

语 信 处 言 号 理

陈永福 王仁寿

2.3

中国科学技术大学出版社

内 容 简 介

语言信号处理是一门由语言学和数字信号处理相结合的、新兴的边缘学科。它主要用于现代通信领域，旨在解决人与人之间以及人与机器之间的通信问题。

全书共分十章。第一章至第三章扼要介绍语言学和数字信号处理的基础知识，以及汉语语言学有关内容。第四章至第六章介绍语言信号处理的基本方法，包括时域、频域、倒谱域、线性预测、矢量量化等。第七章、第八章讨论语音通信中的波形编码和声码技术。第九章讨论语言合成。第十章讨论语言识别。

本书的特点是结合汉语情况，总结了国内外的最新进展，既有较强的理论系统性，又体现一定的应用观点。可作为信号、电路与系统，通信与电子系统，计算机信息处理等学科有关专业的研究生教材，也可作为从事这一领域科研工作的技术参考书。

语 言 信 号 处 理

陈永彬 王仁华

责任编辑：于文良 封面设计：王瑞荣

※

中国科学技术大学出版社出版

(安徽省合肥市金寨路96号)

中国科学技术大学印刷厂印刷

安徽省新华书店发行

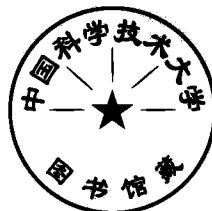
※

开本：787×1092/16 印张：18.25 字数：442千

1990年5月第1版 1990年5月第1次印刷

印数1—4000册

ISBN7-312-00131-9/TN·7 定价：3.70元



前 言

语言信号处理是一门由语言学和数字信号处理相结合的、新兴的边缘学科，它主要用于现代通信领域。这里所指的通信是广义的，即：人至人的通信，机器至人的通信和人至机器的通信；用学术术语来讲，就是包括：语音通信、语言合成和语言识别与理解三个分支。

语言是人类社会特有的一种信息系统。人能够用嘴讲话（发出语言），又能用耳朵听话（接收语言），所以语言是一种最方便的交换信息的形式，语言信号处理在现代信息社会占有重要地位。诸如：数字电话通信、高音质的窄带语言通信系统、报时报警系统、语言学习机、声控打字机、自动翻译电话系统、智能机器人、新一代计算机语声智能终端以及许多军事上的应用等等，都要用到语言信号处理技术。现代集成电路技术的飞跃发展，大大推动了语言信号处理系统的实用化。可以说，开发语言信号处理的各种应用的时代已经来临。

近年来，我国在语言信号处理方面的研究发展十分迅速，并且已取得了一些可喜的成果，许多高等院校都相继开出这一门研究生课程。因此，目前极需要一本教材和技术参考书。国内外虽有几本有关书籍，如：Rabiner 等著的 *Digital Processing of Speech Signals* (1978)、古井贞熙：*デジタル音声処理* (1985)、D.O'schanghnessy: *Speech Communication—Human and Machine* (1987) 等，但是，这些书有的已嫌旧些，并且都没有结合汉语语言学，使用起来总感到有些欠缺。我们自 1982 年以来就开出了这门研究生课程，并持续不断地进行这方面的科学研究工作。现在，在 1982 年开始使用的讲义（陈永彬编）的基础上，结合最近国内外某些具有代表性的科研成果，写成这一本书以飨读者。它既可作研究生教材，也可作为从事这一领域科学研究工作的技术参考书。

全书共分十章。第一章至第三章介绍一些要用到的语言学和数字信号处理的基础知识，以及汉语语言学有关内容。第四章至第六章介绍作为语言信号处理的基础技术的各种分析方法，包括时域、频域、倒谱域、线性预测、矢量量化以及汉语音调声调分析等方法。第七、八章讨论语音通信分支中的各种编码技术和声码技术。第九章讨论语言合成技术。第十章讨论语言识别技术。我们认为，这些章节包括了语言信号处理的基本内容，以及国际上某些最新的有关进展；同时又结合了汉语情况，对国内读者更为适用。作者还曾将其中部分内容在澳大利亚等国的高校、研究所进行过讲学，证明选材也是比较恰当的。

本书的分工基本上是：第一、二、三、四、六、十章由陈永彬执笔，第五、七、八、九章以及 4-1、4-7、10-3、10-6 节由王仁华执笔；但也有一些交叉之处。全书由陈永彬统稿。由于语言信号处理是一门新兴的学科，目前仍在进展之中；所以本书定有不尽之处。此外，错误也在所难免。作者热烈欢迎读者给以指正，以便将来进行修改。

作 者

1989 年 3 月

目 录

第一章 语言学与数字信号处理	(1)
1.1 言语过程及其作用	(1)
1.2 语言学的研究范围	(3)
1.3 语言与文字的关系	(4)
1.4 语言的分类、起源、演变及发展	(6)
1.5 信号处理与数字信号处理	(11)
1.6 语言信号处理及其应用	(12)
第二章 语言的生理、心理及其信号模型	(16)
2.1 语音发音器官的生理	(16)
2.2 语音听觉器官的生理	(18)
2.3 神经系统的生理	(20)
2.4 语音发音的心理	(22)
2.5 语音听觉的心理和两种听觉实验	(24)
2.6 语言信号的数学模型	(27)
2.7 声管模型	(30)
2.8 共振峰模型	(33)
第三章 汉语语音学	(37)
3.1 汉语语音的特点	(37)
3.2 汉语拼音方案	(37)
3.3 汉语的音素	(39)
3.4 汉语的音节及其结构	(43)
3.5 汉语语音的某些特性	(51)
3.6 汉语语音的质量评价	(60)
第四章 语言信号的数字分析	(65)
4.1 语言信号的数字化与预处理	(65)
4.2 短时平均能量, 振幅和跨零数	(69)
4.3 短时自相关函数和平均幅差函数	(74)
4.4 短时傅里叶变换	(76)
4.5 短时傅里叶反变换	(81)
4.6 短时倒谱分析	(89)
4.7 音调检测与汉语声调判决	(99)
第五章 语言信号的线性预测分析	(106)
5.1 线性预测的基本原理	(106)
5.2 线性预测分析的解法	(112)
5.3 斜格法及其改进	(116)

