特性和短时分析技术	12
2.1 MATLAB 中的语音信号分帧	12
2.2 语音分析中的窗函数	15
2.3 语音信号短时时域处理	16
2.3.1 短时能量和短时平均幅度	16
2.3.2 短时平均过零率	18
2.3.3 短时自相关函数	19
2.3.4 短时平均幅度差函数	20
2.4 语音信号短时频域处理	21
2.4.1 短时傅里叶变换的定义	22
2.4.2 语谱图	25
2.4.3 短时功率谱密度	27
参考文献	29
第 3 章 语音信号在其他变换域中的分析技术和特性	30
3.1 语音信号的同态处理和倒谱分析	30
3.1.1 同态处理的基本原理	30
3.1.2 复倒谱和倒谱	31
3.2 离散余弦变换	34
3.3 Mel 频率倒谱系数的分析	37
3.3.1 Mel 滤波器组	37
3.3.2 MFCC 特征参数提取	38
3.4 小波和小波包变换	43

若您对此书内容有任何疑问，可以凭在线交流卡登录MATLAB中文论坛与作者交流。

3.4.1	小波变换	43
3.4.2	小波包变换	44
3.4.3	小波包算法	45
3.4.4	MATLAB 中一维小波和小波包变换函数	46
3.4.5	MATLAB 语音信号小波和小波包变换的例子	49
3.5	EMD 的基本理论和算法	53
3.5.1	EMD 的基本概念	53
3.5.2	EMD 的基本原理	55
3.5.3	EMD 法的完备性和正交性	57
3.5.4	基于 EMD 的 Hilbert 变换的基本原理和算法	59
3.5.5	EMD 法的 MATLAB 函数	60
	参考文献	61
第4章	语音信号的线性预测分析	62
4.1	线性预测分析的基本原理	62
4.1.1	信号模型	62
4.1.2	线性预测方程的建立	64
4.1.3	语音信号的线性预测分析	65
4.2	线性预测分析自相关和自协方差的解法	66
4.2.1	自相关法	66
4.2.2	协方差法	71
4.3	线性预测分析格型法的解法	72
4.3.1	格型法的基本原理	72
4.3.2	格型法的求解	74
4.4	线性预测导出的其他参数	78
4.4.1	预测误差及其自相关函数	79
4.4.2	反射系数和声道面积	79
4.4.3	线性预测的频谱和预测误差滤波器 $A(z)$ 多项式的根	81
4.4.4	线性预测倒谱	83
4.5	线谱对的分析法	86
4.5.1	LSP 的定义和特点	87
4.5.2	LPC 到 LSP 参数的转换	89
4.5.3	LSP 参数到 LPC 的转换	91
	参考文献	95
第5章	带噪语音和预处理	96
5.1	纯语音和带噪语音	96
5.2	信噪比	96
5.3	带噪语音的产生	97
5.4	语音信号的预处理——消除趋势项和直流分量	101
5.4.1	最小二乘法拟合趋势项的原理	102
5.4.2	最小二乘法拟合消除趋势项的函数	103

5.5 语音信号的预处理二——数字滤波器	105
5.5.1 IIR 低通、高通、带通和带阻滤波器的设计	105
5.5.2 FIR 低通、高通、带通和带阻滤波器的设计	109
参考文献	116
第 6 章 语音端点的检测	117
6.1 双门限法	117
6.2 双门限法的改进和推广	123
6.2.1 噪声的影响	123
6.2.2 平滑处理	125
6.2.3 双参数的双门限检测法	127
6.2.4 单参数的双门限检测法	129
6.3 相关法的端点检测	131
6.3.1 自/互相关函数最大值的端点检测	131
6.3.2 归一化自相关函数的端点检测	134
6.3.3 自相关函数主副峰比值的端点检测	136
6.3.4 自相关函数余弦角值的端点检测	138
6.4 方差法的语音端点检测	141
6.4.1 频带方差的端点检测	141
6.4.2 均匀子带分离频带方差的端点检测	142
6.4.3 频域 BARK 子带方差的端点检测	143
6.4.4 小波包 BARK 子带方差的端点检测	145
6.5 谱距离法的端点检测	148
6.5.1 对数频谱距离的端点检测	149
6.5.2 倒谱距离的端点检测	151
6.5.3 MFCC 倒谱距离的端点检测	153
6.6 谱熵在端点检测中的应用	155
6.6.1 谱熵法的端点检测	155
6.6.2 谱熵法端点检测的改进	156
6.7 能零比和能熵比的端点检测	159
6.7.1 能零比的端点检测	159
6.7.2 能熵比法的端点检测	161
6.8 小波变换和 EMD 分解在端点检测中的应用	162
6.8.1 小波变换在端点检测中的应用	162
6.8.2 EMD 分解在端点检测中的应用	164
6.9 低信噪比时的端点检测	167
6.9.1 噪声的估算	168
6.9.2 基本谱减法和方差法的端点检测	170
6.9.3 多窗谱估计谱减法和能熵比法的端点检测	172
参考文献	174

