

基于不相似度评价的水下声目标分类与听觉特征辨识*

杨立学[†] 陈克安 张冰瑞 梁雍

(西北工业大学航海学院环境工程系, 西安 710072)

(2014年1月6日收到; 2014年3月26日收到修改稿)

本文旨在探索人类被试对水下声目标的感知分类及在该过程中所利用的听觉特征. 首先设计了成对比较实验. 然后利用 CLASCAL 算法对实验获得的不相似度评分进行建模, 获得感知空间, 并分析了 3 个公共维度、特异性和 3 个被试潜类各自的特性及其在目标感知分类中所起的作用. 最后, 基于 Gammatone 听觉滤波器组对声样本进行分析, 发现了能够有效描述 3 个公共维度以及节拍特性的听觉特征, 并利用它们构造决策树对新样本实现了分类, 从而为实际中如何应用这些特征提供了指导.

关键词: 水下声目标分类, 不相似度, 多维尺度分析, 听觉特征

PACS: 43.66.+y, 43.30.+m

DOI: 10.7498/aps.63.134304

1 引言

水下声目标自动识别是现代海战中的关键环节之一. 为改善传统特征提取方法, 模仿人耳听觉感知原理提取特征日益受到重视. 目前, 相关研究主要从两个方面开展. 一方面, 由于人耳在低信噪比或存在干扰声源的背景下仍然能稳定地识别目标, 因而人们希望通过模拟人耳对声音的处理过程建立计算听觉模型, 代替传统的信号分析技术作为分析前端, 从而提高特征的鲁棒性. 李朝晖和迟惠生对近 30 年听觉外周计算模型的研究及其在语音识别领域的应用进行了综述, 并讨论了在被动声呐目标识别任务中应用听觉模型的可能性^[1]; Mill 和 Brown 则尝试将计算听觉场景分析 (computational auditory scene analysis, CASA) 理论应用到水下信号的分析中, 以排除干扰声源对目标特征提取的影响^[2].

另一方面, 人类大脑对于声音的表达维度很高, 感觉丰富, 并可以从高维特征空间中找到类别共性和类别异性, 即能够进行高效的特征选择和

降维, 这对于水下多类目标识别是十分有益的^[3]. 因而, 如果能够辨识出人在该过程中所利用的听觉特征, 将大大提高自动识别系统的识别性能和运算效率. 听觉特征辨识需借助于不相似度主观评价实验, 而多维尺度分析 (multidimensional scaling, MDS) 模型则可用相互正交的低维度空间反映不同声音间的不相似关系, 其中每个维度 (称为感知空间维度) 代表着听者在评价时所利用的听觉特征^[4]. 相比于其他统计方法, MDS 的优点是它不需要关于维度数及其性质的先验知识和假设, 因此在分析过程中受到的约束较少.

MDS 法在乐器声听觉感知领域应用较广, 相关研究提取的特征在乐器声识别和音频检索领域得到广泛应用^[5], 而一些研究者也将该方法用于水声信号的相关研究中. Howard 发现由 8 个水下信号构造的 2 维听觉特征空间可分别解释为谱的双峰度和低频周期性^[6]; Tucker 得到一个关于水下瞬态信号的 3 维空间, 分别由谱通量、最低峰值频率和时域质心表示^[7]. 此外, 一些研究者在解释感知空间时利用了语义细分法, 即用一些描述主观感受的形容词来解释感知维度, 例如, 王娜等将水下噪声

* 国家自然科学基金 (批准号: 11074202) 资助的课题.

[†] 通讯作者. E-mail: yanglixue.2008@163.com

音色空间的5个维度分别解释为沉杂的、起伏的、快变的、尖锐的和规律的^[8]。然而,由于这些研究在挑选样本时遵循了不同的标准,并且采用的分析方法也有所差异,因而得出的结论不能相互扩展和印证。同时,这些研究在感知空间特性(如空间维度和被试个体差异等)分析方面没有与目标分类进行有效的结合,因而所提取的特征在实际应用中可能达不到很好的效果^[9,10]。

在利用MDS法提取分类特征时,声样本所代表的目标类型必须精心挑选。水下声目标种类繁多,如潜艇、水面舰艇、鱼雷、商船、渔船,以及水流和海洋生物等^[11],但不存在统一的分类法。一般来说,来自同一类目标的声音在感知上是相似的,而不同类目标的声音则听起来差异较大,因而可基于感知不相似度对目标类别进行划分。本文将通过挑选几种典型目标的声样本,通过基于不相似度评价的多维尺度分析法,获得样本在感知空间的聚类情况,并分析空间相关特性对目标分类的影响,最后提取能够描述感知空间每个维度的听觉特征,而这些特征可直接用于目标分类或为其提供有价值的指导。