5.4	线性预测分析应用	(123)
5.5	线谱对分析	(129)
5.6	极零模型	(133)
第六章	语音信号的矢量量化	(139)
6.1	矢量量化的基本原理	(139)
6.2	失真测度	(140)
6.3	最佳矢量量化器和码本的设计	(144)
6.4	降低复杂度的矢量量化系统	(146)
6.5	有记忆的矢量量化	(148)
6.6	语言波形的矢量量化	(149)
6.7	语言参数矢量量化	(151)
第七章	语音通信之一——编码	(155)
7.1	语言信号的压缩编码	(155)
7.2	脉冲编码调制	(157)
7.3	自适应预测编码	(163)
7.4	频域编码	(173)
7.5	数字语音插空	(182)
第八章	语音通信之二——声码器	(187)
8.1	相位声码器和通道声码器	(187)
8.2	同态声码器	(192)
8.3	线性预测声码器	(195)
8.4	新一代声码器	(199)
8.5	声码器硬件	(205)
第九章	语言合成	(211)
9.1	语言合成原理	(211)
9.2	共振峰合成	(215)
9.3	线性预测合成	(223)
9.4	汉语按规则合成	(225)
9.5	语言合成的应用举例	(230)
第十章	语言识别	(238)
10.1	概述	(238)
10.2	端点检测技术	(241)
10.3	动态时间规正	(246)
10.4	有限状态矢量量化技术	(250)
10.5	隐马尔柯夫模型技术	(253)
10.6	讲话者识别	(260)
10.7	孤立字(词)识别系统及其应用	(264)
10.8	连续语言识别系统	(278)

第一章 语言学与数字信号处理

在讨论学科研究的方向时，常常提到“边缘学科”这个词。所谓边缘学科是指由两个或若干个学科互相交叉的一些新兴学科。现代科学技术发展的总趋向之一就是：那些过去认为是独立的一些学科互相溶汇、结合成多边缘的交叉学科，或者称为科学技术的整体化。因为在这些交叉点上，最可能发出新芽，生长出果实。所以常常可以看到若干学科的学者们合作起来进行某一新学科的科学研究的状况。

语言信号处理就是属于这样一门新兴的边缘学科。而且，它可能主要是语言学和数字信号处理这两个学科相结合的产物。所以，这主要是语言学家和信号处理学家的合作领地。当然，作为一个好的语言学家或者信号处理学家，他们还必须与其它与之有关的学科的学者们进行合作。

现在，让我们简短地介绍一下与语言学以及数字信号处理这两个学科有关的一些问题。

第 1.1 节 言语过程及其作用

语言学是以人类的语言为研究对象的一门科学。语言是从千百万个人的言语中概括总结出来的规律性的符号系统。所以，研究语言首先要了解一下人的言语过程。

人们往往认为我们能够说话和听懂言语是理所当然的事，并未想过进一步去理会言语的过程，就像人们平时并不理会自己的心脏、大脑等器官的动作过程一样。其实，从说话者说一句话给听者听到并理解的过程，虽然十分短暂，但其中却包含了相当复杂的过程。言语过程可以分为五个阶段，如图 1.1 所示：

(1) 想说阶段。人的说话首先是客观现实在大脑中的反映，经大脑的决策产生了说话的动机，接着说话神经中枢选择恰当的单词、短语以及按语法规则组合，以表达他想说的内容和情感。这个阶段与大脑中枢的活动有关。

(2) 说出阶段。由上阶段中枢的决策，以脉冲形式向发音器官发出指令，使舌、唇、颚、声带、肺等部分肌肉协调地动作，发出声音来。与此同时，大脑也发出一些指令给它有关器官，使之产生各种动作来配合言语的效果，如：面部表情、手势、身体姿态…等。另外，还开动了另一个“反馈”系统，来帮助修改言语。这就是：他不但发出言语，而且他自己的听觉系统也在听自己的言语。但是，在这个阶段中，主要与发音器官的活动有关。

(3) 传送阶段。说出来的言语是一连串声波，凭借空气为媒介传送到听者的耳朵里。当然，有时遇到某种阻碍或其它声响的干扰，使声音产生损耗或失真。这阶段中，主要是传送信息的物理过程起作用。

(4) 接收阶段。从外耳收集到的声波信息，经过中耳的放大作用，到达内耳。经过内耳基底膜的振动，激发柯替氏器官内的神经元使之产生脉冲，将信息以脉冲形式传送给大脑。在这个阶段中，主要是与听觉系统的活动有关。

(5) 理解阶段。听觉神经中枢收到脉冲信息之后，通过一种至今尚未完全了解的方式，辨认出说话的人及其所说的信息，从而听懂了讲话者的话。

从五个阶段来看，言语的过程包含着相当复杂的因素，其中有心理的、生理的、物理的以及个人的和社会的因素。这里，个人的因素是指讲话者的口音和用词造句的特色以及听话者的听音和理解的能力；社会的因素则是指讲话者和听话者对于进行交际的手段有共同的理解的社会基础。

下面，让我们讨论言语的作用。言语是人类的交际工具和思维工具。

人类社会愈来愈显示出信息社会的特点。不但在人与人之间，而且在人与机器之间每时每刻都需要进行大量的信息交换。通信或信息交换，已成为人类社会存在的必要条件，正如衣食住行对于人类是必要的一样。

虽然，嗅觉、触觉...等也是人类固有的感觉，人们也可以从中得到某些外界信息；但是可以指出，最重要的、最精细的信息源只有图象和语言两种。从物理上说，就是利用光波和声波直接携带信息并进行信息的交换。

然而，对于光波来说，正常的人不能由其身体的某些器官主动地发出光波信息；这是因为要发出光波，需要原子能级的振动，而这是人体所无法支持的，因而，人类在长期进化中，只能用眼睛被动地接收外界光波信息——图象。

但是，对于声波来说，正常的人既可以用口主动地讲话，这只需要分子能级的振动；又可以用耳被动地接收外界的声波信息——言语。所以，言语是人类最重要的、最有效的、最常用的和最方便的通信方式。

