

基于音色属性辨识声目标的主观评价实验研究

陈克安, 王娜, 伍莹, 马苗, 张冰瑞

西北工业大学航海学院环境工程系, 西安 710072

E-mail: kachen@nwpu.edu.cn

2009-08-14 收稿, 2009-10-31 接受

西北工业大学基础研究基金资助项目(编号: W018104)

摘要 设计了目标辨识、语义细分和自适应分组成对比较等三组主观评价实验, 研究不同音色属性对辨识声目标的作用. 首先对 50 个水下声样本进行等响度、等时长预处理, 完成了三种目标类型的主观辨识实验. 然后完成基于语义细分法和自适应分组成对法的目标辨识实验, 利用相关分析和聚类分析手段对评价数据进行处理, 获得人耳听觉系统在自然音色和本质音色空间下辨识目标类型的结果. 研究结果有助于掌握和理解人类听觉系统听音辨物过程中音色属性的作用, 对音色属性建模及特征提取也有益处.

关键词

音色属性
水下噪声
主观评价
目标识别
聚类分析

声目标的自动识别在军事及民用领域都有重要应用, 在水下作战中尤其如此^[1]. 利用声信号识别目标的关键在于特征提取, 传统的方法基本上是基于信号处理与变换的方法, 如时域内的波形结构特征提取、频域内的谱分析和高阶统计量分析、时频域上的小波分析以及信号的非线性特征提取, 如分形、混沌等. 几十年来, 上述方法的潜力被人们充分挖掘, 取得了良好效果. 然而, 在低信噪比及强干扰条件下的识别效果仍然不能满足需求.

研究表明, 人的听觉系统在听音辨物方面表现出卓越的性能, 引发了人们研究模仿听觉原理辨识目标的强烈兴趣. 长时间以来, 研究者们针对一系列特定的环境声(如掌声^[2]、走路声^[3]、玻璃球的弹跳声^[4]、声纳声^[5]、人工合成的冲击声和空化噪声等^[6-8])进行了听觉辨识的研究, 证明特定条件下人耳识别目标的能力要优于目前的自动识别技术. 至于人耳对环境声识别的基本特点、识别模式及识别机理, Gaver^[9], Ballas^[10]和 Guastavino^[11]尝试进行了系统性的研究, 获得了初步的结论, 认为目标辨识是特征提取与分类方法综合作用的结果. 人对声事件的识别中, 依靠了所谓的广义听觉特征, 包括: (1)声特征(acoustic feature), 主要指声信号本身蕴含的目标特

征, 它与声音的响度、音调和音色相关联; (2)源特征(source feature), 这是与声源、环境、传播介质、声事件关联的特征; (3)高阶语义特征(higher-order semantic feature), 它与声事件的深层次内涵(语义信息)相关联, 主要由人的感知、意识、知觉方面决定的特征. 这些特征共同作用, 相互影响, 使得人耳识别声目标的物理机理异常复杂, 被认为是物理学领域的一大难题^[12]. 相对来说, 人对熟悉声识别(familiar sound recognition, FSR)主要由声特征决定, 其研究容易一些, 因此, 目前人们的研究集中于熟悉声识别, 重点关注声特征的提取.

目前, 基于听觉原理提取目标特征的方法主要有两种: 听觉模型法和心理声学参数法. 前者利用听觉模型计算出声音刺激的听觉响应作为目标特征, 后者利用一些反映人对声音主观感觉的心理声学参数(如尖锐度、波动强度、起伏强度等)作为目标特征. 可听声分为语声、乐声及环境声三类, 水下噪声属于环境声的一个类别. 在水下目标识别方面, 发表了少量基于听觉原理的目标特征提取方法的论文. 陆振波等人^[13]和王娜等人^[14]提出利用临界带宽上的响度(特性响度)作为特征用于水下目标识别; 马元锋等人^[15]提出利用听觉谱特征作为识别特征; 王娜等人

发现特性响度(specific loudness)、特性尖锐度和音调是识别目标的有效特征^[16],同时应用谱质心、谱平整度等6种反映音色属性的量作为听觉特征^[17]; Tucker等人^[18]利用谱通量、频域最低分量对应频率、时域质心等音色特征为目标特征. 归结起来,这些特征基本上都是以临界带宽上的响度值及其变换为基础,本质上都与声音的音色有关.