2 实验设计

本节将通过开展不相似度评价实验研究被试对水声目标的感知分类,其中样本集的构成是影响结果分析的关键因素,需要遵循具有目的性的准则;此外,本文还将关注被试在目标分类中的个体差异,因此同样需要对被试的构成进行探讨。本节最后也对实验设备以及具体的流程进行了介绍。

2.1 样本选择

研究的第一步需要确定一个有代表性的样本集。本文在样本挑选过程中遵循了两个主要原则:1)样本覆盖主要的目标类型,每类样本又包含不同目标,每类的总数要按其常见性和重要性确定;2)总的样本个数不宜过大,因为成对比较次数会随样本数呈平方方式增长,从而引起被试疲劳和结果不稳定。

按以上原则,首先将水下目标类型分为人为目标和自然目标两大类,每类样本均为10个,其中前者又可分为1)水面舰船:驱逐舰(2个)、客轮、拖轮和巡逻艇(2个),2)深度和航速均不同的潜艇声(4个);后者分为1)黄花鱼和虾,2)大型哺乳动物:白

鲸、领航鲸、逆戟鲸和瓶鼻海豚,3)水下背景声:气泡(2个)、水流和小雨。所有的20个样本均来自文献^[12],具有较高的信噪比;每个样本长度为5 s,并进行了等响度处理^[13];实验中信号以较为舒适的声压级播放(接近75 dB SPL)。

2.2 被试

本次实验招募24名西北工业大学航海学院环境工程系的研究生及本科生为被试,男女比例为13:11,年龄在20—25之间。这些被试对水下声音并不熟悉(称为新手),不过其中有3人多次参加过类似实验,而其他被试没有或只有一次实验经历,对实验目的及过程的不熟悉可能导致他们与这3名被试的结果有很大的不同,此外,他们自身也可能因性格、性别等因素的差异存在区别。尽管实验中的被试与声呐员在特定目标的识别上存在较大差距,但在当目标类型差异较大(如人为声和生物发声的辨别)时性能相当^[14];此外,研究表明,在进行不相似度评价任务时,新手和声呐员在空间维度的组成上并无差别,而差异则体现在对不同维度的感知权重上^[15]。因此,选新手作为实验被试所得的实验结果仍然具有指导意义。

2.3 实验设备

实验在绝对安静的房间中进行,声样本由计算机声卡输出给BEHRINGER HA4700耳机放大器,然后通过SENNHEISER HD280高保真耳机播放给被试。

2.4 实验过程

不相似度评价实验需借助于成对比较法。该方法将声音两两组合,评价者在听完一对声音后,对这两个声音的差异进行比较评价。该方法操作简单,理论基础完善,因而在很多社会科学领域获得应用。本文的实验流程如下:首先,顺序播放实验所用的20个样本两遍,使被试熟悉样本,并能感知它们的变化范围;然后,被试将要完成随机呈现的260对声音的不相似度评价,其中包括140对不同声音的单向比较,20对自身比较以及50对不同声音的双向比较(用于数据检验)。每对声音间隔2 s,播放完每对声音后留给被试5 s的时间进行不相似度评价;考虑到样本间的变化较大,评价采用7级尺度,即1代表极度相似,7代表极度不相似。正式实验开始前,被试尝试进行了5个声音对的评价,

均表示能够顺利完成评价任务. 为防止被试疲劳对实验结果可靠性的影响, 每完成65次成对比较后休息5—10 min, 整个实验耗时2 h.

3 数据处理与结果分析

成对比较实验获得了被试的不相似度评价数据, 本节将通过CLASCAL模型对其建模获得感知空间表达, 以观察被试对目标的感知分类, 并分析空间中的公共维度、特异性和潜类权重各自的特性及其对目标分类的影响.

3.1 数据剔除

为保证MDS分析的稳定性, 有必要对由于不认真或者不能理解实验任务的被试的评价数据进行剔除. 该过程包括两个阶段:

1) 被试自身一致性: 计算50个声音对在不同播放顺序时的评分的相关系数, 发现14, 16, 17, 19号被试对应的值没有显著大于0 ($p > 0.05$), 说明他们未能在整个实验中保持统一的评价准则^[8].

