数字声音处理

Shuzi Shengyin Chuli

朱家新 张国海 易武秀 编译
〔日〕古井贞熙 著

数字声音处理

[日] 古井 贞熙 著
朱家新 张国海 易武秀 译

人民邮电出版社

登记证号（京）143号

内 容 提 要

本书是以数字信号处理为基础，用较新的研究方式先后介绍了序论、声音的基本性质、声音形成的数字模型、在时域及频域中的声音处理、线性预测分析、声音波形编码、声音合成、声音识别、讲话者识别以及数字声音处理的今后课题等十章的内容。

本书在数字声音处理方面，是一本系统性、理论性及实用性都较强的专著。可作为从事声音处理、通信、音响等专业的研究人员的参考用书。

デイジタル音声処理
古井 貞熙
東海大学出版会
1985年

数字声音处理
[日] 古井貞熙 著
朱家新 张国海 易武秀 译
责任编辑 张 晏

*

人民邮电出版社出版发行
北京东长安街27号
顺义向阳胶印厂印刷
新华书店总店科技发行所经销

*

开本：787×1092 1/32 1993年5月 第一版
印张：8 8/32页数：132 1993年5月 北京第1次印刷
字数：186千字 印数：1－3 000册
ISBN7－115－04868－1/TN・596
定价：5.50元

前　言

声音——是多么不可思议、深奥而迷人。

笔者从事声学研究的十余年间，尽管得到了许多人的指导，研究过多个课题，发表了几篇论文，但是，十余年来尚无靠近声音本质。

鉴于上述考虑和笔者岗位的环境变化，本人设想不急于取得成果，还是回到声音的基础上，再次仔细地开始研究。此时，承蒙斋藤正男教授委托，执笔编写《数字声音处理》一书。对于笔者这种才疏学浅的后辈，的确难以胜任。但是在一直指导笔者开展声学研究的导师、现任工学院大学教授的斋藤收三博士的极力劝告下，还是接受了斋藤正男教授的委托。

前五年间，随着声音合成器和声音识别装置实用化的实施，声音信息处理技术很快受到了人们重视，而且正期待研究开发成果。声音信息处理技术之所以快速发展，除了对声音的长期研究以外，近几年来数字信号处理技术的进步也是原因之一。最初，数字信号处理技术以复杂的模拟系统仿真，应用于声音处理。后来，由于数字信号处理理论的进步和电子计算机的发展，用模拟系统不可能处理的工作，可以用数字计算方法得以实现了。更进一步将声音信息处理作为一种原动力，随着声音处理技术的深入研究，数字信号处理技术本身也得到了发展。这种处理技术包含有FFT，线性预测分析和对数倒频谱，可以认为声音信息处理技术和数字信号处理技术存在着密不可分的关系。

近五年间，出版了一些有关声音信息处理方面的教科书和参考书，但是，给笔者的印象是从数字信号处理的角度出发，综合论述其技术的书不多，笔者在从事研究的过程中用来作为参考书时，希望深一步了解的问题，常常书中没有充分予以介绍。《数字声音处理》把数字信号处理作为基础，以比较新的技术为轴线，囊括了从建立声音模型的知识到声音编码、声音分析、声音合成、声音识别、话者识别的全部技术内容，而且尽可能使声音信息处理的一系列相关技术彼此衔接。同时，避免过分面面具到，而尽量集中归纳了直到将来也会成为基础的技术。

不用说，数字声音处理的研究，是要涉及到音响、信号处理、通信、语言、心理、生理等学科的广泛领域的研究的，但是涉及面如此广泛，不仅远远超越了笔者的能力，而且不是本书的宗旨。因此，在书中或者会出现见解片面，或者因作者能力所限，陈述不充分，或者会产生意料不到的错误。关于这一点，请读者直言不讳地批评、指责，以便修正。

笔者能坚持对声音的研究，应该感谢斋藤收三博士和现任名古屋大学教授板仓文忠博士以及NTT电气通信研究所的上司和同事们。没有他们的指导，是不可能有本书的。在此，谨向他们衷心道谢。在执笔编写本书的过程中，在不少问题上参考了斋藤收三博士和东京农工大学教授中田和男博士的著作。而且，蒙好田正纪博士、佐藤大和先生、东仓洋一博士、誉田雅彰博士等诸位同事提供资料，过目原稿且提出了许多宝贵的意见。至此也向各位衷心致谢。此外，还得到了东海大学出版会的西田光男先生等人的帮助，谨表谢意。