我们知道,人对声音的主观感觉可以用响度、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,主要取决于声音的幅度、频率和时长,而音色是多维感知属性,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”^[19]. 因此可以推测,音色属性在听音辨物的过程中应该占有主导地位,对音色属性的建模(音色参数化)是模仿听觉系统提取声特征的主要途径. 然而,目前的研究并没有确定给出定量的结论. 同时,在多维度的音色空间中,不同的音色属性在表征目标的能力方面有什么差异,哪些音色属性是辨识目标利用的主要特征,与人的主观感觉有什么联系等, these 问题是提取听觉特征必须解决的基础问题. 但迄今为止,已有的研究没有给出答案.

为此,本文设计三组主观评价实验,以水下噪声为对象,定量研究音色属性在目标识别中的作用,从一个侧面探索听音辨物的物理机理. 根据音色的定义可知,音色属性难以定量描述,因此,音色属性的表述、分类及作用是研究的难点及重点. 本文针对三类不同的水下噪声,在等响度、等时长条件下,研究基于音色属性的目标辨识问题. 需要说明的是,音调也是识别声目标的重要属性(对乐音尤其如此),但对于噪声,音调属性并不显著,且噪声的音色与音调无法截然分开,因此,本文指称的音色属性没有剔除音调成分. 本文首先通过听觉实验,确定人耳识别水下目标的能力;然后通过语义细分实验和聚类分析,获得基于音色属性描述词的目标识别效果;接下来,利用多属性下的自适应分组成对比较实验,获得不同音色属性下的目标识别效果. 最后,对三次实验结果

进行综合分析比较,探索音色属性在辨识水下目标的作用.

1 实验准备

本文共设计三组主观评价实验,分别为:(1)水下目标辨识实验,用于研究人耳利用除响度和持续时间以外的听觉属性识别目标的效果;(2)语义细分实验,用于研究水下噪声音色属性的构成及利用音色描述词分类目标的效果;(3)自适应分组成对比较实验,用于研究基于单一音色属性分类目标的效果. 三组实验所用的声样本、仪器和评价人员均保持一致.

考虑到水下目标主要分军用和民用两大类,以及水下声环境的特点,实验所用声音素材由50个水下噪声样本组成,分为3个类型,分别为水下环境声、潜艇声和民用船声(分别用A, B, C代表这三类声音). 每种类型的声样本数分别为16个、17个和17个. 上述声样本均选自Sound Ideas公司发布的水声音效库(内含98种水下噪声样本),根据产品说明,所有声音不含背景噪声,是为模仿听者在远场条件下水下声源的听音效果而开发的. 样本类型及说明如表1所示.

首先从样本库中挑选并截取特定时长的声信号,然后进行时长及响度调节. 为研究声音时长与辨识能力的关系,挑选了不同类型声样本,经反复实验发现,声音时长5s可满足分辨音色的要求,因此本文实验采用的声样本长度均为5s,采样频率为44.1kHz. 由于声音响度对识别效果有影响,为此,实验前对所有声音样本进行等响调节,响度计算采用Moore模型^[20],调节样本响度为15sone,误差限制在 ± 0.1 sone.

Tuliis等人^[21]提出:在主观评价实验中,20~30个评价者与168个评价者的评价结果的平均相关系数为0.95. 因此,本文被试(subject)选择无听力障碍的在校本科生和研究生24人,平均年龄25岁,男女比例为1:1. 实验前,为降低评价者所给数据的离散性,从每种类型的声样本中随机选出8个声音样本,共

表1 声音样本类型说明

声样本类型	类型代码	声源描述
水下环境声	A类	水下气泡、水的流动、水下冰裂、水下雨声、水流及波浪等
潜艇声	B类	潜艇不同方向行驶、不同运行方式
民用船声	C类	踏板船、独木舟、轮船及民船不同方向行驶和运行方式

24 个. 将这些样本随机排列, 每两个样本之间间隔 2 s, 然后循环播放, 并在播放的同时告知被试将听到的声样本所属类型. 共循环试听 3 遍, 让被试不断学习并熟悉声样本类型, 便于把握实验中评价样本的评分尺度.

实验时, 将选定的声样本由音频播放软件通过声卡输出至耳机放大器(B&K ZE0769), 由监听级耳机(Sennheiser HD250)重放. 评价者在有效隔绝外界噪声干扰的情况下, 聆听声音后填写问卷.

2 水下目标辨识实验

本次实验的目的在于观察被试直接聆听声音辨别声源类别的效果. 测听方式对辨识结果有很大的影响, 研究发现^[5,22]: 比对测听方式能明显提高非受训人员的辨识准确性, 但不影响受训人员的辨识效果. 这说明比对方式是人判断不熟悉声目标的有效手段, 由于本文研究的水下噪声对一般人群而言属于不熟悉声, 为了尽量排除语义特征对识别结果的影响, 实验采用比对测听方式进行目标分类.