2) 被试间一致性: 计算所有被试间260对评价得分的Person相关系数, 通过层次聚类分析发现12, 14, 17, 19号被试与其他被试有显著差异, 说明他们采取了与其他被试截然不同的评价策略或者不能准确理解实验任务^[16].

综合以上考虑, 剔除12, 14, 16, 17, 19号被试的数据. 图1给出了所保留被试的评分数据分布情况的箱线图, 其中线段两端代表数据上下界, 箱体边界分别代表上下四分位数, 箱体中间横线表示中位数, 而圆圈则表示异常点. 可见, 所有被试均利用了所有尺度, 评价分值变化范围较大, 而且大多数被试的平均分值均大于理论均值4, 说明他们感觉大多数声音是非常不相似的.

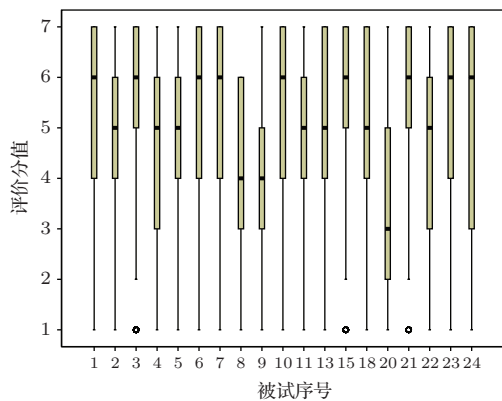


图1 保留被试的评价数据分布箱线图

3.2 CLASCAL模型

MDS是研究样本间不相似关系的有效工具. 从最初的欧氏距离模型开始^[17], MDS模型逐渐将关于样本和被试的各种因素考虑在内^[18,19], 因而得到了不断的精进, 其中CLASCAL模型是一个较为精细的MDS模型, 于1993年由Winsberg和De Soete提出^[20], 能更充分地挖掘实验数据中的信息. 该模型将不相似度数据建模为3部分: 公共维度、特异性和被试潜类. 公共维度由所有样本共有, 代表着被试在进行评价时利用的特征; 特异性为每个刺激所特有, 即其他刺激不具备的特征; 根据对公共维度和特异性分配的权重差异, 可将被试分为不同子群体, 称为潜类. 对于第 t 个潜类, 感知空间中的两个声音 i 和 j 间的距离可定义为

$$d_{ijt} = \sqrt{\sum_{r=1}^R w_{tr} (x_{ir} - x_{jr})^2 + v_t (S_i + S_j)}, \quad (1)$$

其中 d_{ijt} 为声音 i 和 j 间的距离, t 为每个潜类的索引, x_{ir} 为声音 i 沿第 r 维的坐标, w_{tr} 为潜类 t 在维度 r 上的权值, R 为总维度数, v_t 为潜类 t 对特异性的权重, 而 S_i 为声音 i 的特异性值.

利用EM (expectation maximization) 算法可获得这些参数的最大似然估计. 被试的类别结构是潜在的, 不存在关于特定被试属于哪一潜类的先验假设. 最优模型的选择包括公共维度数、特异性存在与否以及合适的潜类数. 其中前两者的确定取决于AIC (akaike information criterion) 和BIC (bayesian information criterion) 信息准则, 理论上模型最优时AIC和BIC同时达到最小值; 而最优潜类数的确定还要借助于Hope提出的Monte Carlo算法^[21].

表1 不同模型的AIC和BIC统计量

维度	无特异性		有特异性	
	AIC	BIC	AIC	BIC
2	14808	15087	11073	11488
3	12887	13295	10845	11390
4	11965	12504	10921	11596
5	11514	12182		
6	11228	12026		
7	11041	11969		
8	10890	11958		
9	10836	12024		

针对本文的数据, 首先根据Monte Carlo实验确定最优潜类数为3. 然后, 分别计算无特异性下2

至9维以及有特异性下2至4维的AIC和BIC统计量的值,如表1所示.在无特异性时,BIC在维度数取8时达到最小值,而AIC却没有;在有特异性时,AIC和BIC同时在维度数取3时达到最小值,且比无特异性下维度数为8的模型相应的值小.因此,最优模型为潜类数为3且包含特异性的3维模型.通过EM算法可获得最优模型对应的公共维度坐标、特异性和潜类权重的估计值,本文接下来将着重分析它们的特性及其对目标感知分类的影响.