第7章 语音信号的减噪	176
7.1 自适应滤波器减噪	176
7.1.1 LMS 算法基本原理	176
7.1.2 基本 LMS 自适应算法	178
7.1.3 LMS 的自适应带陷滤波器	181
7.2 谱减法减噪	184
7.2.1 基本谱减法	184
7.2.2 改进的谱减法	187
7.3 维纳滤波法减噪	195
7.3.1 维纳滤波的基本原理	195
7.3.2 维纳滤波减噪的具体步骤和函数 WienerScalart96	197
7.3.3 维纳滤波的 MATLAB 例子	199
参考文献	201
第8章 基音周期的估算方法	202
8.1 基音周期提取的预处理	203
8.1.1 基音检测中的端点检测	203
8.1.2 基音检测中的带通滤波器	204
8.2 倒谱法的基音检测	205
8.2.1 倒谱法基音检测原理	205
8.2.2 倒谱法基音检测的 MATLAB 程序	206
8.2.3 简单的后处理方法	207
8.3 短时自相关法的基音检测	209
8.3.1 短时自相关函数法	209
8.3.2 中心削波的自相关法	211
8.3.3 三电平削波的互相关函数法	212
8.3.4 基于自相关函数法提取基音的 MATLAB 程序	214
8.4 短时平均幅度差函数的基音检测	215
8.4.1 短时平均幅度差函数法	215
8.4.2 改进的短时平均幅度差函数法	217
8.4.3 循环平均幅度差函数法	218
8.4.4 基于平均幅度差函数法提取基音的 MATLAB 程序	220
8.4.5 自相关函数法和平均幅度差函数法的结合	221
8.5 线性预测的基音检测	223
8.5.1 线性预测倒谱法	223
8.5.2 简化逆滤波法	225
8.6 基音检测的进一步完善	227
8.6.1 主体-延伸法的原理和方法	228
8.6.2 主体-延伸基音检测法的步骤	229
8.6.3 端点检测和元音主体的检测	230
8.6.4 元音主体的基音检测	232

8.6.5 计算延伸区间和长度	239
8.6.6 在延伸区间进行基音检测	241
8.6.7 主体-延伸基音检测法的 MATLAB 程序	248
8.7 带噪语音中的基音检测	251
8.7.1 小波-自相关函数法	251
8.7.2 谱减-自相关函数法	253
8.7.3 谱减法与主体-延伸法相结合	255
参考文献	258
第 9 章 共振峰的估算方法	259
9.1 预加重和端点检测	259
9.1.1 预加重	259
9.1.2 端点检测	260
9.2 倒谱法对共振峰的估算	260
9.2.1 倒谱法共振峰估算的原理	260
9.2.2 倒谱法共振峰估算的 MATLAB 程序	261
9.3 LPC 法对共振峰的估算	262
9.3.1 LPC 法共振峰估算的原理	262
9.3.2 LPC 内插法共振峰的估算	263
9.3.3 LPC 求根法共振峰的估算	266
9.4 连续语音 LPC 法共振峰的检测	268
9.4.1 简单的 LPC 共振峰检测	268
9.4.2 改进的 LPC 共振峰检测	270
9.5 基于 Hilbert - Huang 变换 (HHT) 的共振峰检测	274
9.5.1 希尔伯特变换	275
9.5.2 语音信号的另一种模型——AM - FM 模型	278
9.5.3 对 AM - FM 模型的分析	279
9.5.4 语音信号共振峰特征参数提取的 HHT 方法	279
9.5.5 基于 Hilbert - Huang 变换的共振峰检测步骤和 MATLAB 程序	280
参考文献	283
第 10 章 语音信号的合成算法	284
10.1 语音合成中数据叠接的三种方法	285
10.1.1 重叠相加法	285
10.1.2 重叠存储法	290
10.1.3 线性比例重叠相加法	293
10.2 用频谱参数合成语音信号	296
10.3 线性预测系数和预测误差的语音信号合成	298
10.4 线性预测系数和基音参数的语音信号合成	300
10.4.1 预测系数和基音参数语音合成的模型	300
10.4.2 基音检测函数	301
10.4.3 激励脉冲的产生	303