随着现代科学和计算机技术的发展，除了人与人之间的上述自然语言的通信方式之外，人机对话及智能机器人等领域也开始使用语言了。这些人工语言同样有词汇、语法、句法结构和语义内容等。控制论创始人维纳（Norbert Wiener）在 1950 年就预先指出过：‘通常，我们把语言仅仅看作人与人之间的通信手段；但是，要使人向机器，机器向人以

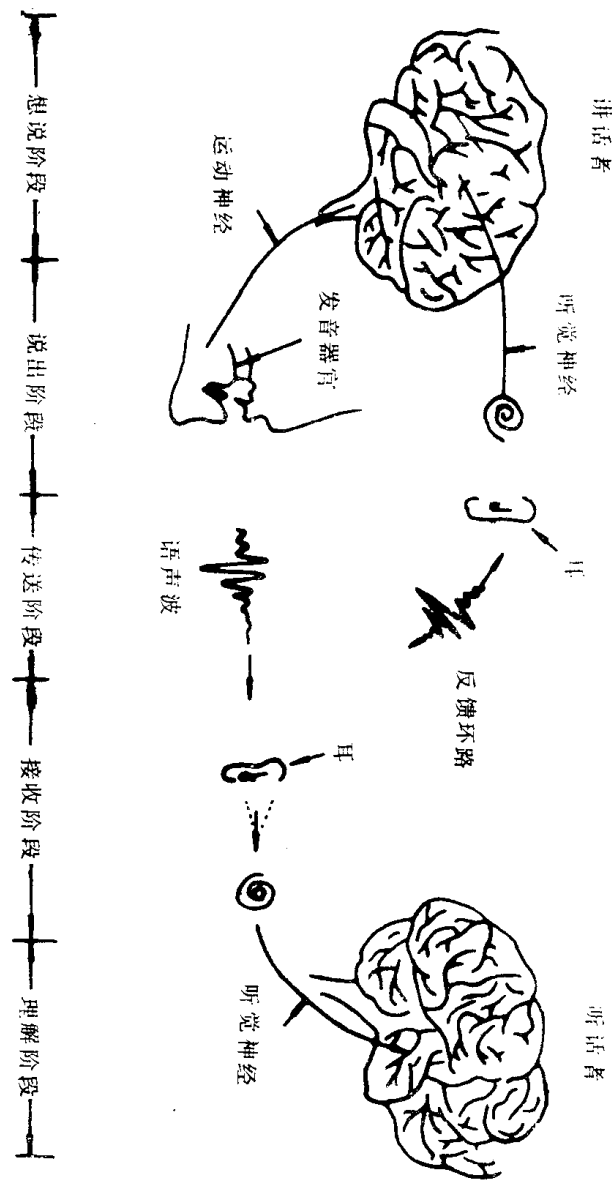


图 1.1 言语过程的五个阶段

及机器向机器讲话，那也是完全办得到的。”这个预见实现的日子终于即将到来。

思维和交际是分不开的。言语过程的第一阶段中就包括了选择和组织言语词汇和句子的思维过程。不仅如此，单纯的思维过程也与语言有关，只是它不象外部语言那样，而是用一种不出声的内部语言。有人做过这样一个实验：用微电极插入舌头或下唇肌肉内，当默算一个数学题或默诵一首诗歌时，所得到的电位变动的记录，与口算这个数学题或朗诵这首诗歌时所得的记录，是完全相同的。

第 1.2 节 语言学的研究范围

语言是由言语中概括出来的一个符号系统。有“形式”和“内容”两个方面的内容，即语音的形式和语义的内容。将这两个语言的基本要素结合起来，可以构成言语的语素、词、短语和句子等言语的不同层次的单位；这个构成规则就是语法。以汉语的一个句子“他热爱祖国”为例：

语音形式：tā rè ài zǔ guó.

语义内容：他 热 爱 祖 国。

这个句子切分出来的三个词的语音形式就相应地表达了这三个词的语义内容。

语言学是一门既是最古老的、又是最年轻的学科之一。研究表明，自从人类从能劳动而分化出来时，就开始有了语言。通常认为，语言至少有几十万年、甚至一百万年的历史了。顺便提一下，语言的出现要比文字早得多，文字的发明可能与农业的产生有关，而农业的发明只是差不多一万年前的事，也就是说，文字至多只有一万年的历史。

虽然，语言的产生和对它的研究比天文学和数学开始得更早。但是，世界学者们公认，从十九世纪开始，人们才开始从各个角度认真地去研究语言，才出现了现代语言学这门科学。由于语言牵涉到人类的种种活动和行为，要涉及到许多学科，所以语言学是一门相当复杂的多边缘学科。最近十几年来出现了这么一个词，叫做“连接号语言学”(Hyphenated Linguistics)。例如：

- (1) 语言—信号处理，研究以语音为形式、以语义为内容的语言信息的编码、传递、储存、合成和识别等技术。
- (2) 语—音学，研究语言声波的各种性质。尤其其中的一个分支，实验语音学，用各种实验方法来研究语音，取得了不少成果。
- (3) 数理—语言学，研究语言符号系统的数理逻辑，尤其注意语义在推理中的作用。
- (4) 心理—语言学，研究言语时，大脑和神经系统的心理活动。
- (5) 生理—语言学，研究言语时，各种发音器官和听觉系统的生理活动。
- (6) 社会—语言学，研究语言的应用对于各种社会因素的依从关系。
- (7) 教育—语言学，研究如何用母语或外语来传授语言能力和语言知识的问题。
- (8) 地理—语言学，研究语言的地理分布，并绘出方言地图。
- (9) 人类—语言学，研究古代的遗迹和人类头骨及发音器官的化石，探讨语言的起源和历史演变的过程。

语言学在研究语义和语言起源时，还与生物学、动物学等有关。限于篇幅，这里简短地介绍几个与语言学有关的研究情况。

第 1.3 节 语言与文字的关系

如果人类选择用语音形式的符号系统来表达语义内容——思想和感情，那么，记录语音的文字，就是第二次符号——符号的符号。这是由于人类发觉语言有其局限性——暂时性。这个局限性使语言在时间上和空间上都受到了限制。所以，随着社会的进化，尤其农业的发生，出现了文字。