3.3 公共维度

公共维度是构成样本间感知差异的基础,首先观察单个维度对水下声信号的分类情况.将样本按

坐标分别映射到 D_1 - D_2 和 D_2 - D_3 平面,如图2所示. D_1 将人为目标和自然目标大体分开,同时也将哺乳动物发声与其他声音明显地区分; D_2 和 D_3 则分别将水下背景声和巡逻艇声与其他声音明显分开.然后,求取所有声音在公共维度空间中的距离,并进行层次聚类,结果如图3所示.可以看出,水下声目标大体可分为水面大型舰船和潜艇、水面小型舰艇、水下背景声、小型鱼虾以及大型哺乳动物,这与2.1节所列举的类别相比,水面舰船和潜艇的分类有些许差异,而其他类别保持一致.本文第4节将利用听觉模型对声信号进行分析,辨识出能够描述每个维度感知意义的听觉特征以及其他有助于分类的重要特性.

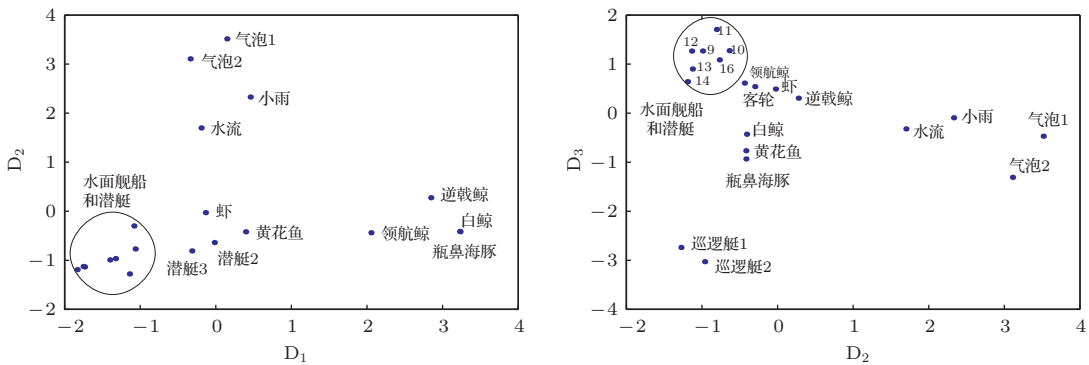


图2 3维感知空间的2维映射

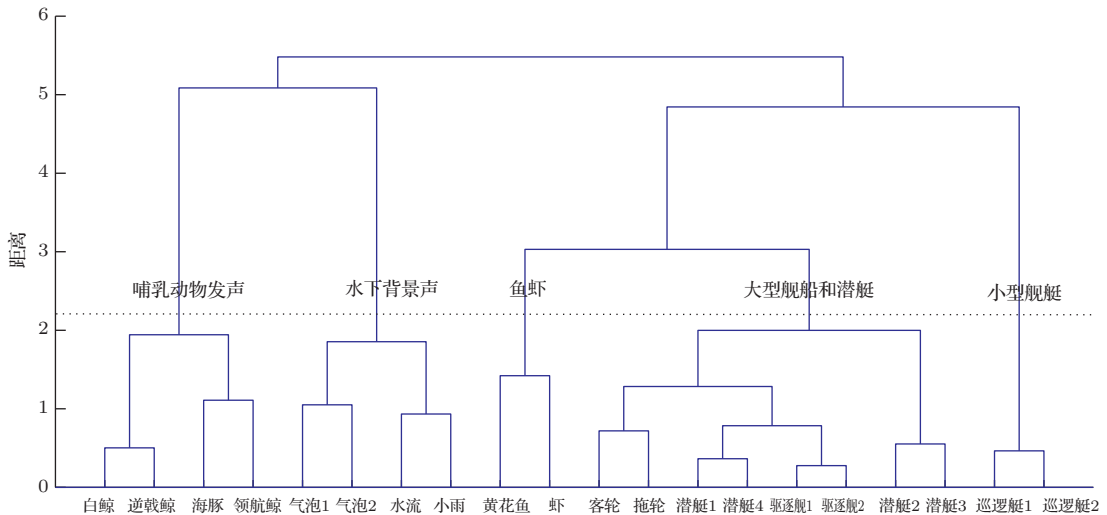


图3 样本聚类图

3.4 特异性和潜类分析

表2给出了所有声样本特异性值的平方根,取值越大,则感受到的特异性越强.可以看出,大多数

声音拥有与公共维度坐标大小可比拟的特异性值,这与先前使用CLASCAL模型的听觉感知研究所得结果差异较大,在这些研究中只存在有限几个样本具有较大特异性值.产生这一现象的原因是,先

