

水下噪声音色属性回归模型及其 在目标识别中的应用*

王 娜[†] 陈克安

(西北工业大学环境工程系, 西安 710072)

(2009 年 7 月 30 日收到; 2009 年 8 月 18 日收到修改稿)

通过对声音的主观评价与客观分析而建立的主观感受数学模型,在许多领域都有重要的应用. 本文采用多元线性回归分析手段对水下噪声音色属性建立回归模型,提取音色特征并改善水下目标的识别效果. 首先,在前期水下噪声音色属性主观评价实验的基础上,将构成音色属性空间的 5 个成分的评价分值作为回归分析中的因变量,提取大量与听觉感知相关的听觉特征作为自变量;然后,通过相关分析和改进的逐步筛选法,挑选出反映音色属性的“最优”自变量子集;最后,利用向后剔除回归分析和水下目标识别实验,确定适当的音色模型,并通过假设检验证明该线性模型不仅正确有效,而且能改善水下目标识别效果.

关键词: 音色,多元线性回归,主观评价

PACC: 8734, 4330

1. 引 言

水下目标自动识别在海战中起到十分重要的作用,一直受到极为关注. 其关键在于特征的提取. 尽管常用的时域波形结构、频谱分析等^[1-3]特征提取方法已取得良好识别效果,但实践中仍然依赖声纳员耳听识别对水下目标类型进行最终判断. 虽然人的听觉系统容易受到生理和心理因素的影响,但在多数情况下其识别性能仍然优于现有的机器自动识别技术^[4]. 因此,模仿人耳听觉原理提取目标特征的方法,为改善目标识别性能提供了一条新的思路.

在人类听音辨物的主观感知过程中,声音的听觉属性起到决定性的作用. 声音的听觉属性可用响度、音调、音色和持续时间来表述,其中响度、音调和持续时间是一维感知属性,基于这些属性已有较为成熟的计算模型^[5]. 而音色是一个更复杂的多维感知属性,受声音时域、频域结构共同影响^[6]. 在我们的前期研究中发现^[7],除了音调可作为有效特征外,反映音色属性的特性响度和特性尖锐度也是有效特征. 也就是说,在水下噪声信号中,由于响度的

大小受目标远近距离和航行工况等条件影响,信号的音调和音色共同反映目标的本质差别,是听觉判断目标差异的主要依据^[8,9]. 所以,模仿听觉提取有效识别特征的主要工作集中在如何提取准确的音色特征方面.

听觉感知研究,早已意识到音色属性的重要性,并针对乐器声开展了大量的研究工作. 常用的研究手段是对声音进行主观评价,利用多维尺度分析(multidimensional scaling,简称 MDS)和主成分分析(principal component analysis,简称 PCA)音色空间的维度. Grey^[10]提出以谱能量分布为主的三维音色空间可以解释一系列乐器的音色组成. 随后,信号特征与音色感知之间的关系也得到了深入研究,并提出与音色维度紧密相关的参数. Daniel^[11]发现有 4 个特征与听觉感知相关,Caclin 等^[12]提出起奏时间、谱矩心、谱精细结构与合成音的音色空间显著相关,而谱通量则相对较差. 然后利用与音色相关的特征实现了乐器的自动识别,McAdams 等^[13]发现利用谱结构中的包络特征可以区分乐器声,Agostini 等^[14]利用谱质心、谱质心带宽、不和谐性和谐波能量的偏斜度实现了对乐器识别.

虽然至今没有公认的标准尺度去衡量音色,但

* 西北工业大学基础研究基金(批准号:W018104)资助的课题.

[†] E-mail: nawang@mail.nwpu.edu.cn

在合成声音的研究中,针对由谐波构成的乐器声也进行了音色建模的研究工作^[15-17].常用的方法是,对乐器声进行谐波分析,通过听觉感知的研究,以正弦相加模型为基础,利用已提出的与音色紧密相关的参数对每个分音建模,改变参数将合成声与录制声比较进行主观评价,进而分析出音色模型的参数,对声音建立物理结构的声学模型,实现乐器声的分析、合成以及自动识别.

由于大多数噪声具有嘈杂性和非谐音的特点,所以相对于乐音而言,噪声音色属性的研究工作相当困难.虽然利用统计分析手段研究发现辨别交通噪声的音色空间是三维空间^[18],水下噪声的音色空间是5维空间^[19],而且Victor等^[20]提取与听觉相关的特征用于海域报警,实现了自动识别,但很少有针对噪声进行音色建模、提取音色特征的研究.其主要原因在于,现有的乐音音色建模方法,不仅不适用于复杂的噪声研究,而且仍然无法建立通用的音色模型来反映主观判断目标的差异.

为提取音色特征应用于水下目标识别领域,本文提出利用回归分析手段,对音色属性听觉感知的评价分值和信号的听觉特征进行数量变化关系的研究,希望找到反映它们之间统计规律的数学模型,在提取音色特征改善目标识别性能的同时,为有针对性地选择声纳员的训练声音提供理论指导.

