# “轻声耳语”和“悠悠鹿鸣”

你喜欢音乐吗？你拥有“高保真”音乐重放设备吗？

如果你有机会到世界上最著名的音乐厅——奥地利维也纳音乐厅、美国波士顿音乐厅或荷兰阿姆斯特丹音乐厅欣赏到著名乐队的演奏，这种实地现场的感受如果与同样曲目的 CD 唱片在环绕立体声音乐重放系统播放的音质相比较，即使是当今最高档的还音设备，也远不能达到音乐现场演奏的优美效果。

人们在音乐厅中欣赏乐队演奏时，人的“双耳聆听”具有立体感，即一方面感受到乐队中各种乐器在不同方位发出既独特又和谐的音乐，另一方面乐队发出的声音还通过音乐厅各个界面的来回反射，在座位上形成音乐的空间感、环绕感，才有“余音绕梁，三日不绝”的美妙境界。

自从 1877 年爱迪生发明留声机以来，录音和放音的技术设备不断改进。20 世纪中期开始出现多声道录音和立体声放音的设备。随着科技进步，特别是数字音频技术的发展，音乐重放系统的指标完全达到了国际电工委员会提出的关于“高保真（Hi－Fi）放声系统的最低要求”的技术指标。到 90 年代出现“5.1 环绕立体声系统”，声音的质量和立体声的感受达到了前所未有的高度。但为什么达不到音乐厅中的效果呢？究其原因，一是扬声器，二是听音的环境效果。

现在的环绕立体声系统，具有前后左右不同方位来的声音

需要全本请访问 www.ertongbook.com

的空间感觉，使你处于音乐包围之中，以期收到在音乐厅的效果，或者根据音乐场景所需的声音方位感的效果。

立体声效果的技术原理如图 1 所示的听觉实验（在江苏省科学宫可以做此实验，有条件的家用还音设备也可做这样的实验）在听音人 M 的左前方和右前方有两只音箱 A 和 B，当同时给 A、B 以相同强度的信号时，所听到的‘声象’位置既不在 A，也不在 B，而是在 A 与 B 的中间 O 的位置。

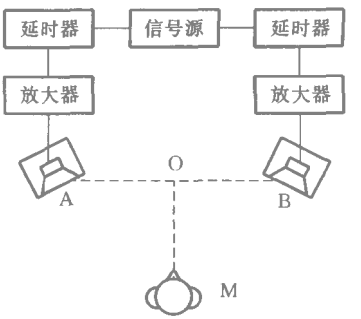



图 1

如果 A 的声音比 B 的声音强，则声象位置从 O 处偏向 A 处 如果让给 B 的信号通过延时器使它比给 A 的信号延迟几个毫秒，则声象位置也感觉偏向 A 处。如果给 A 与 B 的信号分别有不同的时间差和强度(音量)差，就可使声象处于前方从左到右的不同方位，给人以立体声的感觉。这便是双声道立体声的技术原理。

双声道立体声放声系统对听音人来说在他前方展示出一个面的方位感，人处于与两音箱组成正三角形的顶点位置(或稍后一些)有最佳的听音效果。随后发展了多声道立体声系统，近年来最新发展的是称之为‘5.1 环绕立体声系统’：一种是 20 世纪 90 年代初被美国定为高清晰度电视的音频标准（以及数字视盘机及视盘 DVD 的音频标准）的‘AC－3 杜比数字环绕声系统’（Dolby digital surround sound system，AC 是音频编码 audio coding 的缩写）另一种称为 DTS(digital theater system，即数字化影院系统)的多声道处理系统。所谓‘5.1 系统’是指左(L)、中(C)、右

(R)、左环绕 (LS)、右环绕 (RS) 五个声道，再加上前面一个超低音 (SUB) 声道 (频率低于 100 赫的专用音箱),也称‘ 5 + 1 系统 ”,参看图 2。在虚线内为听音良好位置，可以获得身临音乐厅空间之中的美好感觉。