实验中, 实验被试会依次听到 50 对声音, 每个声音时长 5 s, 间隔 2 s, 每对声音播完后, 被试将有 6 s 的时间对听到的声音进行所属类型判断, 然后将自己的主观判断填写到问卷表中. 表中的数字表示选择相应选项的可能性, “5”表示 80%~100%的可能性, 其余数字依次类推, 不填表示“不可能是”. 例如, 当被试听完一段声音后, 认为该声音是水下环境声的可能性有 80%, 是潜艇声的可能性有 20%, 则填表方式如表 2 所示. 由于本次实验所用到的声音均为水下声, 在日常生活中极少出现, 所以实验前先对各类型的声音进行循环试听 3 遍, 以使被试尽量熟悉声音, 这样可将本文的目标辨识实验尽量限制在熟悉声识别(FSR)的范畴里, 排除源特征及高阶语义特征对识别效果的影响, 将识别特征集中在音色属性上.

对 24 份评价数据进行处理. 首先对数据进行一致性检验, 剔除无效数据. 实验中, 50 个样本重复播放两遍, 依据相关分析和重复误判分析进行无效剔除. 同一被试两次判断结果的相关系数如图 1 所示, 2

号和 4 号被试的相关系数 ρ 均小于 0.3, 即对同一声音的两次判断相关程度最低, 故此处剔除 2 号和 4 号被试的辨识数据. 重复误判是根据同一被试前后两次对声样本的分类判断是否一致进行的, 两次分类结果不一样则记误判一次, 将得到的 50 个声样本的误判次数除以样本数便可得到重复误判率(图 2), 剔除重复误判率最大 3 号被试的评价数据. 综合起来, 实验中剔除的无效数据包括 2 号、3 号和 4 号被试的评价结果.

接下来, 对 21 份评价数据进行处理, 计算出三类声样本辨识率, 如表 3 所示. 可见, 对所有声样本的平均辨识率为 66.05%.

由于被试通过耳机聆听声音, 采用比对测听辨识目标, 并且所有声样本的响度及持续时间均相等,

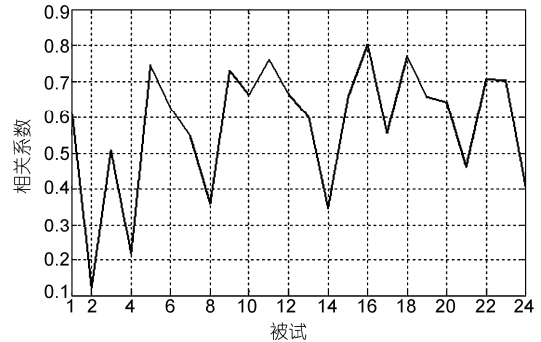


图 1 不同被试两次辨识的相关系数 ρ

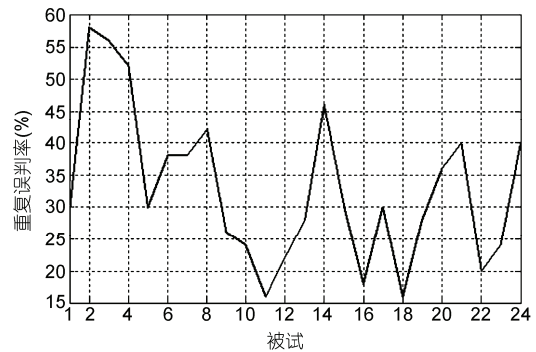


图 2 不同被试的重复误判率

表 2 样本辨识率调查问卷填写示例

序号	水下环境声					潜艇声					民用船声				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
01				√					√						

表3 三类声样本平均辨识率

声音类别	辨识率(%)			平均
	A类	B类	C类	
A类	60.27	27.17	15.97	65.94
B类	30.06	62.61	9.10	
C类	9.67	10.22	74.93	

因此,听觉系统辨识目标依据的特征应该是音色特征. 实验中 A 类和 B 类声目标的辨识率较差,而 C 类的辨识率最高,主要原因在于从音色的角度看, A 类和 B 类目标的音色差异性较 C 类目标声音的小,这一点从三类声音的时频分析及部分音色特征参数计算中可以看出^[14,16]. 另外, C 类声音较另外两类声音而言,在日常生活中接触较多,人们的熟悉程度更强,即使实验中对被试进行了培训,但对 A 和 B 两类声音的熟悉程度较 C 类仍然有差距,这也应该是 A 类和 B 类声目标辨识率较差的原因之一.