2. 音色属性建模及其主观评价

现有的音色建模方法多是利用信号的听觉感知特征针对有规律性的声音进行建模,根据相加模

型或物理模型估计出音色参数,来实现对音色属性的建模.但由于水下噪声的声学特性相当复杂,无法利用现有的建模方法.考虑到要真实反映听觉感知的结果,我们提出对水下噪声在实验室内进行主观评价实验,统计分析音色属性的构成及样本在各属性下的数值表示,利用主观感受的分值与信号的听觉感知特征建立数量关系,实现音色属性的建模.

在文献[18,19]中,作者对等响度、等时长的水下噪声进行了不相似性评价和语义细分实验,并对水下噪声音色属性的空间维度数及其构成进行了详细分析.在此基础上,首先利用主成分分析手段,根据在各描述词属性下评价主体对实验声样本的评价分值,计算获得音色属性的各成分的相应得分;同时计算大量与听觉感知相关的听觉特征(时域、频域特征),并剔除反映声音特性相似的特征;然后改进了回归分析中筛选自变量的方法,获得各成分的“最优”自变量子集,并基于回归分析原理将音色空间的主观感受与听觉特征之间建立数学联系;最后提取音色特征,通过水下目标识别实验研究,确定合适的音色属性回归模型.这有助于了解人耳听觉系统辨别水下目标的信号特性的依据,为进一步研究模仿听觉功能提取有效听觉特征奠定了基础.具体音色属性建模流程如图1所示.

为探求人耳感知水下目标类型的声学因素,有必要先对水下噪声音色属性的构成进行分析.文献[19]中的水下噪声的不相似性评价实验和语义细分评价实验,利用了统计分析手段,对水下噪声音色属性的构成进行了具体分析.

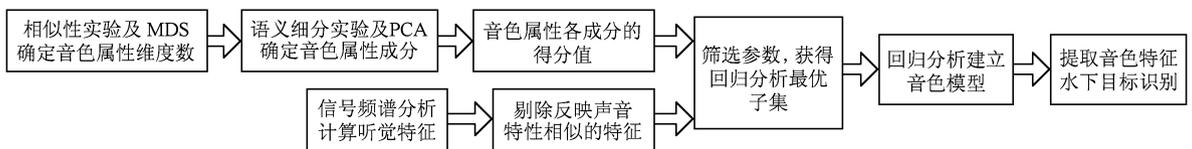


图1 音色属性建模流程

评价实验中样本是采样频率为 44.1 kHz、时长为 5 s、响度为 15 宋的 30 个水下噪声(包括不同工况的潜艇噪声、军舰噪声、民用船噪声以及水下环境声等),27 位无听力损伤的在校大学生和研究生作为评价主体,通过监听级耳机对声音进行主观评价实验.

相似性实验中采用 7 级相似评价尺度进行评价,然后利用 MDS 手段将实验获得的相似性矩阵转

换成目标距离,通过降维处理,用低维空间中点与点的距离来表示声音的差异程度,最大程度地拟合目标距离,最终分析确定水下噪声的音色属性空间为 5 维空间.

语义细分实验仍采用 7 级评价尺度,获得由 30 个样本分别在 15 个描述词属性下的评价价值构成的矩阵 \mathbf{P} ,采用 PCA 手段将多个描述词属性化为少数几个综合指标(主成分 \mathbf{Y}),将相关性较高即联系紧

密的描述词分在同一个成分中,相关性差的分在不同成分,每个成分之间相互独立,最终获得与 5 维空间各维度对应的 5 个主成分,5 个主成分能够解释水下噪声音色属性空间的总 94.63%。第 1 个成分反映了声音的描述词为“沉闷的”、“低沉的”、“粗糙的”和“嘈杂的”的主观感受;第 2 个成分反映描述词为“变化的”、“不单调的”、“不平缓的”和“震颤的”的感受;第 3 个成分反映描述词为“急促的”、“紧凑的”和“短促的”的主观感受;第 4 个成分反映描述词为“刺耳的”和“不柔和的”的感受;第 5 个成分反映描述词为“重复的”和“有节奏的”的主观感受。5 个主成分的分值可以通过各描述词的线性组合来计算,因此 PCA 分析中的单位正交向量确定了 5 个主成分的系数矩阵 C ,如表 1 所示。 C 反映了各成分与 15 个描述词的评价分值之间的定量关系。依据公式 $Y = CP$ 计算获得各样本在水下噪声音色属性的 5 个成分上的评价分值。

5 个成分的分值从数值上定量体现了决定音色属性差异的 5 种相互独立的主观感受,是在音色属性建模回归分析中的因变量。为实现音色属性的定量计算和便于后期提取音色特征进行目标识别研究,在对 5 个成分的分值进行回归分析建模前,必须计算大量与音色感知密切相关的听觉特征(频域、时域特征等)作为自变量,利用回归分析的数学手段分别建立构成音色空间的 5 个成分的数学模型。