10.4.4	预测系数和基音参数语音合成的程序清单	304
10.5	基音和共振峰合成语音信号	308
10.5.1	共振峰和基音参数语音合成的模型	308
10.5.2	线性预测共振峰检测和基音参数的语音合成程序	315
10.5.3	倒谱法与内插法结合的共振峰检测和基音参数的语音合成程序	319
10.6	语音信号的变速和变调算法	322
10.6.1	语音信号的变速	322
10.6.2	语音信号的变调	327
10.6.3	语音信号变速又变调	331
10.7	波形拼接合成技术和时域基音同步叠加	334
10.7.1	波形拼接合成技术	334
10.7.2	时域基音同步叠加(TD-PSOLA)合成技术	335
10.7.3	时域基音同步叠加(TD-PSOLA)的 MATLAB 工具箱	339
	参考文献	344
附录 A	程序的调试和修改	345
A.1	准备工作	345
A.2	元音主体中的基音检测与调试	350
A.3	元音主体前后向延伸中的基音检测与调试	359
A.4	更多中间数据的检测	370
A.5	参数的调整	373
附录 B	本书自编函数速查表	375
附录 C	本书应用的 MATLAB 函数速查表	378

第 1 章

语音的产生和感知

语音是人类所发出的声音,然而这是一种特殊的声音,是由人类讲话时发出的。语音由一连串的音所组成,各个音的排列由一些规则控制。对这些规则及其含义的研究属于语言学范畴,对语音中音的分类和研究称为语音学。

语音中最小的基本单位是音素。各种不同的音是由不同的发音方法和发音部位所产生的。由音素构成音节,又由音节构成不同的词或单词。音素是人类能区别一个单词和另一个单词声音的基础。音素分为元音和辅音两类。

汉语不同于大部分西方语音,有其特殊性。汉语里也有元音和辅音之分。其中,不同的元音是由口腔不同的形状造成的;而不同的辅音是由发音部位和发音方法不同造成的。但是,汉语语音分析中总是把一个汉语音节分为声母和韵母两部分:声母是指一个汉语音节开头的辅音,它们比较简单,只是一个音素;而韵母则比较复杂,它们是汉字音节除了开头的声母以外的部分。汉语中有 21 个声母和 39 个韵母。

汉语的特点是:汉语的自然单位为音节,每一个字都是单音节字,即汉语的一个音节就是汉语一个字的音,这里所指的“字”是一个独立发音的单位;再由音节字构成词(主要是两音节字构成的词),而每一个音节字又都是由声母和韵母拼音而成;最后再由词构成句子。

汉语语音的另一个重要特点是其具有声调(即音调在发一个音节中的变化),这样使用汉语较其他语音更为经济。声调是一种音节在念法上的高低升降的变化。汉语有四种声调,即阴平(ˊ)、阳平(ˊ)、上声(ˇ)、去声(ˋ),括号内表示的是该声调的符号。由于有声调之分,所以参与拼音的韵母又有若干种(包括轻声在内至多有 5 种)声调。

本章对语言学和语音学都不作论述,而主要从声学角度介绍语音的产生过程、语音的声学模型和语音的感知(听觉系统)。

1.1 发声器官

人类的语音是由人体发音器官在大脑控制下作生理运动产生的。人体发音器官由三部分组成:肺和气管、喉、声道。

肺的主要生理功能是使血液和空气之间进行气体交换,即将空气中的氧气吸入血液,而将血液中的二氧化碳气体排出到空气中。这就是人体的呼吸功能。肺还有另一个重要功能,即将压缩空气供给发音器官。人在正常呼吸的情况下,不说话时,呼吸通常是规则的、平稳的、节律性的;而在说话时,为了保持语音有一定程度的连续性,呼吸就不得不有短暂的停顿,其特点是吸气短、呼气长,且呼吸受到句子结构的控制,并没有一个固定的规则。空气由肺部经气管进入喉部,又经过声带进入声道,最后由嘴辐射出声波,这才形成了语音。可见,肺是产生语音的能源所在。