前的研究所用的声音具有相似的声学特性(如乐器声)^[16]或声源结构(如汽车空调噪声^[23]),因而具有不同于其他声音的独特属性较少.然而,本实验所选样本类别差异很大,时域和频域特征也相差极大,加上被试对水下声信号的不熟悉,从而可能导致特异性值较大.此外,自然目标发声由于变化范围相对更大,因而具有比人为目标声更大的特异性值($t(18) = 2.203, p < 0.05$),这也在一定程度上印证了上述解释.由定义可知,特异性反映了样本的独特特性,它同时增加了目标的类间差异和类内差异,而且未知样本的特异性也是无法计算的,因此特异性的量化对于目标分类是无益的.CLASCAL模型将特异性考虑在内,从而排除了不相似关系中的不确定因素,提高了空间维度在目标分类时的有效性.

表2 样本特异性值的平方根

声样本	特异性	声样本	特异性
白鲸	2.90	客轮	2.86
领航鲸	3.06	潜艇1	2.06
逆戟鲸	3.04	潜艇2	2.26
瓶鼻海豚	3.93	潜艇3	2.36
黄花鱼	4.17	潜艇4	1.89
虾	3.56	驱逐舰1	1.90
气泡1	0.74	驱逐舰2	2.34
气泡2	1.85	拖轮	2.55
水流	3.28	巡逻艇1	2.19
小雨	2.72	巡逻艇2	0

图4显示了每个潜类对公共维度和特异性所赋予权重的差异.潜类1和3均包含8名被试,它们的权重模式较为相似,权值在各个维度相对均衡,而潜类3的权重比潜类1更大,说明他们所用的评价分值更大($t(14) = 2.489, p < 0.05$),也反映了潜类1的评价尺度相对于潜类3更加保守.潜类2包含3名被试,他们对D₁赋予的权重显著大于其他维度,而D₁有效区分了人为目标和自然目标,说明此类被试能够重视对于目标分类最重要的维度;此外,该类被试对特异性的重视程度较低,而且他们的不相似评价的均值接近于理论均值4,这表明了他们更能有效辨识出这些声音间的公共差异.对比3个潜类的被试构成发现,潜类2的3名被试比其他两类被试的平均年龄较大,参加此类实验的次数也

较多,能够比较熟悉地完成目标,可见被试的训练程度对于目标辨识的结果是有显著影响的.

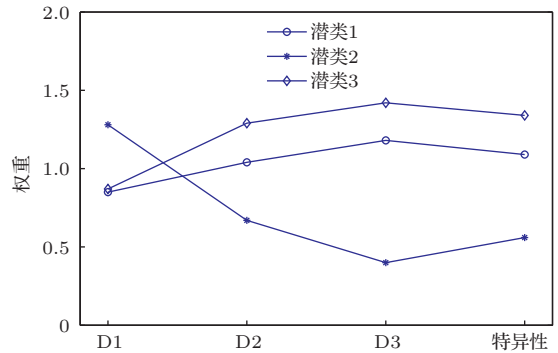


图4 3个潜类被试对公共维度和特异性的权重

4 听觉特征辨识与验证

接下来,本文将分析能区分不同类目标的感知特性,并尝试给出一种可定量描述它们的特征;然后,通过对新样本进行分类验证它们的有效性,并为其在实际中的应用提供指导.为分析这些特征,需基于合适的信号表达方式;由于这里的声目标分类是以听觉感知为基础的,所以可采用基于人耳听觉原理的Gammatone滤波器组作为分析前端^[23].考虑到水下声信号的特点,滤波器组包含100个子带,中心频率按等效矩形带宽(equivalent rectangular bandwidth, ERB)尺度均匀分在50 Hz和6400 Hz之间.声信号在通过该滤波器组后,会进行进一步变换,以获得最终的听觉特征.

4.1 维度1

D₁有效区分了人为目标和自然目标.首先,通过耳听比较,自然目标,包括哺乳动物、鱼虾发声以及一些水下背景声(如气泡)由于具有偶发的瞬态性,而人为目标发声较为规律平稳.为描述这一特性,将声样本Gammatone滤波器组每个子带的输出进行Hilbert变换和100 Hz低通滤波,获得听觉时频包络;图5显示了典型人为目标和自然目标声的听觉时频包络,可以看出,自然目标(如白鲸和虾)的声音包络一般在某些子带中变化较大,并有短暂的瞬态峰值,而人为目标(如潜艇和驱逐舰)声的听觉包络起伏较为平缓.