表 1 语义细分实验的 5 个主成分的系数矩阵 C

描述词	成分				
	Y_1	Y_2	Y_3	Y_4	Y_5
震颤的	0.24	0.25	-0.02	0.07	0.17
不平缓的	0.04	0.26	-0.10	0.11	0.19
刺耳的	-0.14	0.11	-0.14	0.53	-0.05
嘈杂的	0.13	-0.12	0.20	0.09	-0.17
沉闷的	0.31	0.07	-0.11	-0.04	0.15
急促的	-0.00	-0.07	0.39	-0.05	-0.14
有节奏的	0.01	0.02	-0.05	-0.02	0.41
紧凑的	0.01	-0.10	0.42	-0.15	-0.14
变化的	-0.03	0.35	0.02	0.09	-0.24
粗糙的	0.21	-0.10	0.18	-0.03	-0.09
不柔和的	0.05	0.07	-0.04	0.42	-0.01
不单调的	-0.02	0.25	0.02	-0.12	-0.18
低沉的	0.32	0.09	-0.11	-0.09	0.15
短促的	-0.03	0.05	0.19	0.02	0.05
重复的	0.05	-0.17	-0.01	-0.09	0.51

3. 音色属性的多元线性回归分析

多元线性回归分析用于考察一个因变量与多个自变量的相互依赖关系。其基本思想是:通过考察变量之间的数量变化规律,以方程形式描述和反映这种变化关系,以通过控制自变量来达到对因变量进行预测和控制的目的。因此,在获得水下噪声音色属性的评价分值的基础上,利用多元线性回归分析的数学手段,分析声音样本的听觉特征与音色属性的评价分值之间的数量变化关系,获得音色属性的回归模型。

3.1. 听觉感知特征的计算

在听觉感知研究领域中,认为音色属性是人耳辨识目标的主要依据^[9]。美国国家标准协会定义:音色是听觉敏感的属性,是具有相同响度和音调的两种声音在听觉感知属性上显示出来的差异。通过大量实验研究发现,音色属性主要依赖于激励的频谱、波形、声压、频谱的频率位置、激励的时域特性等,即声音的时域结构和频域结构共同影响音色属性。

在乐音领域的研究中发现时域特征、频域特征和谐波特征与音色属性密切相关。例如,所有关于音色属性研究的文献中都提及的谱质心和起奏时间,文献[6],[13],[15]和[21—24]中提出的谱通量、谱下降值、谱平整度和零交点比率等;文献[5]和[25]提出的尖锐度和粗糙度等;此外文献[6],[11],[21]和[26]中还提出了分段特征及其统计量。这些特征多种多样,在不同程度上影响了音色属性,而且不同的特征组合能够实现目标的自动识别。由于水下噪声的谐波特性不明显,为了后期的回归分析,本文计算了听觉感知特征中常用的时域特征和频域特征。

通过对多种类型的声学目标进行自动目标识别研究发现,对于水下噪声、频域特征更为有效,而适用于乐音领域的起奏时间等特征并不合适,因此确定了适用于环境声识别的听觉感知特征及其分段特征均能有效识别水下目标^[27,28]。本文提取的特征、计算模型及物理含义见表 2 所示。

为保证回归分析中的听觉特征具有一定的完备性,文献[6],[11],[21]和[25]的分段特征的计算方法,本文以表 2 列举的特征为基础,计算了反映

信号时频包络结构变化的分段特征,并参见文献[29]的计算方法计算其统计值,其中最小值、最大值比 f_r 和方差 f_v 反映量值之间的差异程度,分段特征包络的斜率(slope of envelope, f_{se})反映量值的变

化快慢程度,偏度(skewness) f_{sk} 反映量值的分布形态的对称性,峰度(kurtosis) f_k 反映量值分布形态的陡缓程度,几何平均值(harmonic mean) f_h 反映量值的平均水平.

表2 听觉感知特征名称及简称

特征名称	英文名称	表示符号	计算模型	物理含义
谱质心	spectral centroid	β_{sc}	本文对模型进行改进,见文献[28]	描述声音的明亮度,一定频率范围内通过能量加权平均的频率值
谱质心带宽	spectral centroid bandwidth	β_{scbw}	本文对模型进行改进,见文献[28]	反映了声音能量集中的频带宽度
谱通量	spectral flux	β_{sf}	见文献[6]和[22]	声音信号频谱的包络面积,反映信号的全频带能量
谱熵	spectral entropy	β_{se}	见文献[24]	描述谱的平坦特性
谱下降值	spectral roll-off	β_{sro}	见文献[6]和[22]	描述频谱倾斜程度的量值
谱不规则性	spectral irregularity	β_{si}	见文献[6]和[22]	谱包络的形状,信号能量大的谐波次数越多其值越小,即越嘈杂的声音的谱不规则性值越大
谱平整度	spectral flatness measurement	β_{sfm}	见文献[6]和[22]	描述谱的平滑程度
sfm的扩展	extension of SFM	β_{α}	见文献[6]	描述声音有调子的程度, β_{α} 越接近1,信号越有调子, β_{α} 越接近0,信号越像噪声
mel倒谱系数	mel-frequency cepstrum coefficients	β_{mfcc}	见文献[21]和[27]	基于人耳听觉特性的信号倒谱中各频带的特征
振幅因子	crest factor	β_{cf}	见文献[6]	最大振幅在整个包络中所占的重要程度
零交点比率	zero-crossing rate	β_{zcr}	见文献[6]	反映声音的吵闹程度
尖锐度	sharpness	β_s	本文对模型进行改进,见文献[7]	描述高频成分在声音频谱中所占比例的参数,反映着声音信号的刺耳程度
粗糙度	roughness	β_r	本文对模型进行改进,见文献[7]	反映信号的调制幅度的大小、调制频率的分布情况等特征