古井 贞熙

1985年7月

译者的话

电话的发明不仅是人类通信史上极其重大的变革，而且可以说是把声波作为人类工程学来进行处理的第一步。在PCM方式未发明之前，对声音的研究只能停留在原封不动的模拟方式上。继PCM方式出现之后，特别是随着数字技术和计算机技术的飞跃发展，把声波数字化，并将其作为数字信号来处理的研究方式诞生，这种处理技术主要包含有快速付里叶变换（FFT），线性预测（LPC）分析和对数倒频谱DFT（CEPSTRUM）分析等。可以认为声音信息处理技术与数字信号处理技术是密不可分的。

近几年来，国内先后出版了一些有关声音处理的教科书和参考书，但是作为综合性专著并不多。本书把数字信号处理作为基础，以较新的研究方式为轴线，综合论述了从声音建模，到声音编码、声音分析、声音合成、声音识别、话者识别的全部技术内容，可以说是一本系统性、理论性与实用性都较强的专著。可作为从事声音处理、通信、音响等专业的研究人员的参考用书。由于译者水平所限，专业知识亦较欠缺，错讹之处在所难免，恳请读者批评指正。

在本书翻译过程中，得到了本书作者NTT古井贞熙博士和日本东海大学出版社加藤千曼树次长的支持与帮助，同时也得到了人民邮电出版社各位先生的支持与帮助。至此一并表示衷心的感谢。

译　者

1992年9月于武汉

目　录

第一章 序 论

通过声音相互传递信息，这是人类最重要的基本功能之一。也可以说，声音几乎是人类不使用工具，而相互传递信息的唯一手段。常言道："百闻不如一见"。这就是说，人们从外界所接受的信息，用眼睛看到的要比用耳朵听到的多得多，即所谓的眼如同嘴一样传情。但是，与用声音传递信息相比较，显然用视觉相互传递信息，其效果要差得多。声音除包含实际发音内容的语言信息之外，还包含发音者是谁及其喜怒哀乐等各种信息。因此，在人类生活中，通过声音来进行信息交换是极其重要的。声音的音响和语言的结构与人的智力活动密切相关，与文化和社会的进步紧密相连。因此，世界上文化高度发达的区域必然是电话网十分发达的区域。

"华逊，我没有给您讲过，我有一个一定让您吃惊的想法，即假定能制作出一种装置，使其电流强度的变化如同声音通过空气时，空气密度发生变化一样，那就能把任何声音，比如讲话声传播到远方去……。" —A·G·Bell[1]

这是贝尔在1875年初的某一晚上，就研究人耳的结构所获得的预测，同他一起从事研究传送音乐的电信机械的助手T·A华逊的一次谈话。这段话即为1876年贝尔发明电话前的重要设想。电话的发明，不仅是通信史上极其重大的变革，而且可以说是把声音波作为人类工程学来进行处理的第一步。研究声音的历史可以说开始于从发明电话之前的18世纪末的机械式声音合成器和19世纪中期就声带振动和听觉的研究，然而直到

1938 年发明 PCM 为止，声音波一直按模拟信号原封不动地处理着。继发明 PCM 之后，数字电路和电子计算机等也发展起来，因而，可以对声音进行数字处理，从而带来了 1960 年以后的声音信息处理的重大飞跃。

到目前为止，声音信息处理，即数字声音处理技术日新月异，其间，具有划时代意义的是 1968 年在东京召开的第 6 届国际音响学术会议上发表的高水平的论文：一篇是日本电报电话公司（现在的 NTT）电气通信研究所发表的采用最优法的声音分析合成系统；另一篇是美国贝尔研究所发表的预测编码法。这两篇论文都是将声音波进行线性预测、信息压缩，以随机过程数理方法作为背景。后来被人们总称之为线性预测分析法（Linear Predictive Coding：LPC），构成了一个大的学科体系。除了属于这种 LPC 的方法外，还开发了各种数字声音处理方法。到目前为止，相继实现了声音编码、声音分析、声音合成、声音识别、话者识别等各种具体系统。人与人之间、人与机器之间声音信息处理的流程图，如图 1.1 所示。

与声音信息处理相关的著作已经出版不少[2]—[7]。本书是从数字处理的观点出发，以比较新的技术作为中心，为了今后研究声音的方便，就不可缺少的技术进行了集中解说。对于预学习研究声音处理的人来说，本书可以作为指导书和教科书；而对于已经从事研究的人来说，本书可以作为参考书。关于古典的论述，可参看其他书籍。对于重要的论述，包括公式的展开，我们尽量做到论证详细。数字声音处理技术种类繁多，但它们相互关连。所以，理解它们相互之间的关系，对于研究声音是极其重要的。因此，全书的叙述方法尽量统一、相关的内容尽量进行说明。即使相关的参考文献，只要是重要的部分，也尽量采用。