这里我们计算每个声音每个子带包络变化的归一化方差,取其中的最大值作为D₁的解释,其定

义式为

$$\max_{1 \leq h \leq 100} \frac{\frac{1}{N} \sum_{n=1}^N (e_h(n) - \mu_h)^2}{\mu_h^2}, \quad (2)$$

其中 h 为子带索引, N 为子带包络长度, $e_h(n)$ 为时刻 n 处的包络幅度, μ_h 为每个子带对应的包络均

值. 经计算, 样本的特征值与其 D_1 上的坐标值间的相关系数为 0.835. 图 6 显示了所有样本的特征值, 除小雨声外, 其他自然目标声的值均大于人为目标. 小雨声特征值较小的原因是, 该声音由无数个水滴无规律地撞击叠加产生, 其特性与白噪声相似, 从而其时频包络无较大起伏.

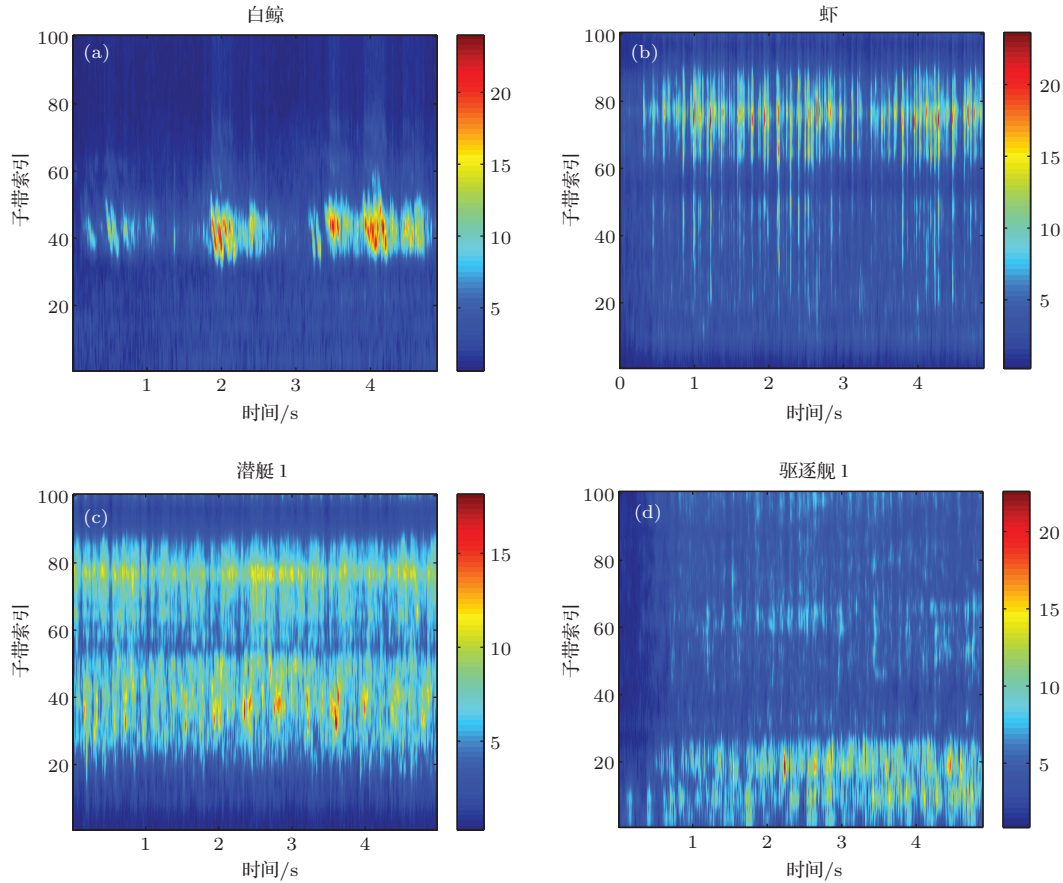


图 5 (网刊彩色) 典型人为目标和自然目标声的听觉时频包络

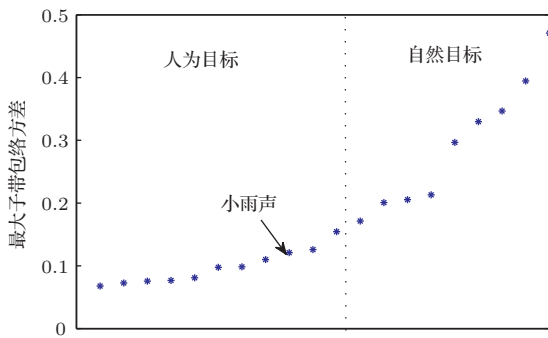


图 6 所有声样本的最大子带包络方差值

4.2 维度 2

D_2 有效分离了水下背景声与其他声音. 图 7

左右两图显示了气泡声和水流声的听觉时频包络, 它们与其他声音有两个截然不同的特点: 时频无规律性和较高的低频能量. 其中前者无法用单个特征表示, 而后者可用谱下降值描述.