文字是从语音中发展出来的，但是不同语言发展出来的文字，所走的道路是不同的，因而与语音的关系与距离是不同的。也许可以分为四大类文字：

(1) 与语音关系最密切、距离最近的大概要算朝鲜的谚文。它与朝鲜语语音的区别特征有关。写的时候，元音与辅音分开；元音中怎样写是属于前元音、怎样写又是后元音；怎样写的辅音是送气的、怎样写又是不送气的…，在谚文中都能表达出来。

(2) 其次，是印欧系语言的文字。这些文字中的单位是音段 (Segment) 而不是区别特征。当然，各种文字与音段的关系程度也不相同。例如，英文的音段大部分是用语素音位 (词音位, Morphophoneme) 表达的。如：sane 和 sanity 这两个英文的词中的字母 a，前者表示 [ei]，后者表示 [æ]，代表了两个不同的音位。这是因为要保持词、词位和语法之间的简单关系，所以读音变了、音位变了，而文字却没有变。又如：electric 和 electricity 两个词中最后的字母 c，前者为 [k]，后者为 [s]，但是为了保持词形的联系，文字上没有作相应的变化。

(3) 日文。这是用假名来作为语音单位的文字。这种单位比印欧系语又大了一些。因为假名是表示音节的。例如：假名 kana 这个字，日文写为かな，有两个音节，而每个音节包括了两个音段。日文是典型的音节文字。

(4) 中文。各种文字中距离语音最远的就是汉字。汉字是象形的语义的文字。当然，它也有一定的语音扩散的关系。例如，亭、停两个字，后者的读音与前者相同，但意义却完全不同。这就是“同音字”。而且，由于汉字的文字改革，与语音联系时可用“拼音文字”来书写。

下面，让我们讨论汉语与汉字的情况。

汉语在中国各地的读音相当复杂，有许多种方言。但是，文字却是统一的；也就是说，虽然一个汉字可以有多种读音，但都可以阅读同样的书报。这完全是由于汉字与语音的关系不怎么密切，可以直接由象形字或形声字联系到意义的缘故。汉字的这个特点为中国文化能高度一致有很大的贡献。

世界上大部分地区或国家的文字和语义都不直接发生关系，它们之间都是通过语音发生关系的。已经发现有一种病症，叫做“失读症” (Alexia)，轻一点的叫做“诵读困难症” (Dyslexia)。患这种病的儿童，说话没有困难，但到了十几岁还不能认字和阅读。中国就很少这类问题。

关于这个问题，有两篇文章值得介绍一下。1968 年日本学者 K. Makita 发表在美国《行为精神病学学报》 (America Journal of Orthopsychiatry) 上文章名为“日本儿童丧失阅读能力的实际情况” (The Rarity of Reading Disability in Japanese Children)。他认为，由于日本学的是汉字和假名，所以日本儿童根本就不存在这个问题。这篇文章曾引起

一些争论，后来，1971年3月，一位美国心理学家 P. Rozin 在美国的《科学》(Science) 杂志上发表了另一篇文章，题目是：“有阅读问题的美国儿童能容易地学会阅读汉字表达的英语”(American Children with Reading Problems Can easily Learn to Read English Represented by Chinese Characters)。他在美国东海岸的一所有较多患有失语症的儿童小学中做了实验。他用汉字代替英文，教他们用英语语音来阅读，例如：文字“妈妈放书”读为“The mother put the book”。也就是让文字与语音分了家，实验相当成功。他认为，因为汉字是不怎么依赖语音的，所以小孩不必一个字母一个字母地去拼读，从而避开了失读症的毛病。也就是说，这个实验证实了中国汉字与语音的关系是较远的。

此后，近年来又有一些心理学家和学者就这个问题作了更深入的各种实验，得到了类似的结论。[注]

[注]：有四篇文章，介绍了这些实验，值得提一下。

(1) 日本女学者 Sumiko Sasanuma, 1974 年在《中国语言学报》^[1]上发表了一篇文章，题目：“日本失语症患者在书面语上的障碍：假名和汉字的处理”。她研究一种由于大脑受伤引起的失语症 (Aphasia)。有二类病人，一类是大脑的卜洛卡 (Broca) 区受伤，患者，43 岁的工程师，伤后能写汉字 (日本文字由汉字和假名混用的)。请他写“那个女人在写信”。日文应为：“女の人が手紙を書いている”，他却写成“女の人手紙书”。这在日文是完全不通的。这表明，比汉字更为简单的假名他却错了。这个实验说明他对语音拼写文字的能力丧失了。再叫他写假名ど，却写成の (do: 写为 no:)，错误就在语音的区别特征上。另一类是大脑的维尔尼克 (Wernick) 区受伤，情况则相反。这是一个 29 岁的中学生，写假名没有问题，但写不出或写错汉字。例如，写一个只有训读的“田舍”应写为いなか (inaka)，他却写成でんしゃ (densha)。这个实验证实了 Broca 区管语音；Wernick 区管图形文字；左脑管语音；右脑管文字。但这仅是初步的看法。

(2) 美国华人 Ovid J. L. Tzeng (曾志朗)：1977 年在《实验心理学学报》上发表了一篇文章^[2]，题目为“阅读汉字时言语的再编码”。实验方法是：先放映汉字，然后再听录音，看看他受语音干扰后记忆情况。如果汉字的确不通过语音，那么听录音不会干扰幻灯的认字和第三段的记忆。测试的字有三组：SC (同辅音的字) 如：克、康…；SV (同元音的字) 如：七、吉、西、你…；SVCV (同辅音同元音的字) 如：石、示、十、事、史…，得出结论是：(1) 汉字也通过语音，因为它也受到语音的干扰。(2) 尤其元音的干扰比较厉害。(3) 同组字的干扰最大。

(3) R.A.Treiman, J. Baron and Kenneth Luk (陆仲真)：1981 年《中国语言学报》，题目：“文字，语音与阅读理解：中英文比较”(Speech Recoding in Silent Reading: A Comparison of Chinese and English)。两组人 (美国人和中国人) 接受实验，如听：A pier is a fruit (码头是一种水果)，(干扰：pear 梨)；如听：煤是一种水果 (干扰：梅)。此实验证实，汉字与英文都与语音有关，但英文的关系比较密切。