本文提取分段特征的方式是对点数为 N 、采样频率为 f_s 的时域信号 x ,加长度为 L 的矩形窗,将声信号分为 N/L 段后计算每段时域信号的时频特征,再计算统计量;或者对在 $[0, f_s/2]$ 频域内的频域信号 E ,采用依据听觉频率辨别特性提出的临界频带划分方式(见(1)式)将频域分段,计算每个临界频段的频域特征后计算统计量.

$$z = 13 \arctan(0.00076f) + 3.5 \arctan(f/7500)^2, \quad (1)$$

客观参量的具体计算公式参见文献[6]和[27],这里以频域特征谱质心为例,计算时域分段特征和频域分段特征.其公式分别为

$$\beta_{sc}^i = \frac{\sum_{n=C_{i-1}L}^{iL} f(n)E(n)}{\sum_{n=C_{i-1}L}^{iL} E(n)}, \quad (2)$$

$$\beta_{sc}^j = \frac{\sum_{z \in C_{j-1,j}^1} f(z)E(z)}{\sum_{z \in C_{j-1,j}^1} E(z)}, \quad (3)$$

其中 i 是以 L 长度对时域信号分段的分段数,其中 $i = 1, 2, \dots, N/L, L = 1024; j$ 是以临界频带 z 分析方式对频域信号分段的分段数,其中 $j = 1, 2, \dots, 24$.

在回归分析前,利用(4)式对听觉特征进行数

值标准化处理,保证客观参量 X 的数据在0—1之间.

$$X_i^n = \frac{X_i - \min\{X_i\}}{\max\{X_i\} - \min\{X_i\}}, \quad (4)$$

其中 i 是指30个样本计算的某一类客观参量.

在回归模型中如果漏掉对因变量影响显著的自变量,那么回归模型用于预测时会产生较大偏差,但模型中若包含太多对因变量影响不大的自变量,回归模型不仅使用不方便,而且会影响预测的精度,因而优化自变量子集,建立一个最优的回归方程是十分重要的问题.

3.2. 回归分析

回归分析的因变量是经过主成分分析获得的主观分值,不具有强相关性,因此对音色属性的5个主成分分别进行多元线性回归分析,即可获得5个音色的回归模型.

目前,广泛使用的筛选自变量方法是逐步筛选法^[29].该方法每次引入对因变量 Y 影响最显著的自变量 X ,并对方程中原来的变量逐个进行检验,把变

为不显著的变量逐个剔除掉,最终得到“最优”集合. 但逐步法的局限是筛选的优劣受到第一个引入方程的自变量的影响,会忽略某些“次优”自变量,而且该方法忽略了可能存在的效果相同(甚至更优越)的其他自变量子集. 因此,如图 2 所示,我们提出一种改进的逐步筛选法,循环从 X 中逐步筛选出显著变量,获得数量最大的“最优”自变量子集. 其具体步骤如下:

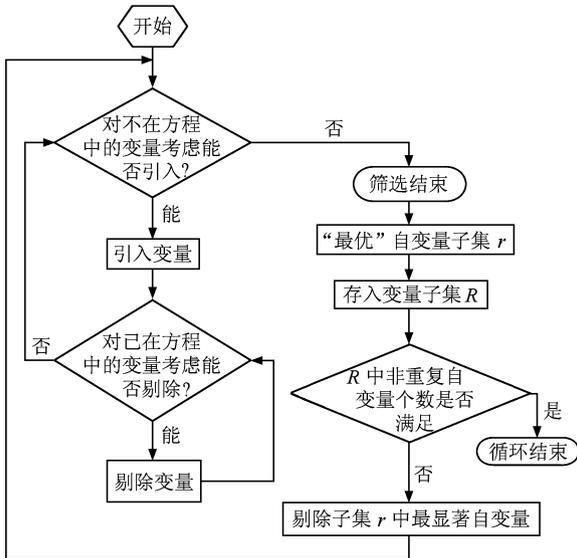


图 2 循环逐步筛选的步骤框图

1) 为保证回归分析的准确性,有必要对强相关的自变量进行剔除. 对 14 个时域特征和频域特征进行相关分析,相关系数 $r > 0.9$ 时,认为存在强相关性,任意剔除其中的一个,然后剔除与剩余 12 个特征强相关的分段特征的统计量,共获得 97 个自变量集合 X ;