谱下降值定义为累积能量达到声音总能量的 95% 时对应的频率 [24], 此处需将对应的索引数按照 ERB 尺度转换到普通的频率尺度. 该特征与 D_2 的相关系数为 -0.788 , 而直接通过功率谱求得的谱下降值与 D_2 的相关系数为 -0.720 , 从而说明了听觉模型更接近人类的听觉判断. 水下背景声的谱下降值显著低于其他声音, 除了小雨声; 而当去掉这个声音时, 该特征与 D_2 的相关系数则会变为 -0.837 . 因此, 该特征不能完全将水下背景声与其他声音完全分开, 需要其他特征作为辅助.

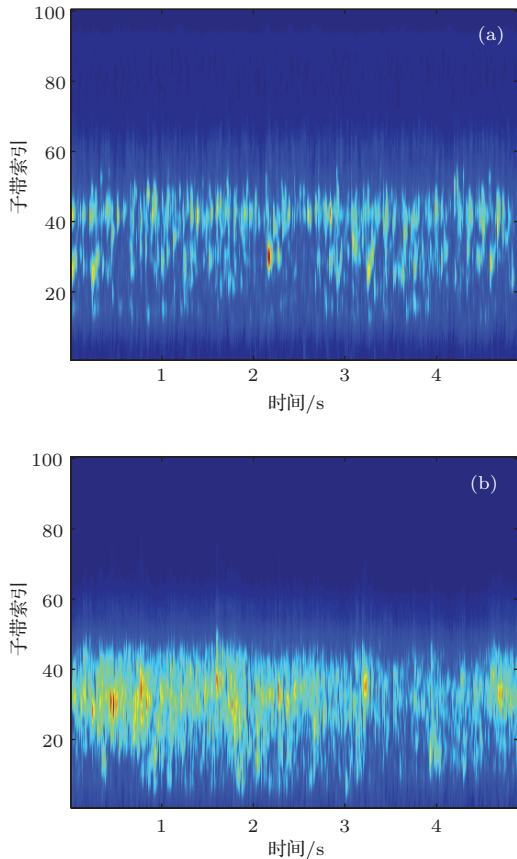


图7 (网刊彩色) 气泡和水流的听觉时频包络

4.3 维度3

D_3 将巡逻艇和其他声目标单独分开. 通过初步的耳听对比, 发现巡逻艇的声音具有频率较低的平稳音调. 因此, 描述该维度的特征与音调有关. 本文在提取信号音调时用到了相关图 (correl-ogram) 模型^[25]. 信号的相关图通过计算滤波器组每个子带输出的自相关函数产生. 信号的自相关是通过与自身的时间延迟版本相乘获得. 如果信号是周期的, 那么自相关函数会在时间延迟为重复率周期时显示出峰值. 自相关函数定义为

$$ac(\tau) = \sum_{t=0}^{\infty} x(t)x(t - \tau), \quad (3)$$

其中 $ac(\tau)$ 为信号 $x(t)$ 在延迟为 τ 的自相关函数. 为计算相关图, 需计算滤波器组每个子带输出的自相关函数, 从而获得一个时间和频率的二维函数

$$ac(h, \tau) = \sum_{t=0}^{\infty} x(h, t)x(h, t - \tau), \quad (4)$$

其中 h 表示子带索引. 沿频率加和相关图获得其总结:

$$sac(\tau) = \frac{1}{H} \sum_{h=1}^H ac(h, \tau), \quad (5)$$

其中 H 为滤波器组子带总数. 该函数的最大峰值延迟对应于信号音调的倒数, 而音调强度定义为该延迟时刻的值与 0 延迟时刻的值之比. 为获取音调随时间的变化情况, 需对信号分帧 (每秒 16 帧), 每帧对应一个音调值和音调强度值. 图 8 显示了巡逻艇声所对应的值随时间变化的情况.