(4) 曾志朗、洪兰等人在《语言》一书的“汉字的视觉”中，做了一个称为“两眼不同视觉实验”(Dichoptic Experiment)^[3]。结论是：右大脑管总体上不需分析的东西，左大脑管需分析判断的东西。

第 1.4 节 语言的分类、起源、演变及发展

世界上有多少种语言？最保守的估计有几百种。据东德《语言学及语言交际工具问题手册》的介绍有五千六百五十一种。但有些正在衰亡，如：澳大利亚有 250 种语言正在衰亡，只有四万人使用…。但也有些语言一直在使用，尽管用的人数很少，如天主教会以拉丁语为正统语言，医生处方则也用拉丁语。人数方面来说，用汉语的人有十亿，用英语的有两亿；而世界上还有 3/4 的语言没有文字。

那么，应如何将语言加以分类？这叫做谱系分类法，就是根据语言历史来源或亲属关系来对语言进行分类。分为语系、语族、语支。基本的分布是：

亚洲：汉藏 (Chinese-Tibetan) 语系、阿尔泰 (Altai) 语系、伊比利亚—高加索 (Iberian-Caucasian) 语系、马来—玻里尼亚 (Malaya-Polynesian) 语系、达罗毗荼 (Dravidian) 语系、印欧 (Indo-European) 语系、含—闪 (Hamito-Semitic) 语系。

欧洲：印欧 (Indo-European) 语系、乌拉尔 (Uralic) 语系。

非洲：含—闪 (Hamito-Semitic) 语系 (主要是阿拉伯语)、尼日尔—刚果 (Niger-Congo) 语系、沙里—尼罗 (Chari-Nile) 语系、科依撒 (Khoisan) 语系、印欧 (Indo-European) 语系。

美洲：印欧 (Indo-European) 语系、印第安 (Indian) 语系。

澳洲：印欧语系中的英语和土著的语言。

下面，只列举最大的四个语系的分支情况。

一、汉藏 (Chinese-Tibetan) 语系

1. 汉语 (Chinese) 语族：汉语，使用人口十亿。

2. 壮侗 (Chuang-Tung) 语族：

(1) 壮泰 (Chuang-Tai) 语支：壮语 (Chuang)、布依语 (Puyi)、泰语 (Tai)、泰语 (Thai)、老挝语 (Lao)、单语 (Shan)。

(2) 侗水 (Tung-Shui) 语支：侗语 (Tung)、水语 (Shui)、么佬语 (Mulao)、毛难语 (Maonan)。

(3) 黎语 (Li)。

3. 苗瑶 (Miao-Yao) 语族：

(1) 苗语支：苗语 (Miao)。

(2) 瑶语支：瑶语 (勉语) (Yao)。

4. 藏缅 (Tibetan-Burmese) 语族：

(1) 藏语支：藏语 (Tibetan) (有十三种方言)、嘉戎语、阿卡语。

(2) 缅甸语支：缅甸语 (burmese)、库启钦语。

(3) 彝语支：彝语 (Yi)、傈僳语 (Lisu)、纳西语 (Nakhi)、哈尼语 (Hani)、拉祜语 (Lahu)、白语 (Pai)、阿昌语 (Achang)。

(4) 景颇语支：景颇语 (Chingpaw)、波多语、拿加语。

5. 越南语：未定，似应属此语系。

二、印欧 (Indo-European) 语系

1. 印度 (Indic) 语族: 印地语 (Hindi)、乌尔都语 (Urdu)、孟加拉语 (Bengali)、
茨岗 (吉普赛) 语 (Romany) 等二十多种。使用人口近五亿。
2. 伊朗 (Iranian) 语族:
 - (1) 东部语支: 阿富汗语 (Afghan)、沃舍梯语。
 - (2) 西部语支: 波斯语 (Persian)、库尔德语 (Kurdish)、塔吉克语 (Tadzhik)。
3. 斯拉夫 (Slavic) 语族:
 - (1) 东部语支: 俄语 (Russian)、乌克兰语 (Ukrainian)、白俄罗斯语 (Belorussian)。俄语使用人口两亿少一些。
 - (2) 南部语支: 保加利亚语 (Bulgarian)、马其顿语 (Macedonian)、塞尔维亚—霍
尔特语 (Serbian—Croatian)、斯洛文语 (Slovenian)。
 - (3) 西部语支: 波兰语 (Polish)、捷克语 (Gzech)、斯洛伐克语 (Slovak)。
4. 日耳曼 (Germanic) 语族:
 - (1) 东部语支: 峨特语, 已消亡。
 - (2) 西部语支: 荷兰语 (Dutch)、德语 (German)、英语 (English)、佛兰德语 (Flemish)、伊狄语 (Yiddish)。英语使用人口有两亿多人。
 - (3) 北部语支: 瑞典语 (Swedish)、丹麦语 (Danish)、挪威语 (Norwegian)、冰岛
语 (Icelandic)。
5. 塞尔特语族:
 - (1) 高卢语, 已衰亡。
 - (2) 布里索尼 (Brythonic) 语支: 布列塔尼语 (Breton)、威尔士语 (Welsh) (均已
衰亡)。
 - (3) 盖尔语支: 爱尔兰语 (Irish)、苏格兰语 (Scottish)、曼语 (Man)。
6. 波罗的 (Baltic) 语族:
 - (1) 东部语支: 立陶宛语 (Lithuanian)、拉脱维亚语 (Latvian)。
 - (2) 西部语支: 古普鲁士语 (Prussian) (已衰亡)。
7. 意大利 (Italian) 语族:
 - (1) 奥斯干—温布利安语支: 奥斯干语、温布利安语。(均已衰亡)。
 - (2) 拉丁语支: 后来发展为拉丁语族。(又称为罗曼语族)。
8. 拉丁 (Latin) 语族:
 - (1) 伊比利亚—拉丁 (Iberian—Latin) 语支: 西班牙语 (Spemish)、葡萄牙语 (Portuguese)、卡达伦语 (Gatalan)。
 - (2) 高卢—拉丁语支: 法语 (French)、普鲁旺斯语 (Provençal)。
 - (3) 东部语支: 罗马尼亚语 (Rumanian)、摩尔达维亚语 (Moldarian)。
 - (4) 意大利语 (Italian)。
 - (5) 撒丁语 (Sardinian)。
 - (6) 勒多—罗曼语 (Rhaeto—Romanic)。