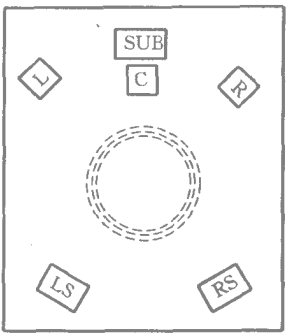



图 2

有了好的音乐重放设备,还需在听音的房间作适当的声学处理。你可能有这样的经验 在浴室里放声歌唱,似乎感觉声音变得‘ 浑厚 ”了。这是因为浴室中吸声量极少,声音就发‘ 混 ”(用声学术语描述是‘ 混响时间 ”太长)。在空的房间内拍一下手,可能会听到在两平行墙面之间有‘ 啪、啪…… ”好多次来回反射的声音,声学术语称之为‘ 颤动回声 ”。所以在一个未经装饰、家具很少的房间里听音乐,可能会产生这些不良的声学现象,虽然空间内充满了声音,但却不是应有的环绕立体声,而是混杂一片。

专业听音室的要求高，要经过建筑声学方面的设计。既要按计算布置各种吸声材料和结构，使从低频到高频的混响时间都控制在合适的范围,还要对吸声面、反射面、散射面作合理安排,使中心区座位上获得最好的听音效果。此外,还需噪声控制设计,使外界的噪声和振动,包括通风系统的噪声,不干扰室内的听音效果。因为吸声材料的物理性能和几何尺寸对吸声的频率特性有关,材料内部有微孔并且相连通,才能使声波的振动能量转化为摩擦热量被损耗。材料越厚，吸声的频率越低。但不能任意增大厚度,所以低频吸声技术目前主要依靠各种‘ 共振吸声 ’的方法,如薄板吸声、薄膜吸声、帘幕与墙面(窗户)一定距离

时的共振吸声以及用得较多的穿孔共振吸声结构。突破已有吸声理论,研究宽频带高效吸声结构,这是声学装修工程和噪声控制工程非常需要的研究课题。

为什么很好的高保真重放设备,你仍然一听就知道‘这是喇叭里出来的”?这主要是将电信号转变为声信号的电声换能器——扬声器,还不能完全‘保真’地还原音乐信号。单只扬声器(箱),它的频率范围不够宽,不同频率的辐射方向特性也有不同,音量大了也还可听到失真。用高、低频甚至高、中、低频组合扬声器,可以使频率范围展宽,但不同频率的辐射特点不同,频率分割点附近的功率匹配、位相匹配也不易做好。总之,大半个世纪以来,扬声器的材料、工艺、设计做了很多改进,但根本原理没有变化。要使重放音质有进一步的飞跃,需要在电声换能器的机理上有大的改革,这是新世纪的一个迫切需要解决的研究课题。

目前立体声技术都基于双耳聆听的方位感的基理。不同方向来的声音,两耳接收时在强度上或位相上有差别,所以能判别方位。因为两耳在人头的左右两侧,所以对水平方位声音的分辨率高,大约在 3°左右;对垂直方位声音的分辨率低,大约在 10° ~ 15° 以上才能辨别。通过图 1 所示的听觉实验,从两扬声器到达某个听音点的强度差和时间差(即位相差)产生声象位置变化的原理,获得了制作立体声重放的技术。但人的听觉还有辨别声音远近的功能。声音的响和轻并不代表声音的远和近。同样是很小的声音,轻声耳语,知道声源就在耳边;悠悠鹿鸣,大概在远处的树林中。但人耳能辨别声音远近的机理尚不清楚,也即还没有找到物理参量来解释人耳远近感的判别依据。所以现在的立体声,即使是环绕立体声,也只是左、右、前、后方位的立体声,还没有声音的远近深度变化的感觉,这是

**要从人的听觉机理的基本研究开始的使立体声技术更上一层楼的研究内容。**

**（孙广荣）**

需要全本请访问 www.ertongbook.com

# 语音处理技术

## 什么是语音处理

语言是人类最重要的交际工具。语言有两种基本形式，即口头语言和书面语言。人们用口头语言进行交流的过程如下图所示。设有甲、乙两人，甲说乙听，首先是甲在大脑中形成语言，并由大脑控制发音器官产生语音。语音离开甲的口以后，作为一种声波在空气中向听话人乙传递过去。语音传到乙的听觉器官耳之后，为乙所接受。听觉器官分解并传递语音信息至大脑，由大脑正确地感知语言的内涵。




人类的发音器官由声带、咽腔、鼻腔和口腔等组成。模仿人类发音器官的功能，研究人工产生语音的方法，是语音合成技术。研究在空气中传播语音的物理性质，如语音的频谱分布等，是属于声学语音学范围。模仿人类听觉系统的功能，从接收到的语音信号，辨认出语音所承载的语言信息，即口头语言的内容和意义，是语音识别技术。在计算机技术飞速发展的情况下，人