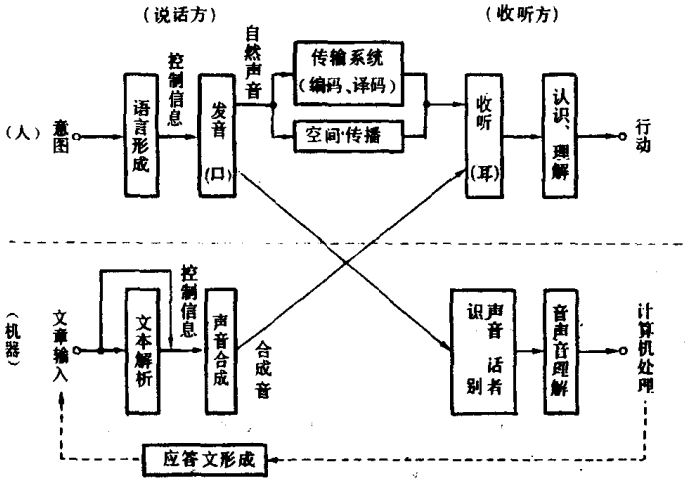

图 1.1　人与人之间、人与机器之间声音信息处理的流程图

本书的第二章到第四章是论述数字声音处理的基本技术；第五章到第九章，分别论述线性预测分析、声音波形编码、声音合成、声音识别及话者识别等重要技术；最后第 10 章概括性地介绍了今后的研究课题。如上所述，数字声音处理的各种技术是相互有联系的，因此，我们希望读者从第二章起，逐章阅读。另外，关于声音波的叠加噪声除去和回波抑制等问题，也是重要的领域，但由于篇幅所限，本书也只得省略。

关于数字声音处理的重要基本技术—傅里叶变换（FFT、DFT）、Z 变换、频谱推定等，请参考这套丛书的第二卷《数字信息处理的基础理论》；关于硬件及 LSI 技术请参考第三卷《数字信号处理系统》及第五卷《SC 电路网的设计和应用》；关于数字滤波器的设计，请参考第四卷《数字滤波器的设计》；关于与通信技术相关的部分，请参考第八卷《数字通信技术》。

第二章　声音的基本性质

2.1　声音和语言

人类用于交流的手段中，声音是最基本、最有效、且最重要的手段。声音中包含着说话者是谁的个性信息、表现说话者感情的情绪性信息等。显然传递到对方的意思内容、即语言信息是最重要的信息。具有语言能力以及制作工具的能力，是区别人类与其他动物的基本所在。语言的发达与文明的发达是不可分割的。语言中，当然写下来的语言，即书面语言作为知识的交流方法也是最有效的，且是持久的。但是，通过声音交流信息的方法更多。尽管书刊、杂志作为单方面的信息传递手段是有效的，但在信息相互交流方面，它们都无法与声音相比。

人类生成声音语言过程的第一阶段是决定想传给对方的内容是什么，然后将内容变换成语言的形式。选择表现其内容的适当语句，将它按文法规则排列，便能构成语言的形式。由大脑对发音器官发出运动神经指令，发音器官的各种肌肉运动，振动空气而形成声音波。这个过程可划分为神经和肌肉运动的生理学阶段和产生声音波、传递声音波的物理阶段。作为物理现象的声音性质是连续的，而作为传递信息的编码体系的语言却由具有离散特性的单位构成。

形成文章的基础是单词（word），各单词由音节（syllable）组成，音节由音素（phoneme）组成，音素有元音（vowel）和

辅音（consonant）。音节的定义不一定明确，但是，一个音节一般可以是1个元音和1～2个辅音组合。除表现外来语的特殊情况以外，日本语有101个音节，并且对应着各自的假名。日本语有5个元音和20个辅音。一种语言所用的音素数一般都不超过50个。实际上，各种音素组合而构成语言时的连接方法，有几种限制，并不是所有的组合都存在。因此，一种语言中所用的音节数，远少于音素的组合数。

世界各国的语言，所用的单词种类非常多。而且经常出现新的单词，但比起可能的音素和音节的组合数，它们却少得多。其中，经常使用的单词一般不超出2000～3000个，一般人使用的单词数在5000～10000个范围内。

重音（stress）和语调（intonation）也是构成语言学的一部分，它们或者用来表示一句话中重要的单词，或者用来表示疑问句，或者用来表示说话人的感情。重音和语调是一种附加的信息。

如上所述，从语言学的（音韵论的）定义，声音的最小基本单位叫做“音素”，从实际发音所产生声音的声音学的最小基本单位叫做“单音”，前者采用音韵符号，后者采用声音符号，分别用/a/、[a]表示。例如，单音[ɛ]和[e]在日本语中解释为同一音素/e/。但是，在法语中，却解释为不同的音素/ɛ/和/e/。

2.2 听觉和声音

发音的目的就是让对方听懂并且理解这种声音。因而发音方法与人的听觉能力密切相关。人的听觉能力，既有人工智能机器无法模仿、高能力的一面，也有无能为力的一面。所谓高