3 基于语义细分的目标分类实验

声音的音色属性具有丰富的内容,可以从不同侧面用形容词加以描述(成为音色描述词). 描述的方式有两种:一种是从日常及文学语言中提炼出的所谓“自然音色”;另一种是在自然音色描述词的基础上,通过数学变换(如主成分分析)获得的相互独立的所谓“本质音色”. 自然音色属性的描述词语义清晰,其含义易于掌握;本质音色属性在音色空间是彼此独立的,其个数是支撑音色空间的最小维度数,但其音色描述词不易掌握和理解^[23]. 为了观察不同音色属性对目标分类发挥的作用,接下来的两个实验分别从自然音色和本质音色两个角度出发.

对于自然音色属性,考虑到评价工作量,我们选择采用语义细分法进行评价实验. 语义细分法是目前常用的主观评价方法之一,它用几个在语义上不同的形容词来描述对一个声音的主观感觉,用于判断声音的哪一种属性更为测试人群所看重^[24]. 语义细分法的实现分为两个阶段:首先确定评价术语(音色描述词);然后在选定的尺度上,对每个评价术语进行等级评价,确定声样本在该评价术语空间上对应的点^[25];最后,利用评价数据得到每个音色描述

词下的目标识别效果.

3.1 水下噪声自然音色描述词的获取

在水下噪声音色属性的研究中,适当、全面的评价用语是获得准确的声音主观评价实验结论的关键. 本文在汉语语境下进行了水下噪声听觉属性评价用语的调查研究. 首先根据已有的声品质评价词和问卷调查结果获得大量噪声描述词,通过声音在听觉感受上的“三要素”(响度、音调和音色),确定其中 97 个词适用于评价噪声的听觉属性. 然后采用成对比比较实验和聚类分析手段,依据人对词的理解差异将词汇分成 10 类,避免了研究者个体分类词汇的不确定性. 最后为确定少量的、语义不重叠的、具有代表性的水下噪声听觉属性评价用语,通过水下噪声重放,进行选词问卷调查,从中确定水下噪声听觉属性的评价术语词库.

根据水下音色属性的分析结论、评价术语及其选择概率的研究,从上述词库中选取 10 个评价词作为水下噪声音色描述词^[25],分别为“震颤的”、“粗糙的”、“刺耳的”、“急促的”、“变化的”、“不平缓的”、“嘈杂的”、“沉闷的”、“紧凑的”和“重复的”,分别用 f_i ($i=1, 2, \dots, 10$)表示. 上述音色属性也称为自然音色属性.

3.2 语义细分实验及目标分类

实验中,向被试依次播放 100 个声音. 每个声音播完后,被试将有 5 s 时间对该声音在不同的描述词下进行打分判断,按照表 4(以“震颤的”为例)所示,根据主观感受给出相应的分数.

采用与上节一样的方法进行数据的有效性检验,以“震颤的”为例,剔除 9 号、22 号和 23 号被试的评价结果.

语义细分的实验数据表现了每个描述词下评价样本的距离信息,因此,可以利用聚类分析方法将声样本进行分类. 聚类分析能够将一批样本(或变量)数据根据其诸多特征,按照其性质上的亲疏程度、在没有制定分类标准的情况下实现自动分类,产生多个分类结果. 评测聚类效果常用的方法是:选择人工已

表4 语义细分实验评分标准

震颤的	非常	有点	不知道	有点	非常	不震颤的
	1	2	3	4	5	

表 5 聚类结果评估表

聚类类别	人工类别		
	A	B	C
1	n_1^A	n_1^B	n_1^C
2	n_2^A	n_2^B	n_2^C
3	n_3^A	n_3^B	n_3^C

经分好类或者做好标记的集合作为测试集合, 聚类结束后, 将聚类结果与已有的人工分类结果进行比较. 本文中强迫聚类为三类, 分别为 1, 2 和 3 类, 然后由于样本类型已知分为 A, B 和 C 类对聚类结果进行评估, 如表 5 所示. 聚类结果的第 1 个类别中, 标记为第 A 个类别的数据数为 n_1^A , 标记为 B 类的的数据数为 n_1^B , 标记为 C 类的的数据数为 n_1^C , 寻找三者中最大值作为聚类类别 1 的代表, 即聚类结果为该数据所属的人工类别. 如果存在两个相同的最大数据数时, 按照列数据数最大值作为代表进行类别判断.

根据上述聚类结果与人工类别的分析, 可以计算出每个自然音色描述词下样本的分类结果, 见图 3. 可以看出, 每个自然音色属性都有一定的辨识目标的能力, 不过, 不同的音色属性所起的作用大不相同. 除 f_1, f_3 和 f_7 外, 有 7 个自然音色属性对 B 类目标的识别率都在 80% 以上; f_5 和 f_6 对 C 类目标的识别率高于 70%; f_1, f_3 和 f_7 对识别 A 类目标起主要作用, 其中 f_3 的作用最突出. 综合 10 个自然音色描述词, 计算其距离信息并作聚类分析, 其平均分类率为 55.39%.