2) 计算 Y 和 X 的相关系数,选择具有最大相关系数的自变量进入回归模型,对回归系数进行显著性检验,若 $\alpha \geq 0.05$ 则终止筛选,获得子集 r ,将 r 存入集合 R 中,进入第 5 步,否则进入下一步;

3) 将引入的自变量作为控制变量,分别计算 Y 与未引入的自变量的偏相关系数,从中找出偏相关系数绝对值最大的自变量引入回归方程,再进行显著性检验,若 $\alpha \geq 0.05$ 则计算结束,获得子集 r ,将 r 存入集合 R ,进入第 5 步,否则将该自变量正式引入方程,进入下一步;

4) 除刚引入方程的变量外,对方程中其他变量分别计算它们对方程的贡献,从中找到最小值对应的变量进行显著性检验,如果 $\alpha < 0.1$ 则没有要剔

除的变量,回到第 2 步,否则剔除该变量,重复第 4 步;

5) 判断 R 中非重复变量的个数 m_0 是否大于 30 个(m_0 必须小于样本个数),若通过则终止循环,“最优”自变量子集为前一个集合 R ,否则在自变量集合中,剔除贡献最大的自变量,回到第 2 步.

通过上述步骤,分别对构成音色空间的 5 个主成分进行自变量的筛选,获得 5 组不同的优化子集,自变量个数在 26—28 之间. 然后用优化自变量子集进行回归分析,首先将子集中所有变量全部引入回归模型,不断剔除回归系数 t 检验的概率 P 值 α 最小的自变量,以 $\alpha \leq 0.05$ 、自变量个数和拟合优度检验中的调整决定系数 \bar{R}^2 (见(5)式)作为回归分析的终止条件,确定回归模型.

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2 / (n - p - 1)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)}, \quad (5)$$

其中 y_i 是因变量, \hat{y} 是建立模型预测值, \bar{y} 是因变量平均值, n 是样本个数, p 是模型中自变量个数. \bar{R}^2 取值在 0—1 之间, \bar{R}^2 越大说明回归模型对样本数据点的拟合优度越高,预测精度越高. 如果自变量个数增多,不会增加 \bar{R}^2 值,只有当回归方程中引入对因变量有重要“贡献”的自变量时,才会使 \bar{R}^2 值增加.

因此,在要求不同变量个数或者 \bar{R}^2 值的条件下,回归分析建立了不同的数学模型,为确定合适的音色属性模型,有必要建立不同情况的回归模型,提取不同条件下的 5 维音色特征,进行目标识别研究.

4. 音色模型在目标识别中的应用

利用实测三类(标记为 A, B 和 C)水下目标的平稳声音样本进行目标识别研究. 采样频率为 22050 Hz, 每个样本长度为 8192 点,其中 A 类 304 个, B 类 176 个, C 类 208 个. 根据不同终止条件建立的数学模型,提取音色特征作为识别特征,用基于误差反向传播训练算法三层神经网络分类器进行分类识别. 随机选取各类的 1/3 作为训练样本,选择与训练样本完全不重复的其他样本作为测试样本,取 50 次随机测试结果的平均值作为最终识别

结果,不同条件的识别效果如图 3 和图 4 所示.

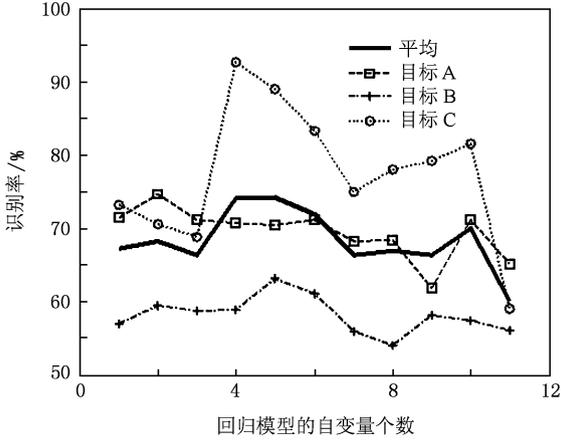


图 3 回归变量个数与识别效果关系图

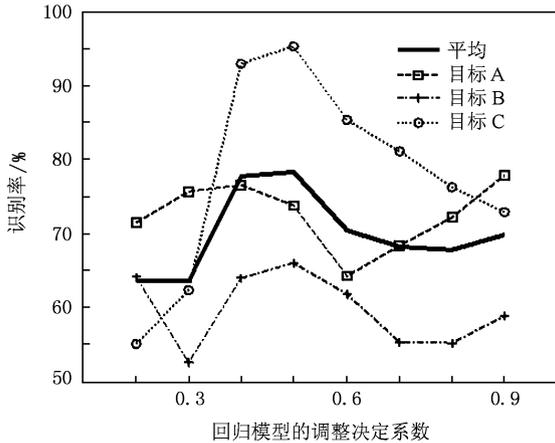


图 4 调整决定系数与识别效果关系图

理论上,音色模型提取的音色特征应该是目标的本质特征.模型越准确越能有效提取识别目标的音色特征.因此,在图 3 和图 4 上识别率越高的情形意味着建立的回归模型越接近真实的音色属性.分析图 3 和图 4,可以发现:1) 当变量个数较少或者较多时, \bar{R}^2 偏低或者偏高时,虽然可以正确识别目标,但识别效果相对较差. 2) 控制自变量个数时,依据 4,5 和 6 个自变量构成的回归模型所提取音色特征的平均识别效果较为接近;控制 \bar{R}^2 大小时,0.4 和 0.5 时识别效果较好. 3) 控制调整决定系数提取的音色特征的识别效果明显优越于控制自变量个数的识别效果.