三、阿尔泰语系

1. 突厥 (Turkic) 语族:
 - (1) 布尔加尔语支: 楚瓦什语 (Churash)。

- (2) 奥古慈语支：土耳其语 (Turkish)、土库曼语 (Turkmen)、阿塞拜疆语 (Azerbaijani)、撒拉语。
- (3) 克普恰克语支：哈萨克语 (Kazakh)、塔塔尔语 (Tatar)、巴什基里亚语 (Bashkilian)、吉尔吉斯语 (Kirgiz)。
- (4) 葛逻禄语支：维吾尔语 (Uigur)、乌兹别克语 (Uzbek)。
- (5) 回鹘语支：裕固语、雅库特语 (Yakut)、图瓦语 (Turinian)、绍尔语、哈卡斯语 (Khakass)。
2. 蒙古 (Mongolian) 语族：本部蒙古语、布里亚特蒙古语 (Buryat)、莫戈勒语、达斡尔语、土族语、东乡语、保安语。
3. 通古斯-满洲 (Tungus-Manchu) 语族：
 - (1) 通古斯语支：埃文尼语 (Evenki)、鄂温克语、鄂伦春语等。
 - (2) 满洲语支：满语 (Manchu)、纳奈语 (赫哲语)、锡伯语 (Siba)。朝鲜语 (Korean) 似应属于此语系。

四、含一闪 (Hamito-Semitic) 语系

1. 闪米特 (Semitic) 语族：
 - (1) 东部语支：阿长德语、巴比伦语 (Babylonian)。(均已衰亡)
 - (2) 北部语支：古伽南语、腓尼基语、古希伯来语 (Hebrew)。(均已衰亡)
 - (3) 南部语支：阿拉伯语 (Arabic)、马尔他语 (Maltese)、埃塞俄比亚语 (Ethiopic)。
2. 柏伯尔 (Berber) 语族：柏伯尔语、什卢赫语 (Shluh)、塔马齐格特语 (Tamazight) …，分布于北非的摩洛哥、阿尔及利亚等地。
3. 库施特 (Cushitic) 语族：盖拉语 (Galla)、锡达莫语 (Sidamo)、索马里语 (Somali) …等。
4. 埃及 (Egyptian) 语族：古埃及语 (已衰亡)、科普特语 (Coptic) (一种宗教语言)。
5. 乍得 (Chadic) 语族：豪萨语 (Hausa)、科托科语。

中国语言分布：

- (1) 汉藏语系：除汉语外，有二十多种。
- (2) 阿尔泰语系：有十七种。
- (3) 马来-玻里尼西亚语系：台湾高山语。
- (4) 印欧语系：新疆的俄语。
- (5) 朝鲜语、京语。(未定属于何种语系)。

人类的语言是如何起源的？又是如何演变的？

1976年，美国纽约州的科学院主持召开了一次规模很大的学术会议——语言和言语的起源与演变的学术讨论会。有几千人参参加，有十几个学科有关的内容，并发表了一本论文集^[4]。从这个论文集中可以看出三个问题的研究成果：

1. 语言起源的时间。地球形成时间为46亿年前。生物出现为36亿年前。3亿年前出现恐龙，6000万年前出现灵长目动物，2000万年前出现猿类。1000万年前开始人和黑猩猩

猩与猿类的分化。42 万年前北京周口店猿人开始用火。所以大致可以说，语言的发生大约是在几十万年至 100 万年前发生的。

2. 语言与语音的关系。由于地球气候变化，古猿的一部分灭绝；一部分找到新森林进化为现代的黑猩猩、大猩猩等。另一部分则移居平原成为人类的远祖。古猿用四只脚走路，黑猩猩用三只脚走路，猿人则只用两只脚走路了。这一点与语言发生有密切关系。用四只脚时，嘴要吃东西、移动东西、打架，负担很大，不可能发出各种不同的声音，产生语言的可能性很小。站起来之后，拿东西、打架可以用手了，嘴逐渐变成进行交际的工具。同时，由于站立，喉头受地心吸力的影响逐渐下移，除口腔之外形成了咽腔，增加了发音的能力。例如，元音[i]，[u]，[a]都需要声门相当低才能发出来，[k]，[g]，[ŋ]这一类舌根音也是这样，有了发音的条件，才可能有语言的产生。但是，语言不等于语音。手势语、符号语也都是语言。如从这个观点出发，有人已证明黑猩猩也可以有语言；当然，这只是简单的语言。所以除了发声器官之外，大脑是产生语言的根本因素，正是由于人类大脑的发达，才使人类有现在的语言。

3. 语言与大脑的关系。人的大脑最大最重（1500g）。人的大脑皮层面积最大（有皱纹，使脑神经细胞达 140 亿个左右），人的大脑发育快（婴儿大脑 350g，成人 1500g）。过去认为左大脑与语言关系密切，现在发现语言中枢有三个区，各有职能，对于大脑的语言功能哪些是先天的、哪些是后天的，也有人在研究。下面，简单地讨论语言的演变。

大家都承认，在人类社会的演变过程中（包括自然分布的变迁和人为的分开与归并），语言是在变化的。语言的演变包括语音的变化、语义及词汇的变化以及语法的变化等。举一些例子来说明如下。