4 基于成对比较的目标分类实验

前面得到的自然音色属性, 是人们为了形容水下噪声音色而采用的最为集中的词汇. 但是, 从数学的角度看, 这些自然音色属性之间彼此包含的信息有冗余, 相互之间并不独立. 为此, 可以采用统计分析的手段将其变换为相互独立的成分, 在此基础上进行目标分类.

4.1 水下噪声本质音色属性提取

主成分分析是一种常用的压缩维度的方法, 用于寻找几个能解释原有变量的线性组合函数(主成分), 这些线性组合一方面能够保持原来变量的信息, 而且主成分之间相互独立不会重叠, 更重要的是, 它能够以“较少”的主成分代替原来“多个”解释变量, 以达到精简的目的.

对上一节的语义细分结果进行 KMO 检验及 Bartlett 球度检验, 实验数据的 KMO 值为 0.792, 已经达到“适中”的标准; 而 Bartlett 的球度检验统计量的观测值为 742.341, 相伴概率为 0.000, 因此可以拒绝零假设, 实验数据适合进行主成分分析. 各成分解释原有变量总方差的情况如表 6 所示, 可见采用 4 个主成分时可解释 96.768% 的变异量. 因此, 对于水下噪声的 10 个自然音色描述词, 可以用这 4 个相互独立的主成分来表示.

采用最大方差法对成分负荷旋转, 得到旋转后 4 个主成分的负荷矩阵, 如表 7 所示. 表 7 给出了各变

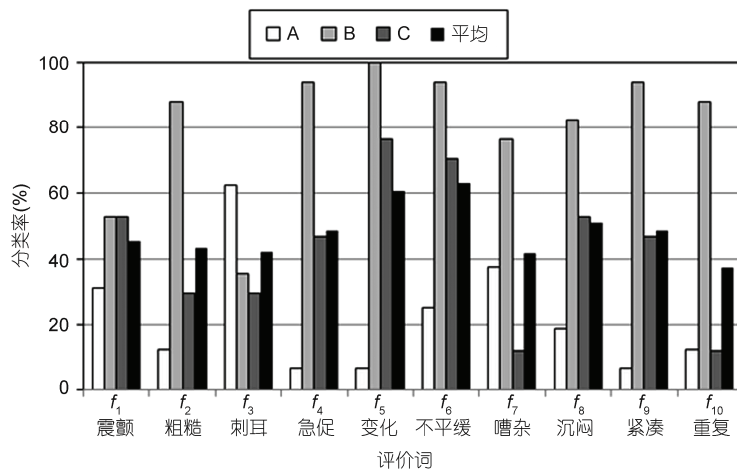


图 3 不同评价词的评价结果聚类分析的平均分类率

表6 成分解释原有变量总方差的情况

成分	初步主成分提取			旋转后主成分提取		
	特征值	方差贡献率(%)	累计方差贡献率(%)	特征值	方差贡献率(%)	累计方差贡献率(%)
1	5.294	52.939	52.939	4.049	40.491	40.491
2	2.700	27.000	79.939	2.531	25.305	65.796
3	1.440	14.400	94.339	2.117	21.166	86.963
4	0.243	2.429	96.768	0.981	9.805	96.768

表7 旋转后的主成分负荷矩阵

描述词	符号	成分			
		F_1	F_2	F_3	F_4
震颤	f_1	0.958	0.138	-0.174	0.036
粗糙	f_2	0.873	0.306	-0.281	-0.034
刺耳	f_3	0.823	0.489	0.059	-0.151
急促	f_4	0.219	0.959	0.116	0.035
变化	f_5	-0.353	-0.092	0.871	-0.290
不平缓	f_6	-0.121	0.345	0.901	-0.188
嘈杂	f_7	0.835	0.516	-0.017	-0.040
沉闷	f_8	0.843	-0.098	-0.493	0.055
紧凑	f_9	0.296	0.909	0.086	0.206
重复	f_{10}	-0.095	0.176	-0.410	0.888

量在主成分上的负荷(表中加粗部分的数据),这就是水下噪声本质音色属性与自然音色属性的关系.由此可知,第一主成分由“震颤”、“粗糙”、“刺耳”、“嘈杂”和“沉闷”构成;第二主成分由“急促”和“紧凑”构成;第三主成分由“变化”和“不平缓”构成;第四主成分则由“重复”构成.