们加紧研究语音合成技术和语音识别技术。希望能为计算机配备合成语音输出终端和具有语音识别功能的输入终端。这样人们便可以应用自然的口头语言而不是现在使用的键盘，来控制计算机并与计算机进行对话。为此语音合成技术和语音识别技术的研究,从 20 世纪 60 年代以来一直是一个热门的研究课题。

通常在语音从发音端人的口到接受端人的耳之间的传递空间是开放的,在这空间除所传递的语音之外,还会存在着其他的声音同时在传播,因此接受端的耳,既接收到语音信号,同时也接受到其他声音。这些声音在听觉系统中对语音的正确感知造成干扰。尤其在干扰声很强的情况下，对语言的感知会引起困难。这些干扰声音被称为噪声。在某些工厂如纺织厂的车间内,在战斗机的座舱内,在舰艇的机舱内等等,都存在着强的干扰噪声。有没有什么处理方法可以在耳端接受到的混有强噪声的语音信号中,把噪声抑制下去,或是把语音提取出来呢？研究在含噪语音中抑制噪声的方法称为语音增强技术。语音增强技术在国防上和工业上都是极被重视的研究课题。 20 世纪 70 年代以来,由于数字技术的发展,也研究出一些语音增强的具体方法,在一定的条件下,得到一定的效果。然而语音是一种随机信号,而噪声往往也是随机的,要把两个混在一起的随机信号分离开，在理论上至今还没有很好地解决，在技术上当然更难实现。因此语音增强技术是当前语音处理技术中一个〃待研究和攻克的最难的课题。

人的语音具有很强的个人特征，因此可以通过语音辨别出说话人是谁。这种通过语音辨认说话人的识别被称为话者识别，它在公安和保卫部门有着重要的应用。由于语音的个人特征随时间、情绪、健康状况等变化而变化,这给话者识别技术带来很大的困难。话者识别也是语音处理技术中一个至今尚未很

好解决的难题。

## 语音合成技术

早在 19 世纪，人们就开始研究语音合成方法。当时应用管子长度的变化模仿人的发音过程中口腔、鼻腔的共鸣作用，研制出能发几个元音的机械语音合成器。在 20 世纪 30 年代初，采用模拟电子技术研制出琴键式电子语音合成器。这些合成器都只能供表演用而不具有实用价值。直到 60 年代，应用数字技术的语音处理方法得到发展以后，合成语音才成为可能。最早成功地应用语音合成技术的是保密语音通讯中的声码器。 80 年代以来，语音合成技术走向实际应用，大致有以下几类 第一类是词汇量不多而且是固定词的语音合成器，要求合成语音的音质清晰可懂，如公交汽车上的语音报站，117 语音报时等。第二类是在保密通信中应用的声码器，它要求高的保密通信和低的传输码率和一定要求的语音质量。第三类是在文语转换系统中的应用，它是将文字材料直接转换成清晰自然的口头语言。用在计算机的信息查询系统，辅助教学系统中，可以用自然语音输出说明文本，可以用口语方式回答问题等。还可以通过它让盲人阅读普通人阅读的书刊和报纸。这种文语转换的语音合成技术是语音合成所追求的目标。

从语音合成的技术上讲，上述第一类应用中的语音合成技术较为简单，属于波形存储方法，已经成熟，只要将需要合成的词组或句子应用已有生产的语音合成芯片（在一般电子技术杂志上可以查到）处理后，转换成数字量并存储起来，在存储器中建立一个所需词汇的语音库。在发音时，从这语音库中取出对应词句的数据，再通过语音合成芯片恢复成语音输出。第二类

保密通信用的声码器，它采用参数合成的方法。这方法是模拟人类发音器官产生语音的过程来合成语音。它采用噪声和周期脉冲代替肺和声带的激励作用，用数字滤波器替代声道的共鸣作用。参数合成方法，在发话端，先将人的自然语音切成小段(约 20 毫秒一段),对每一小段语音提取代表激励特征的音调参数及代表声道共鸣特征的滤波器参数。把代表该语音段的参数进行编码加密，然后发送出去。在接收端，将这些参数恢复后，通过语音产生的模型合成出语音段，然后拼接成合成语音。现在这种参数合成语音的方法，已经可以用超大规模的集成芯片来实现。这种合成方法可以极大地压缩数据量，减少通信码率，但语音质量受到一定的影响。进一步提高语音质量和降低码率一直是这种合成方法需要继续研究的问题。第三类文语转换系统则采用规则合成的方法。因为它所需合成的词汇量和句子不受限制。这方法将代表语音的每个音素存储起来，合成时从音素库中取出音素进行拼接。在拼接过程中需根据语句的内容，应用音调规则、韵律规则、音素的变化规则等等,对原有音素进行调整后再连接。因此这种合成方法称为规则合成。虽然音素的数量不是非常多，但是要拼接出高度自然的合成语音却是非常困难的。这种规则合成现在还处在研究阶段。研究的重点是和语言学家配合，建立音素在自然语言中的各种变化规则。要有完善的规则才能合成高度自然的语音。当然也不排除将来会提出更新的合成方法来实现词汇不限、音质高度自然的文语转换系统。