气管连接着肺和喉,是肺与声道联系的通道。

喉是一个由软骨和肌肉组成的复杂系统,其中包含着重要的发音器官——声带。声带受

喉部软骨和肌肉的控制。声带绷在喉头的前后壁上,前端由甲状软骨支撑,后端由杓状软骨支撑。杓状软骨与环状软骨的上部相连接。这些软骨由附在环状软骨上的一组肌肉所控制,可以移动声带的末端使之开启或闭合。当声带末端分离开启时,这就是正常呼吸的状态。两片声带之间的空间叫做声门。当声带末端闭合在一起时,肺道便被封闭起来构成了一个密闭的小室。声带具有生物学和声学两种功能。它的生物学功能是封闭气管以保护肺道(例如在吞咽食物时防止食物进入肺道),或在胸腔和腹腔建立一定气压(例如为了帮助排泄和分娩)。声带的声学功能是为产生语音提供主要的激励源。讲话时声带先合拢,后因受声门下气流的冲击而开启;但由于声带韧性又迅速地闭合,随后又开启与闭合,这样不断重复。不断地开启与闭合的结果,使声门向上送出一连串喷流而形成一系列脉冲。声带每开启和闭合一次的时间就是声带振动的周期,也就是音调周期或基音周期;它的倒数称为基音频率,简称为基音。基音决定了声音频率的高低,频率高则音调高,频率低则音调低。基音的范围为 $60 \sim 500 \text{ Hz}$,它随发音人的性别、年龄而定。老年男性偏低,小孩和青年女性偏高。

声道是指声门至嘴唇的所有发音器官,其剖面图如图 1-1-1 所示。声道包括咽喉、口腔和鼻腔。口腔包括上下唇、上下齿、上下齿龈、上下腭、舌和小舌等部分。上腭又分为硬腭和软腭两部分;舌又分为舌尖、舌前、舌中和舌后四部分。鼻腔在口腔上面,靠软腭和小舌(在软腭后端)将其与口腔隔开。当小舌下垂时,鼻腔和口腔便耦合起来;当小舌上抬时,口腔与鼻腔是不相通的。口腔和鼻腔都是发音时的共鸣器。口腔中各器官能够协同动作,使气流通过时形成各种不同情况的阻碍并产生震颤,从而发出一些不同的音来。声道可以被看成一根从声门一直延伸到嘴唇的具有非均匀截面的声管,其截面积主要取决于唇、舌、腭和小舌的形状和位置,最小截面积可以为零(对应于完全闭合的部位),最大截面积可以达到 20 cm^2 。在产生语音的过程中,声道的非均匀截面又是随着时间不断地变化的。成年男性的声道的平均长度约为 17 cm 。当小舌下垂使鼻腔和口腔耦合时,将产生出鼻音。

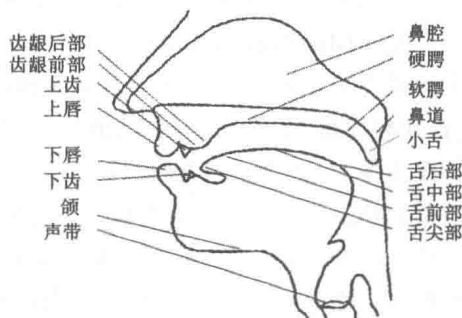


图 1-1-1 声道剖面图

1.2 语音信号的数字模型^[1,2]

表示采样语音信号的离散模型是特别重要的。为了定量描述语音处理所涉及的某些因素,虽然已经假定了许多不同的模型,但是可以肯定,目前还没有找到一种可以详细描述人类语音中已观察到的全部特征的模型(由于它的复杂性,也许不可能找到一个理想的模型)。建立模型的目的是要寻求一种可以表达一定物理状态下的数学关系,而且要使这种关系不仅具有最大的精确度,还要最简单。