4.2 多属性下的自适应分组成对比较实验

由于本质音色描述词比自然音色描述词少了许多,因此可以采用更为准确的主观评价方法评价不同音色属性下样本之间的几何距离.不同于语义细分法对单个声音进行语义尺度上的判断,成对比较法(paired comparison, PC, 又称为对偶比较法)是一种两两成对比较的评价方法,它将声音样本两两成对组合后,评价者听完一对声音后,对这两个样本的某一听觉感知属性进行比较并给予(不)相似程度的评判.该方法简单易行、评价结果准确度高,在众多社会科学领域已获得广泛应用.因此,为了尽可能地消除识别方式对结果的影响,这里提出利用PC法在不同本质音色属性下进行成对比较,从而获得不同本质音色属性下的分类效果.

需要指出的是,PC法的比较次数随样本数量的平方增长,对大样本量的评价,实际操作并不可行.因此,人们提出了分组成对比较法(grouped paired

comparison, GPC)法^[26],将样本分成多个样本组,再对各样本组独立进行成对比较评价,通过在各样本组之间预设关联样本,由各组关联样本的评价结果重建获得整体样本评价价值,从而有效降低实验工作量.

在实现GPC法时发现,当预设的关联样本不合理时,该方法评价结果的准确性会降低甚至失效,为此,文献[27]提出了自适应分组成对比较(adaptive grouped paired comparison, AGPC)法.该方法是GPC法的改进,它将待评价的样本分成多个样本组后,不预设固定的关联样本,而是按照一定的准则,根据前一样本分组的评价结果选出最合适关联样本(也称为种子样本)放入下一分组以建立联系,在整个实验的进行过程中不断地调整关联样本,最终通过每两组之间的关联样本,由各组样本的评价结果进行重建获得全体样本的评价结果.在有效减小评价时间和实验强度的同时,最大程度地避免了由于关联样本选取不当造成的误差^[27].另外,当评价对象(如本文中的音色)具有多个维度的属性时,AGPC的种子选取规则需要考虑到不同属性的共同作用.

实验开始前,利用语义细分法实验获得的声样本不相似度,将声样本尽可能均匀地分到不同小组中.将50个声样本分6组进行,除第一组为10个声样本外,其他5组均为8个声样本.从第二组开始,认为不同音色属性对AGPC法种子选取的影响相等,然后按AGPC法种子选取原则及操作步骤,依据前一组的实验结果挑选出两个声样本作为最佳种子,评价术语采用本质音色(表7中的 $F_1 \sim F_4$)描述词.由于本质音色描述词无法用单一词汇或简洁的语言表述,实验中按表7中成分给被试进行语义解释,然后让被试依据不同的本质音色属性进行评价,评价打分如表8所示.

通过以下三个方面的标准检验不同被试评价结果的可靠性检验,即相同声事件比较($i-i$ 比较)、不同

表 8 AGPC 实验不相似性评价的评分标准^{a)}

打分可能性	1 非常相似	2 很相似	3 有点相似	4 不太相似	5 完全不相似
-------	-----------	----------	-----------	-----------	------------

a) 如果感觉两个声音完全一致可以打 0 分

回放顺序的比较(ij - ji 比较)、三角循环误差检验. 采用加权一致性系数的方法进行数据的筛选, 以全面反映评价中各种误判对统计结果的影响. 加权误判率 C_w 为

$$C_w = \frac{\sum C_i \cdot E_i}{\sum E_i}, \quad (1)$$

其中 E_i 为第 i 种误判可能产生的次数, C_i 为第 i 种误判的实际误判率. 加权一致性系数为

$$\zeta_w = 1 - C_w. \quad (2)$$

最终剔除加权一致性系数低于 0.7 的被试的数据, 即剔除约 10% 被试的结果后, 经过反演获得各声样本在不同音色属性下的评价分值.

4.3 基于成对比较的目标识别

根据上一节得到的不同音色属性下的评价分值, 计算出声样本的欧氏距离信息, 然后进行聚类分析, 获得不同本质音色属性下的分类率(表 9), 其中 F 表示由 4 个主成分 $F_1 \sim F_4$ 构成的音色空间. 分析表 9 数据可见, 4 种不同音色属性的聚类结果比较接近, 其中 F_4 属性的分类效果较好, 三类目标的平均分类率为 40.44%.