这说明在评价样本个数较少的情况下,当变量个数较多或者 \bar{R}^2 较大时,虽然建立的模型对实验

样本的音色属性具有较高预测精度,但模型中包含了一些冗余自变量和干扰自变量,不能正确反映水下噪声的音色属性,所以识别率会较差.也就是说在小样本量下建立音色属性的回归模型时,变量个数不宜多, \bar{R}^2 不宜接近 1.

因此本文在少样本量情况下,回归分析的终止条件为 $\bar{R}^2 > 0.5, 6 > m_0 > 4$ 即可,分别建立的 5 个主成分的回归模型如下:

$$Y_1 = 0.1 + 0.9f_k\{\beta_\alpha^i\} + 0.82f_r\{\beta_{zcr}^i\}, \quad (6)$$

$$Y_2 = -0.57 + 0.72f_r\{\beta_{zcr}^i\} + 0.39f_k\{\beta_{sf}^i\} + 0.6f_{se}\{\beta_s^j\} + 0.45f_r\{\beta_{mfcc}^i\}, \quad (7)$$

$$Y_3 = -0.05 - 0.83f_v\{\beta_{sc}^i\} + 0.37f_{se}\{\beta_{sf}^i\} - 0.54f_r\{\beta_{sfm}^i\} + 0.84f_k\{\beta_s^j\}, \quad (8)$$

$$Y_4 = 0.13 + 0.82f_h\{\beta_{scbw}^i\} - 0.38f_h\{\beta_{se}^i\}, \quad (9)$$

$$Y_5 = 1.1 - 0.37f_{sk}\{\beta_{sfm}^i\} - 0.82f_k\{\beta_{si}^i\} - 0.59f_h\{\beta_{mfcc}^i\}, \quad (10)$$

其中 $Y_i (i=1, 2, \dots, 5)$ 分别表示音色属性的 5 个主成分, i 是时域信号分段的分段数, j 是以临界频段分析方式对频域信号分段的分段数,其中 $f_k(\beta_\alpha^i)$ 表示不同时段谱平整度扩展值的峰度系数, $f_r(\beta_{zcr}^i)$ 表示不同时段的过零点比率的最小值与最大值之比, $f_{se}(\beta_s^j)$ 表示不同临界频段的尖锐度的包络曲线斜率值, $f_v(\beta_{sc}^i)$ 表示不同时段的谱质心的方差值, $f_h(\beta_{se}^i)$ 表示不同临界频段的谱熵的几何平均值, $f_{sk}(\beta_{sfm}^i)$ 表示不同时段的谱平整度的偏度值,其他客观量值的含义可以通过表 2 和统计量的计算模型获得.

音色属性 5 个成分的回归模型((6)–(10)式)不能仅用某一个特征值来代表音色属性的 1 个成分,信号的时域和频域结构共同影响了音色属性,如(6)式定量描述了音色属性中的粗糙和沉闷的程度,其中谱平整度的扩展值越接近 1, $f_k(\beta_\alpha^i)$ 越大,反映声音不规则性的变化越陡峭,声音的粗糙程度越严重;(8)式定量描述了音色属性中的刺耳程度, $f_k(\beta_s^j)$ 值越大,反映声音不同频段的尖锐程度变化越陡,即声音的刺耳程度越严重.信号特征与音色之间的具体关系较为复杂,有待后期进行深入分析研究.

为保证上述模型的准确性,对其进行假设检验,结果见表 3.

表 3 回归模型的检验结果

音色属性的回归模型		Y_1	Y_2	Y_3	Y_4	Y_5
回归模型	调整决定系数	0.58	0.61	0.52	0.57	0.56
	F 检验概率 P 值	0.00	0.00	0.00	0.00	0.00
回归系数	最大 t 检验概率 P 值	0.00	0.00	0.01	0.00	0.01
	最大方差膨胀因子	1.11	1.50	2.11	1.18	1.35
	残差非参数检验概率 P 值	0.72	0.56	0.99	0.71	0.98

从表 3 数据可得出如下结论:1) \bar{R}^2 说明上述回归模型可以预测音色属性的主观评价分值. 2) F 检验的概率 P 值小于显著性水平 0.05, 认为各回归系数不同时为零, 因变量与自变量的线性关系是显著的, 同时标准化残差非参数检验概率 P 值大于显著性水平, 表明标准化残差与标准正态分布不存在显著差异, 认为残差满足了线性模型的前提要求, 所以可以建立线性回归模型. 3) t 检验统计量的概率 P 值均小于 0.05, 认为偏回归系数与零存在显著差异; 同时各自变量的方差膨胀因子远小于 10, 可以认为自变量之间不存在多重共线性问题, 不会影响方程的最小二乘估计. 因此, 建立的回归模型是可用的. 所以根据表 3 的检验结果, 说明对水下噪声音色属性建立的多元线性回归模型是正确、可用的, 并且能准确识别水下目标.