## 语音识别技术

从语音识别系统研究的发展过程看，它沿着以下三个方面

需要全本请访问 www.ertongbook.com

不断地前进：

(1) 在发音的方式上，由孤立语音向连续语音方向发展。孤立语音是指待识别语音发音时，以每一个词一停顿的间断方式发音。早期的识别系统都是孤立语音方式的词识别系统。它可以应用在如计算机控制命令的识别。连续语音识别是指待识别语音的发音方式不作一词一间断的限制，可以用较自然的方式发音。它应是语音识别中发音方式的主流。然而在连续语音中要正确地切分出字或词来是非常困难的。

(2) 在识别系统能识别的词汇量上，由小词汇量向中词汇量、大词汇量发展。最初的识别系统都是小词汇量的，只能识别几十个词，如计算机的控制命令、地名和数字等。词汇量的增大会引起词和词之间的混淆，因而增加识别的困难。如果要求系统能识别口述的报告或文章，相应的词汇量必须很大，当然这也是识别系统最终要达到的目的。

(3) 在语音识别系统的使用对象上，由特定人的语音识别向非特定人的语音识别发展。早期建立的语音识别系统是以某特定人的语音进行训练的，它对该特定的人使用时，具有最高的正确识别率，当其他人使用时识别率就会明显地下降。然而应该要求识别系统同时能为众多的人服务，所以语音识别系统的使用对象必需是非特定人的。

事实上，连续语音识别、大词汇量的语音识别和非特定人语音识别是发展语音识别的研究中需要攻克的三大难题，直到现在还在继续努力之中。

语音识别研究的历史还不长，最早的是 20 世纪 50 年代初在贝尔实验室建立的特定人、孤立发音的英语数字识别系统。系统采用模式匹配的方法。训练时先由特定的使用人，对要求识别的词逐个发音。对每个词的语音提取特征参数，如语音的

共振峰频率等。在存储器中建立每个词的语音特征参数样本库。识别时,使用人再发待识别的词汇,输入到识别系统中。系统用相同的方法提取语音特征参数，将它与库中的样本逐个进行比较，以最接近的样本作为识别输出。

在 20 世纪 60 和 70 年代语音识别研究的重点是在识别方法上。对语音特征参数选取上，相继有频谱参数、线性预测参数、倒谱参数等。现在用得最多的是倒谱参数。选择的特征参数应最有效地显示出各个语音词汇的不同，以利于模式匹配中进行划分。在待识别语音和样本的匹配过程中，一个很大的困难是每次发音中,发音的快慢都会有变化,造成语音长短的不一致，使得待识别语音和样本的比较在时间的对准上发生困难。动态时间规整( DTW)方法可以较好地克服这一困难,但运算量比较大。

20 世纪 80 年代是语音识别技术发展的关键时代，在识别方法上引入了隐马尔柯夫( HMM )模型为框架的概率统计识别方法。它是先对词汇表的语音特征进行隐马尔柯夫模型训练并建立模型库。识别时,用 HMM 模型计算待识别语音的概率，以概率最大者作为识别输出。隐马尔柯夫模型识别方法可以较好地解决语音发音快慢的变化和语音声学特性的变化而引起的困难。至此语音识别技术取得了长足的进步，初步攻克了非特定人、大词汇量和连续语音三大难点。然而基本的 HMM 模型在训练算法及模型本身都存在一些不完全合理的假设，制约了识别性能。因此还在继续研究以改善 HMM 模型的识别方法。

20 世纪 90 年代出现了不少语音识别系统，然而诸如环境噪声的干扰,说话者的口音,识别率的提高等等一系列的问题尚待解决。因此语音识别系统的实际应用还需有一个过程。