模型既是线性的又是时不变的,才是最理想的。但是语音信号是一连串的时变过程,根据语音的产生机理,不能精确地满足这两种性质。此外,声门和声道相互耦合,形成了语音信号的非线性的特性。然而,在做出一些合理的假设、在较短的时间间隔内表示语音信号时,可以采用线性时不变模型。下面给出经典的语音信号数字模型,这里,语音信号被看成是线性时不变系统(声道)在随机噪声或准周期脉冲序列激励下输出。

长期研究证实,发不同性质的音时,激励的情况是不同的。大致分为两大类:发浊音时,气流通过绷紧的声带,冲激声带产生振动,使声门处形成准周期性的脉冲串,并用它去激励声道;发清音时,声带松弛而不振动,气流通过声门直接进入声道。

产生语音信号的示意图如图 1-2-1 所示。

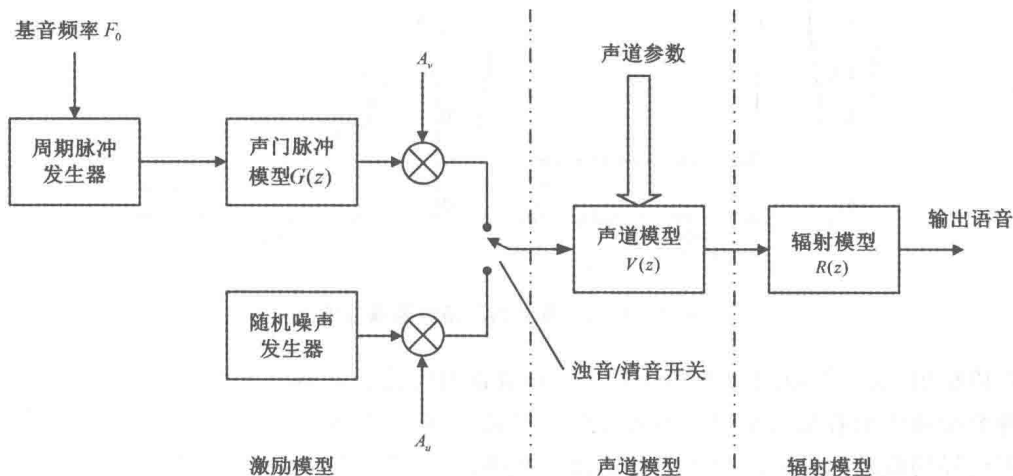


图 1-2-1 语音信号产生的时域模型

下面详细讨论语音信号数字模型的各部分。

1.2.1 激励模型

发浊音时,由于声带不断开启和关闭,将产生间歇的脉冲波。根据测量结果,这个脉冲波类似于斜三角形的脉冲,如图 1-2-2(a)所示。因此,此时的激励信号是一个以基音周期为周期的斜三角脉冲串。单个斜三角波形的数学表达式为

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/N_1)] & 0 \leq n \leq N_1 \\ \cos[\pi(n - N_1)/2N_2] & N_1 \leq n \leq N_1 + N_2 \\ 0 & \text{其他} \end{cases} \quad (1-2-1)$$

式中, N_1 为斜三角波上升部分的时间; N_2 为三角波下降部分的时间。单个斜三角波形的频谱 $G(e^{j\omega})$ 如图 1-2-2(b)所示。由图可见,它是一个低通滤波器,通常更希望将其表示为 Z 变换的全极点模型的形式

$$G(z) = \frac{1}{(1 - e^{-CT}z^{-1})^2} \quad (1-2-2)$$

式中, C 是一个常数。

显然,式(1-2-2)表明斜三角波可描述为一个二极点的模型。

因此,斜三角波脉冲串可被看作加权的单位脉冲串激励上述单个斜三角波模型的结果。而该单位脉冲串及幅值因子可表示成下面的 Z 变换形式

$$E(z) = \frac{A_v}{1 - z^{-1}} \quad (1-2-3)$$

所以整个激励模型可表示为

$$U(z) = G(z)E(z) = \frac{A_v}{1 - z^{-1}} \cdot \frac{1}{(1 - e^{-CT}z^{-1})^2} \quad (1-2-4)$$