1. 语音的变化。例如，汉语句子“叶公好龙”，古时读“叶”为[shə]（蛇），现在则读为[jiɛ]。汉语中更重要的是音位的变化，例如，m，n，ŋ，这三个音位，原来在词头词尾都可以用。现在 m 只能在词头，ŋ 只能在词尾使用了。又如，古汉语中的“附”字，读为[bu:]“部”。而不象现在读为[fu:]，也就是重唇音变为轻唇音了。其实，古汉语中所有浊塞音、浊塞擦音和浊擦音，在现代北京话（中国一些方言除外）都已全部清化了。英语中，语音的变化也很多，直至现在也还在变化。如：动词被名词化之后，其读音的变化，affix[ə'fiks]（贴上）变为 a'fix['æfiks]（附件）。这个词在 1612 年字典上已有名词解释，但读音未变；直到 1775 年（150 年之后）的字典里，才注上变化的读音。又如，英语中非重读元音的丢失现象，armory（军械库）这个字，既可读为['a:məri]，也可以读为['a:mri]。又如，英语中不规则动词的过去式，开始向规则动词转化，也用词尾-ed 表示了，有的已完全变化稳定。如 bide[baid]（忍耐）已不太用 bode[boud]，而用 bided[baidid]；reap[ri:p]（收割）已不用 reapt[rept]，而用 reaped[ri:pt]；mow[mou]（割草）已不用 mew[mju:]，而用 mowed[moud]。有的还在变化之中，可以共用两种读法。如 creep[kri:p]（爬行），既可以用 crept[krept]，也可以用 creaped[kri:pt]；leap[li:p]（跳跃），既可用 leapt[lept]，也可用 leaped[li:pt]；dive[daiv]（跳水），既可用 dove[douv]，也可用 dived[daivd]；dream[dri:m]（做梦），既可用 dreamt[dremt]，也可用 dreamed[dri:md]。当然，也有未改变的，如：keep[ki:p]仍用 kept[kept]，know[nou]仍用 knew[nju:]，rise[raiz]仍用 rose[rouz]，lose[lu:z]仍用 lost[lɒst]等。

2. 词义的变化。这有扩大、缩小、转移三种情况。

(1) 扩大: 如汉语中“江”, 原仅指长江, 现泛指江。如英语中“manuscript”(底稿) [ˈmænʃkript] 原指手写的底稿, 现泛指所有底稿, 包括打字机打的底稿。法语中“arriver”原仅指船的靠岸, 现泛指到达。

(2) 缩小: 如汉语中“瓦”, 原指一切土器, 现仅指屋顶上的瓦。如英语中的“meat” [mi:t], 原指一切食物或用餐 (如 green meat 青菜; before meat 餐前), 现仅指“肉”。如俄语中“жена”, 原指妇女, 现仅指妻子。

(3) 转移: 如汉语中“兵”, 原指武器, 现转指人——战士。如英语中 book, 原指一种可用来写字的山毛榉 (beech) 树的树皮, 现转指书本。如法语中 plume, 原指羽毛, 现转指钢笔尖。

此外, 语法上也有许多演变, 这里就不详细介绍了。

下面, 让我们讨论语言学的发展与近况。

中国是世界上最早进行语言研究的国家。早在汉代 (公元前 206 年至公元 220 年, 即 206 B.C. ~ A.D. 220, B.D.: Before Christ, A.D.: Anno Domini) 许慎在《说文解字·序》里, 就已经在许多处提到汉语演变的有关问题了。

语言学的发展大致上可以分为四个时期。

(1) 历史比较语言学时期。本世纪之前, 就不断有人研究语言问题。其方法是从世界各地搜集和整理各种语言材料, 进行分类和具体的分析、归纳与总结, 从而得到某些结论。例如, 证实了印度、伊朗的语言与欧罗巴语言之间的有机联系; 证实了拉丁语、希腊语也属于现代印欧语系等。大约一百年前, 德国的新语法派 (Neogrammarians, 简称为 NG), 已开始研究语音的规律及其变化, 也就是跳出了“比较”的界限, 开始了结构的研究。

(2) 结构语言学时期。以瑞士的语言学家索绪尔 (Ferdinand de Saussure 1857~1913) 为代表, 开始对各种语言的结构以及各种语言之间的关系进行了研究; 同时开始了各种实验, 建立了实验语音学这个学科, 提出了语言的符号系统。他们的最主要贡献在于: 语言在结构上被科学地划分为句、词、音节、音素等不同层次, 也就是说, 在语音方面, 建立了音位学和区别特征理论; 在语义方面, 建立了语素 (morpheme), 语素音位 (morpho-phoneme) 等。此外, 也探讨了语言与思维过程的关系等。这一时期一直延续到 1957 年。在这个时期的中国语言学家, 主要的有: 赵元任 (著有: 论音位标音法的多能性。The Nonuniqueness of Phonemic Solution of Phonetic System)。罗常培 (著有: 普通话语音学纲要。1957 年, 科学出版社) 等人。在五十年代, 已有人研究句法分析。如, Z.S.Harris 已开始提出用转换 (Transformation) 方法把不同句型联系起来。

(3) 转换生成语言学时期。这个时期的代表人物是美国的乔姆斯基 (Noam Chomsky 1928~)。他是 Z.S.Harris 的学生, 其代表作是: 《句法结构》(Syntactic Structures)。这本书宣告了结构语言学从此没落和转换生成语音学的兴起。在此期间诞生的两门学科: 信息论 (C.E. Shannon 提出, 1916~) 和控制论 (Norbert Wiener 提出, 1894~1964), 渗透到语言学中来了, 产生了一些边缘学科。Chomsky 提出, 建立一套语言演绎系统, 他认为语法是一套规则, 根据这些规则 (好象数学里的一套假设一样), 可以推导出一个句子。有限的规则经过反复使用, 可以生成出无限多的合法句子。这就是转换生成语法 (Transformation Generative Grammar, 简称 TG)。这些规则如: 短语结构规则、助