表 4 不同识别特征的识别结果

序号	识别特征	正确识别率 / %				归一化 方差
		A	B	C	平均	
1	谱质心	77.29	56.44	95.21	76.31	0.32
2	谱通量	65.10	44.61	30.47	46.73	0.26
3	谱熵	84.01	26.31	87.66	66.00	1
4	谱不规则性	57.08	24.61	36.27	39.32	0.23
5	谱平整度	68.03	46.29	96.36	70.23	0.53
6	零交点比率	67.91	55.93	97.34	73.73	0.38
7	振幅因子	58.23	23.31	59.60	47.05	0.36
8	音色	73.79	66.017	95.18	78.33	0.19

有效的识别特征不仅应具有较高的平均识别率, 三类目标各自的识别率也必须较高, 即离散程度小. 因此, 为验证音色特征的有效性, 将听觉特征中谱质心、谱熵、零交点比率等常用的识别特征, 与上述音色模型提取特征的识别结果进行比较, 具体识别结果见表 4 所示, 其中归一化方差反映了三类目标识别率的离散程度. 图 5 刻画了表 4 中的不同特征的归一化平均识别率和方差的变化情况.

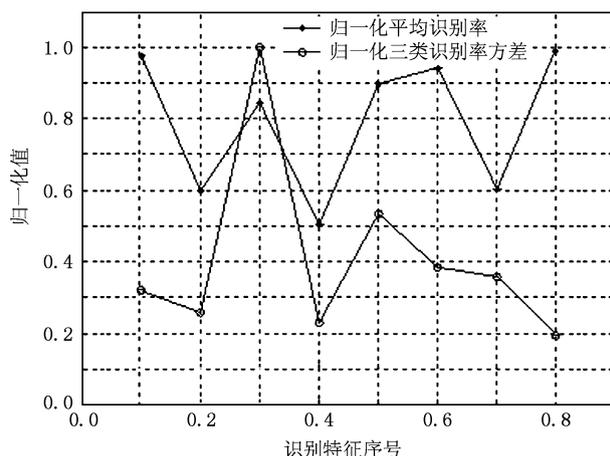


图 5 不同特征的归一化平均识别率和方差变化图

表 4 和图 5 表明, 本文建模提取的音色特征不仅具有高识别率, 同时方差值最小. 也就是说该特征有效改善了三类目标各自的识别效果, 达到提高平均识别率的目的. 故而, 音色特征的识别效果优越于某些常用的听觉特征, 适用于改善水下目标的识别效果.

5. 结 论

为模仿听觉功能提取音色特征用于目标识别领域, 通过主观评价计算获得音色属性空间的 5 个主成分的评价分值作为因变量, 同时计算能准确识别水下目标的听觉感知特征作为自变量; 然后提出一种改进的逐步筛选法, 从自变量集合中优化出自变量子集, 以自变量个数、调整决定系数和显著性水平作为终止条件建立音色属性回归模型; 最后, 以水下目标识别率作为衡量模型有效性的准则, 确定适当的音色模型. 研究发现, 在小样本量回归分析的情况下, 自变量个数和调整决定系数大小适中时, 音色模型较为合适, 提取的音色特征具有较好的识别效果和优越性, 并且通过对模型的假设检验证明模型是正确的、可用的.

本文为模仿听觉功能提高目标识别率的进一步研究建立了基础, 而且在海军培养声纳员时, 可依据声音音色属性差异对训练样本进行人工分类排序, 有针对性的训练声纳员, 提高耳听辨识目标的准确性, 同时也为乐音等领域建立音色属性数学模型提供了研究途径.

建立水下噪声音色属性的数学模型是一项长期的研究工作. 为完善本文工作, 后期还需要扩充

主观评价实验的样本类型和样本量,并考虑声纳员对水下噪声音色属性的主观感受,对音色属性的回

归模型进行改进,使之具有通用性和较高的预测精度,能够准确地、定量地计算水下噪声的音色属性.