5 分析与总结

本文设计了三组主观评价实验, 探讨水下噪声音色属性在识别目标类别中的作用. 实验一让被试通过耳机聆听等响度及等持续时间的声样本, 采用比对测听的方式辨识目标, 获得人类听觉系统利用音色特征辨识目标的结果, 三类目标的平均辨识率为 66.05%. 实验二利用语义细分法对水下噪声的 10 种自然音色属性进行等级评分, 通过聚类分析获得不同自然音色属性下的平均辨识率. 将自然音色属性利用主成分分析法压缩为彼此独立的 4 种本质音色属性后, 实验三利用多属性下的 AGPC 法进行本

质音色属性下的声样本不相似度主观评价, 随后由聚类分析获得 4 种本质音色属性下的平均辨识率. 为了便于比较, 将实验二中的自然音色属性下的辨识率通过表 7 的关系转换为本质音色属性下的辨识率, 在 $F_1 \sim F_4$ 尺度上的平均分类率分别为 49.27%, 53.06%, 61.28% 和 39.58%, 由此构成的音色空间中(F)的目标辨识率为 55.39%, 结果标示于图 4 中.

图 4 说明, 实验一获得的识别效果最好, 单个音色属性的辨识率都要低. 对于语义细分实验和 AGPC 实验而言, 将其识别结果统一到本质音色空间中, 总体而言, 语义细分实验的辨识效果要好于 AGPC 实验辨识效果. 本文的实验及结果表明:

(i) 在完成实验一时, 其实验条件的设置是使音色属性特征(声特征的一部分)与其他特征尽量分离开来. 在现实的声学场景中, 人类听觉系统还将利用除音色以外的其他声特征、源特征和高阶语义特征对目标进行综合性辨识, 这些特征的表述及其量化对提高目标自动识别系统的性能将有很大帮助, 但更加困难.

(ii) 人对音色描述词的理解与掌握程度对其辨识目标有很重要的影响. 虽然在音色空间中, 本质音色属性彼此独立, 维度数更少, 但由于被试对自然音色描述词的理解和掌握要优于本质音色描述词, 因此, 利用自然音色获得的目标辨识率要高于本质音色的辨识率. 从评价方法上讲, 在评价过程中使用语义细分法, 由于不同被试对本质音色的理解不尽一致, 致使其使用的评价标准将不能保持统一, 从而使评价准确度低于成对比较法; 反过来说, 如果被试对音色描述词的理解程度准确, 保持一致, 语义细分法获得的目标辨识率将低于成对比较法, 图 4 中 F_4 的辨识结果说明了这一点, 因为本质音色 F_4 与自然音色 f_{10} 是基本等同的.

(iii) 从图 4 可以看出, 利用单一音色属性获得

表 9 AGPC 法各音色属性下的平均分类率

音色属性	F_1	F_2	F_3	F_4	F
平均分类率(%)	42.40	40.44	40.44	46.32	40.44

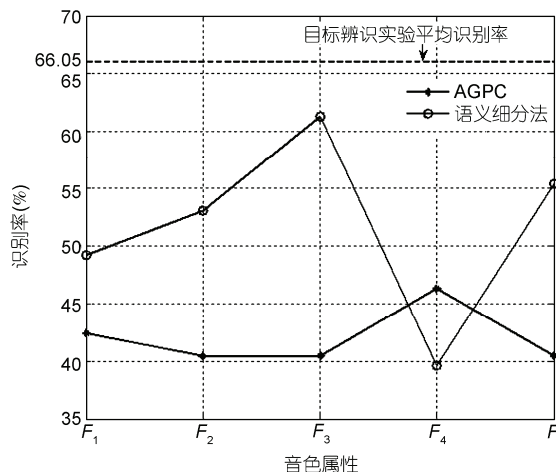


图4 不同音色属性评价结果聚类分析的平均分类率

的最高识别率(如语义细分法下的 F_3)比实验一获得的识别率要低,这说明不同音色属性的共同作用对提高识别率有帮助,但作用机理及效能需进一步研究.另外,为进一步完善本文工作,还需要扩大实验样本量和样本的类型,同时针对音色属性的评价改进并设计更加准确的主观评价实验方法.

(iv) 本文结果说明,在对声目标进行主观辨认时,人们利用了易于掌握的自然音色属性.在音色参数化过程中,应该首先以自然音色属性参数化为基础,然后将自然音色属性模型转换到本质音色属性上.这样,音色属性建模获得的目标特征既可以保持较高的目标识别率,又能保持较低的维度数,降低硬件系统的运算量.

致谢 感谢所有参加主观评价实验的西北工业大学航海学院的本科生和研究生.