动词转换规则...等。这种用有限规则生成无限句子是人的内在能力的表现。所以，这个时期开始了对人的语言能力的研究，这也促进了对人脑与人工智能的研究。

(4) 目前时期。到了六十年代，Chomsky 已发现他的理论的缺陷。于是，他不断进行规则的修补，（如：提出变异转换，Modification transformation、连接的句子，Conjoining、嵌入的句子，Embedding 等）。到 1968 年之后，已不断有人评论这些缺陷。尤其，发现规则和修补规则都是很抽象的研究，与现实社会的语言之间脱节十分严重；语言是活生生的东西，比逻辑系统要复杂得多，不可能成天关在房子里用画箭头的方法研究出来。目前，美国是语言学研究的中心，投入的人力最多，发表的文章最多。1973 年在美国密执根（Michigan）办了一次语言学讲习班，是一个重要的转折点。会上提出要改变抽象的研究，要研究语言在时间和地理上的变化以及它与社会的关系。从此，思想活跃起来了，各种人有各种不同的想法。如：William Labov 提出变项规则（Variable Rule），企图用它判断语言变化的趋向。此外也出现了一些新刊物，如《大脑与语言》（Brain and Language），《语言与社会，社会语言学》（Language and Society, Socio-linguistics）；不少语言学家开始走入实际，注重语言的应用（如应用语言学 Applied Linguistics 这个学科），有的与医生合作，研究失语症、阅读困难症等病症...等。

初步回顾了语言信号处理的一个支柱——语言学的一些问题之后，让我们开始转向另一支柱——数字信号处理这个学科的情况的介绍。

第 1.5 节 信号处理与数字信号处理

信号处理是研究对携带有信息的信号进行处理，以得到所希望的信息的一门学科。处理就是变换。变换有许多方法，如：光学处理、声变换等。但是，在现代电子技术和计算机技术十分发达的今天，信号处理主要依靠电路的（模拟的）和计算机的（数字软件或数字硬件）技术。

从原理上来看，信号处理的内容或方法基本上有两种：第一种是用电子系统对输入该系统的信号进行滤波，滤除不需要的分量（包括各种干扰以及不必要的那些信号分量）。最简单的是频率滤波器。它将信号中不需要的各频率分量（不需要的信息）滤除，因此，系统的输出信号中仅含有人们希望有的那些分量。例如，受工频干扰的信号输入到一个“点阻滤波器”去，这个滤波器在 50Hz 处有一个尖锐的低谷的幅频特性，因而可以滤除干扰。又如，分配频带的通信系统中（例如单边带系统），也可以将信号通过一个带通滤波器，抑制带外的频率分量...等。从广义上来说，不限于频率滤波，而是由信息论观点，滤除各种不需要的信息，提取出需要的信息。例如卡尔曼滤波，在研究了干扰的统计特性之后，就可以滤掉它们。第二种是用信号分析技术，对信号进行各种方法的分析，从而得出信号中携有信息的各种参数。这样，可以得出信号中所含信息的较全面的了解，从而可以灵活地使用它们来达到人们的目的。分析的方法就是变换。最简单的是频谱分析。例如，用极窄带通的滤波器组来得到信号的各种频率分布的幅度相对值，也可以用快速傅里叶变换的方法来得到信号各频率分量的幅值和相角。又例如，用这个方法可以得出语言信号的基频及其变化、共振峰频率、带宽和幅值及其变化；甚至可以由此画出语谱图、共振峰图、声调语调图等。除了频谱之外，利用信号分析技术还可以得到被分析信号的其它参

数。虽然，频谱（包括随机过程信号的功率谱）是一种最重要的信号参数；但是，诸如倒谱参数、信号模型参数…等参数，在某些场合下，具有特别的用途。例如，寻找石油矿藏时，倒谱参数特别有用，它可以方便地实现某些卷积信号的解卷。又例如，语言信号分析中，自回归（AR）模型信号参数特别有用，它可以使信息被压缩在尽可能小的区间，对储存和传输它们非常有利。

显而易见，数字信号处理是研究用数字技术的方法来对携带信息的信号进行处理的一门学科分支。尽管数字处理的基本原理，早在一个世纪之前就已有人提出过，但是，直到本世纪六十年代之前，人们却一直沿用模拟的方法去进行信号处理，这主要由于电子计算机技术的发展，包括各种运算方法和集成技术的发展的结果。

模拟处理有其重要的优点，这就是直观、实时。例如，两个模拟信号的相加或相乘的运算，几乎是在整个时间变量范围内即时完成的，并立即显示出运算的结果。现代模拟计算机求解高阶方程的能力也是数字电子计算机所望尘莫及的。所以，目前有一种混合计算机（Hybrid Computer）（这实际上是一台模拟计算机和一台数字计算机，其间设置了通信接口），它能充分利用模拟与数字的各自优点，在空间技术中发挥了很大的作用。但是模拟处理有其缺点，例如用电感或电容作为储存记忆部件体积庞大，更重要的是模拟处理难于达到先进的技术指标。例如，模拟频谱仪难于做到 10Hz 以下，而数字的频谱仪已可达 10^{-3} Hz。

数字处理则相反。其优点是：只要字长及时间许可，可以达到很高的技术指标，存储部件简单灵活；其缺点是难于实时。不过，由于运算方法和集成技术的进展，实时处理基本上已得到解决，目前已处于实用化的阶段。或者说，数字信号处理已成为一个桥梁，将各种传统的信号处理设备更新换代成超大规模集成设备的桥梁。以语言通信终端这样一种信号处理设备来说，利用语音学和数字信号处理的目前研究成果，已经可以将信道设备（包括调制解调器和均衡器）与声码器做在一个 $100 \times 120 \times 180\text{mm}$ 的体积之内了。

数字信号处理已经渗透到许多学科领域中，形成了许多边缘学科，如：雷达信号处理，声纳信号处理，地震信号处理，图象信号处理，语言信号处理…等新的边缘学科。种种新的算法软件和种种超大规模集成的专用硬件和通用硬件，层出不穷，日新月异，发展迅速。

第 1.6 节 语言信号处理及其应用

数字信号处理与语言学（包括语音学和语义学）相结合，产生了语言信号处理这个新的边缘学科。如果这个结合的目的是为了进一步研究语言学中的问题，那就是“实验语言学”；如果这个结合的是为了解决人与人之间，人与机器之间的信息交换的问题，那就可称为“语言信号处理”^①。

我们认为，语言信息交换大致上可以分为三大类：1) 人与人之间的语音通信（包括语音压缩与编码、语音增强等）。2) 第一类人机语言通信问题，指的是机器讲话、人听话

^①在语言信号处理中，有时只涉及话音信号，或者将语言信号当作纯粹的物理声学信号来处理时，也可称之为语音信号处理。例如语音编码，语音合成等。本书中这二个名词混用处不再另注。