- [1] Zhou L W 2004 *Targets Detection and Recognition* (Beijing: Beijing Institute of Technology Press) (in Chinese) [周立伟 2004 目标探测与识别(北京:北京理工大学出版社)]
- [2] Cai Y B, Zhang M Z, Shi X Z, Lin L J 1999 *Acta Electron. Sin.* **27** 129 (in Chinese) [蔡悦斌、张明之、史习智、林良骥 1999 电子学报 **27** 129]
- [3] Fan Y Y, Sun J C, Li P A, Xu J D, Shang J H 1999 *Acta Acoustica* **24** 611 (in Chinese) [樊养余、孙进才、李平安、许家栋、尚久浩 1999 声学学报 **24** 611]
- [4] Collier G L 2004 *Speech Commun.* **43** 297
- [5] Zwicker H E, Fastl H 1999 *Psychoacoustics: Facts and Models* (Berlin Heidelberg: Springer-Verlag Press)
- [6] Park T H 2004 *Ph. D. Dissertation* (Princeton: Princeton University)
- [7] Wang N 2006 *M. S. Thesis* (Xi'an: Northwestern Polytechnical University) (in Chinese) [王 娜 2006 硕士学位论文(西安:西北工业大学)]
- [8] Houtsma A J M 1997 *J. New Music Res.* **26** 104
- [9] Chen K A, Wang N, Wang J C 2009 *Acta Phys. Sin.* **58** 5075 (in Chinese) [陈克安、王 娜、王金昌 2009 物理学报 **58** 5075]
- [10] Grey J M 1977 *J. Acoust. Soc. Am.* **61** 1270
- [11] Daniel J F 1990 *J. Acoust. Soc. Am.* **87** 311
- [12] Mcadams S, Winsberg S, Donnadiou S, Soete G D, Krimphoff J 1995 *Psychol. Res.* **58** 177
- [13] Caclin A, McAdams S, Smith B K, Winsberg S 2005 *J. Acoust. Soc. Am.* **118** 471
- [14] Agostini G, Longari M, Pollastri E 2003 *Eu. Assoc. Sign. Proc.: J. Appl. Signal Proc.* **1** 5
- [15] Jensen K 1999 *Ph. D. Dissertation* (Copenhagen: University of Copenhagen)
- [16] Aucouturier J J, Pachet F, Sandler M 2005 *IEEE Trans. Multimed.* **7** 1028
- [17] Burred J J 2008 *Ph. D. Dissertation* (Berlin: Technical University Berlin)
- [18] Chen K A, Ma M, Zhang Y N, Wang N, Yan L 2008 *Acta Acoustica* **33** 348 (in Chinese) [陈克安、马 苗、张燕妮、王 娜、闫 靛 2008 声学学报 **33** 348]
- [19] Wang N, Chen K A, Huang H 2009 *Acta Phys. Sin.* **58** 5730 (in Chinese) [王 娜、陈克安、黄 凰 2009 物理学报 **58** 5730]
- [20] Victor W Y, Paul C H 2007 *J. Acoust. Soc. Am.* **122** 1502
- [21] Eronen A, Klapuri A 2000 *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing* Istanbul, Turkey, June 5—9, 2000 p753
- [22] Zhang X, Zbigniew W R 2007 *International Conference on Multimedia and Ubiquitous Engineering* Seoul, Korea, April 26—28, 2007 p3
- [23] Tindate A, Kapur A, Fuinaga I 2004 *International Computer Music Conference* Miami, Florida, USA, November 1—6, 2004 p592
- [24] Toh A M, Togneri R, Nordholm S 2005 *Proceedings of Postgraduate Electrical Engineering and Computing Symposium* Perth, Australia, September 7—10, 2005 p22
- [25] Giannakis K 2001 *Ph. D. Dissertation* (Middlesex: Middlesex University)
- [26] Liu D, Lu L, Zhang H J 2003 *Proceeding of International Symposium on Music Information Retrieval* Baltimore, Maryland, USA, October 27—30, 2003 p81
- [27] Wang N, Chen K A 2009 *J. Syst. Sim.* **21** 3128 (in Chinese) [王 娜、陈克安 2009 系统仿真学报 **21** 3128]
- [28] Wang N, Chen K A 2009 *Acta Armament.* **30** 144 (in Chinese) [王 娜、陈克安 2009 兵工学报 **30** 144]
- [29] Gao H X 2005 *Applied Multivariate Statistical Analysis* (Beijing: Peking University Press) p105 (in Chinese) [高惠璇 2005 应用多元统计分析(北京:北京大学出版社)第 105 页]

Regression model of timbre attribute for underwater noise and its application to target recognition *

Wang Na[†] Chen Ke-An

(*Department of Environmental Engineering, Northwestern Polytechnical University, Xi'an 710072, China*)

(Received 30 July 2009; revised manuscript received 18 August 2009)

Abstract

Timbre attribute is the most important feature to recognize a target. This paper presents a model of timbre features by multiple regression analysis applied in the recognition of underwater noise. At first, timbre attribute as a dependent variable is analyzed by the semantic differential evaluation and principal component analysis. And then an extended stepwise variables selection is proposed to select the optimal set as independent variables from auditory features that have been discussed in previous researches. Finally, the timbre features extracted by the regression model are used to recognize the underwater target. The results show that the extended regression analysis as a statistical method can find the relationship between timbre attribute and the auditory features. And the modeling timbre features calculated by several statistics of the sub-spectral features and the sub-temporal features are more effective than other features.

Keywords: timbre, multivariate linear regression, subjective evaluation

PACC: 8734, 4330

* Project supported by the Fundamental Research Foundation of Northwestern Polytechnical University (Grant No. W018104).

[†] E-mail: nawang@mail.nwpu.edu.cn