参考文献

- 1 周立伟. 目标探测与识别. 北京: 北京理工大学出版社, 2004
- 2 Repp B H. The sound of two hands clapping: An exploratory study. *J Acoust Soc Am*, 1987, 81: 1100—1110
- 3 Li X, Logan R J, Pastore R E. Perception of acoustic source characteristics: Walking sound. *J Acoust Soc Am*, 1991, 90: 3036—3049
- 4 Warren H, Verbrugge R. Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *J Exp Psychol Hum Percept Perform*, 1984, 10: 704—712
- 5 Collier G L. A comparison of novices and experts in the identification of sonar. *Speech Commun*, 2004, 43: 297—310
- 6 Lutfi R A, Oh E, Storm E, et al. Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians. *J Acoust Soc Am*, 2005, 118: 393—404
- 7 Lutfi R A, Liu C J. Individual differences in source identification from synthesized impact sounds. *J Acoust Soc Am*, 2007, 122: 1017—1028
- 8 Howard J H. Perception of simulated propeller cavitation. *Human Factors*, 1983, 25: 643—656
- 9 Gaver W. What in the world do we hear? An ecological approach to auditory event perception. *Ecol Psychol*, 1993, 5: 1—29
- 10 Ballas J. Common factors in the identification of an assortment of brief everyday sounds. *J Exp Psychol Human Percept Perform*, 1993, 19: 250—267
- 11 Guastavino C. Categorization of environmental sounds. *Canad J Exp Psychol*, 2007, 61: 54—63
- 12 陈克安. 人类听觉系统识别声音的物理机理及其定量表达. 10000个科学难题·物理学卷. 北京: 科学出版社, 2009. 368—369
- 13 陆振波, 章新华, 胡洪波. 水中目标辐射噪声的听觉特征提取. *系统工程与电子技术*, 2004, 26: 1801—1803
- 14 王娜, 陈克安. 心理声学参数提取及其在目标识别中的应用. *计算机仿真*, 2008, 25: 21—24
- 15 马元锋, 陈克安, 王娜, 等. 听觉模型输出谱特征在声目标识别中的应用. *声学学报*, 2009, 34: 142—150
- 16 王娜, 陈克安. 分段谱质心特征在水下目标识别中的应用. *兵工学报*, 2009, 30: 144—149
- 17 王娜, 陈克安. 听觉感知特征在目标识别中的应用. *系统仿真学报*, 2009, 21: 3128—3132
- 18 Tucker S, Brown G J. Classification of transient sonar sounds using perceptually motivated features. *IEEE J Ocean Eng*, 2005, 30: 588—600
- 19 ANSI. Psychoacoustical terminology S3.20. New York: American National Standards Institute, 1973
- 20 马元锋, 陈克安. Moore 响度计算模型的改进. *应用声学*, 2009, 28: 59—63
- 21 Tullis T T, Wood L. How many users are enough for a card-sorting study? In: *Proceedings of the UPA Conference*, Minneapolis, USA, 2004. 1—10
- 22 陈克安, 王娜, 王金昌. 人耳辨识声目标能力的实验研究. *物理学报*, 2009, 58: 5075—5082

- 23 王娜, 陈克安, 黄凰. 水下噪声听觉属性的主观评价和分析. 物理学报, 2009, 58: 7330—7338
- 24 孟子厚. 音质主观评价的实验心理学方法. 北京: 国防工业出版社, 2008
- 25 陈克安, 马苗, 张燕妮, 等. 汉语语境下的车辆噪声听觉属性评价与分析. 声学学报, 2008, 33: 348—353
- 26 毛东兴, 高亚丽, 愈悟周, 等. 声品质主观评价的分组成对比较法研究. 声学学报, 2005, 30: 515—520
- 27 黄煜, 陈克安, 闫靓, 等. 自适应分组成对比较法: 原理及种子的选择. 声学学报, 2008, 33: 443—449

Subjective evaluation experiments of timbre attribute based acoustic target identification

CHEN KeAn, WANG Na, WU Ying, MA Miao & ZHANG BingRui

Department of Environmental Engineering, College of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Three subjective evaluation experiments involving in target recognition, semantic differential (SD) and adaptive grouped paired comparison (AGPC) were designed and performed to investigate the effect of timbre attributes on acoustic target discrimination. Firstly, 50 sound samples of equalized loudness and duration were used to subjectively identify three kinds of targets. Then SD and AGPC based target discrimination experiments were carried out, and the experimental data were processed using coherence and clustering approach to explore the roles of the two kinds of timbres, natural timbre and essential timbre, in the identification of acoustic targets. This work provides some insight into timbre mechanisms inherent in auditory object identification and timbre modeling applied to auditory feature extraction and target recognition.

timbre attribute, underwater noise, subjective evaluation, target recognition, clustering analysis

doi: 10.1360/972009-1485