近年来人工神经网络的研究蓬勃发展。根据神经网络具有

强的学习能力和分类能力，如何将它应用到语音识别系统中以提高系统的识别能力,也是一个重要的研究课题。还有,人的听觉系统对环境噪声的干扰和不同人的口音都具有很强的适应性，所以很多人对研究听觉模型在语音识别中的应用感到兴趣。另外,以汉语来说,从音节构成词,从词构成句子都不是随意的，有语言学上的约束,必须进行语言模型的研究,把语言学中约束规则结合到语音识别系统中，对提高系统的性能有非常重要的作用。

（余崇智）

# 大海中声传播的奥秘

蓝天、白云、一望无际的大海,是各式各样的声音打破这一片沉寂,将我们带进海中声传播的世界里,去遨游、去探索……

众所周知,在海水中,以声波传播性能为最佳,在混浊和含盐的海水中,无论光波或电磁波都远较声波的衰减为大,不能有效地传递信息。因此在这海洋世界里,作为声信息传输的通道(称之为信道),其海面、海底及海水介质都具有十分复杂和多变的特性:风和日丽,海面宛如一面镜子;狂风巨浪,海面犹如一头咆哮的雄狮;一年四季不一样,一天早晚都不一样,海中声波传播速度因温度、深度、盐度而异;海底世界更是奇妙无比,致使声波在海水中传播的规律相应也是十分复杂和多变。了解和掌握水下声场的这种复杂多变的规律并作出合理的预报,对水下探测和通讯系统的设计、制造、操纵和使用,以及增进人类对海洋环境的认识,海洋资源的开发是必不可少的。那么声音在海中究竟是怎样传播的呢?它能传多远呢?如何通过声音去探测海中的目标——舰艇、潜艇、鱼雷、水雷、水下暗礁、鱼群及其他海中之物呢?让我们去揭示海空的声传播的奥秘吧!

## 舰船的水下耳目——声纳

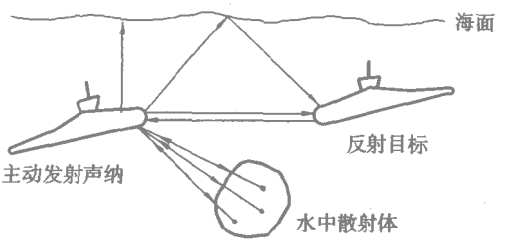
声纳是英文“ sonar ”一词的译音,由声音( sound )、导 航

(navigation)和测距(ranging)三个英文字母的字头构成,意即用发射声波的方法来发现目标和测定它在水下的位置，是第二次世界大战后仿照当时颇具魅力的‘雷达’一词而来的。虽然声纳的“新时代”可以追溯到第二次世界大战初期,然而声纳的起源却是很早的,在 1490 年即哥伦布发现美洲大陆的前两年，列昂纳多·芬奇是一个多才多艺的造型工程师，在他的一本摘记中曾写到‘如果使船停航,将长管的一端插入水中,而将管的另一开口放在耳朵旁,则能听到远处的航船”,这种被动声纳的雏形具有极其简单的优点，这项引人注目的发现包含了现代被动声纳系统的全部主要器件。它指出了航行的船，无论是用帆还是用桨，所产生的声音在水中能传播到相当远的距离。它描述了一种接收装置，这里是充满空气的长管，将水中的声音转变为空气中的声音,耦合到耳-脑组合，结果是收听到并识别出远处的船只。它还指出这样的事实：若减少混在远处航船声音中的自噪声(如使探测船本身停航),则将获得较好的效果。然而就是这种在海中和听测者之间用了一根长管的听音方法一直到第一次世界大战还广为采用，那时，在另一只耳朵和海水中的另一点之间又加了一根长管，这样就能测得方向，而目标的方位也就能确定了。在 1912 年,由于受‘泰坦尼克’号沉没的驱使，费森登(Fessenden)将其设备用于对两英里处的冰山进行回声测距……

随着时代的变化,换能器件的更换,电子新技术的出现,计算机的应用等促进了声纳设备的不断更新。时至今日，声纳的定义已远远超出原来的定义,广义地说,凡属对水中目标进行探测、定位、跟踪、识别、通讯、导航制导、武器的射击指挥和探测对抗方面的水声设备皆属声纳设备的范畴。

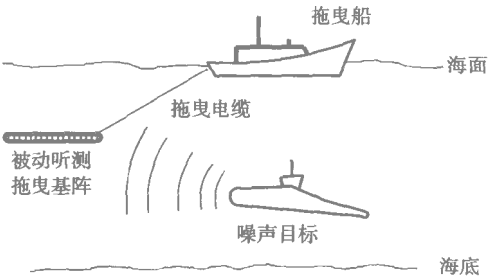
声纳在海中是怎样工作呢？声纳具有两种工作方式，一类

称之为主动声纳,由声纳发射某种探测信号(声音),在水中传播的路上遇到障碍物或目标,反射回来到达发射点被接收,由于目标的信息保存在被目标反射回来的回波之中，可以根据回波信号获得目标距离、方位、速度轨迹、目标性质等参数。它能精确测定目标距离，可探测固定目标（海底的潜艇沉船、飞机残骸等)。但隐蔽性差,因工作时发射信号,易被发现。作用距离较近,工作是双程的,传播损失大,故回波信号弱,当有其他噪声干扰时就会掩盖有用的回波信号。另一类称之为被动声纳，是用接收基阵搜索海区内目标发出的噪声，并以此来测定目标距离、方位和性质的设备，用作警戒或测距。它能较准确地判断目标性质,除能测向外,还能测距。其隐蔽性好,当潜艇处于潜航状态时可一直开机工作，不断地监视水下周围情况，以防遭受袭击。惟一的不足是无法测定静止目标。




主动声纳工作示意图

声纳在军事和民用事业中得到广泛的应用。当代声纳的发展正处于空前高涨时期，由于潜艇的迅速发展，出现了下潜深、长期潜航并载有各种导弹武器的高速核动力潜艇，引起对反潜问题的重视，再加上电子技术飞速发展使声纳发展在技术上的实现成为可能。国外军用声纳的发展已经历了三代，正向第四代发展,国内正研制的军用声纳基本上属于第二代,某些声纳已




被动声纳工作示意图

具备第三代声纳的部分特征。当代先进的声纳具有更高的综合性能：大功率发射(200～1 000 千瓦)、工作频率低频化(0.1～3.5 千赫)、探测距离远(在深海声道传播 主动式声纳达 30～35 海里，被动式声纳达 120～125 海里）、定位精度高（方位精度 ±0.2～±0.1 度，距离精度 ±1%～±10%，在 3～5 海里距离上，深度为倾角 1 度)、扫描范围大(水平 0°～360°，垂直 +35°～−35°)、多目标探测、多路信号处理、多路传播途径、多种信号显示、多种换能器材料和多种安装法、广泛应用电子计算机。

尽快地研制出跨越某些发展阶段的声纳换代产品，赶上世界先进水平，是当代青年的光荣使命。历史告诉我们：落后就要挨打，就要受人欺负。为了建设一支强大的海军，保卫祖国的海疆，青年朋友们要刻苦钻研技术、敢于创新，造出最先进的声纳设备，去迎接新世纪的新挑战。

（龚淑仪）

# 水下目标的识别

一声巨响,豪华客轮‘泰坦尼克’号撞到冰山上,轮船沉没,致使成千旅客葬身大洋,这是由于没有及早发现冰山这个目标。

在大海中有水面舰艇、水下潜艇、鱼雷、水雷、暗礁、鱼群及生物群等各式各样的目标,如何通过大海中传播的声音去识别目标,将目标分类,这在军用或民用上都具有很重要的意义。尤其在海战中,对声纳目标能否进行检测和分类,对于海军所有的水面舰艇和潜艇的性能和生命力都是至关重要的,是海军作战能力的重要标志。如果在识别敌对目标中有任何失误和延误,本舰就可能遭到突然袭击,或失去攻击的机会。当今的时代,声纳作用距离愈来愈远,在作用距离范围内目标很多,确定哪个是感兴趣的目标或要攻击的目标,就非常重要。在民用方面,随着海洋开发活动的日益增多,如海底资源开发、石油勘探、海底地貌的绘制、海底铺设电缆、水下机器人和探鱼等等。为了进行水下作业,水下目标识别是必不可少的,也受到水声工作者对此的关注和青睐。

水声目标分类识别是国际公认的最具有难度的学科之一。由于目标的多样性和复杂性,使目标特性很难描述和分类识别。如对舰艇的分类识别,目标分水面舰船、水下潜艇和其他水下运动物体,如鱼雷;水面舰艇又分为战斗舰艇和民船;战斗舰艇又可分为大舰和小舰艇;大舰又分为驱逐舰、护卫舰、巡洋舰和航空

需要全本请访问 www.ertongbook.com