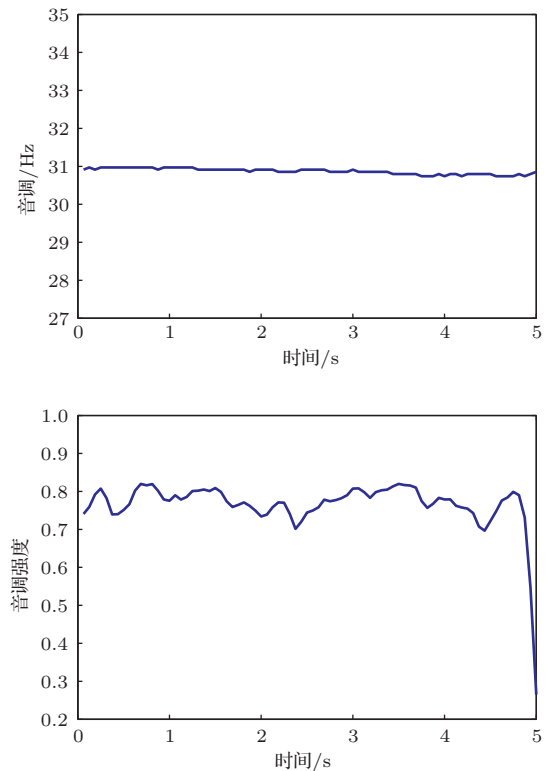


图8 巡逻艇声的音调和音调强度

考虑巡逻艇声音调较低的特点, 本文只计算每个声音 20—100 Hz 范围的音调. 取 0.3 为音调显著性阈值, 低于该值的时刻则认为没有音调. 取音调持续时间为 D_3 的测度, 相关系数为 -0.878 , 其中两个巡逻艇声的值显著大于其他声音.

4.4 节拍特性

虽然空间的 3 个公共维度具有明显的分类能力, 但上面提取的特征 (如谱下降值) 无法与其感知特性完全符合, 因而在目标分类时无法得到理想结果, 我们还可以发掘感知空间中的其他显著特征作为补充. 观察发现, 水面大型舰船和潜艇也在

空间聚为一类, 因而描述它们的共同特征对于目标分类也是非常重要的. 通过耳听对比, 发现此类声音均具有较强的节拍特性. 为描述这一特征, 本文利用到了 Foote 和 Uchihashi 提出的节拍谱 (beat spectrum) 模型 [26]. 该模型主要分为 3 步: 首先, 利用某种谱表达对信号进行参数化, 这样便获得一系列随时间变化的特征向量, 本文利用了前面提出的听觉时频包络; 然后, 计算任意两个时刻特征向量间的相关系数, 获得相似性矩阵 S ; 计算 S 主对角线和各副对角的加和平均, 便得到信号的节拍谱, 其中的峰值便对应信号自身的周期.

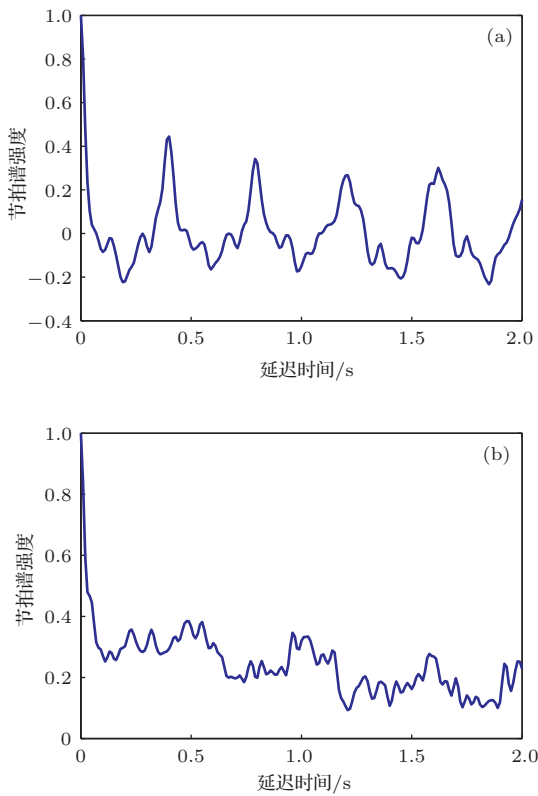


图 9 潜艇声和驱逐舰声的节拍谱

图 9 (a), (b) 分别显示了一潜艇声和一驱逐舰声的节拍谱, 它们均呈现出周期性的幅值. 为描述这一特性, 可对节拍谱进行 FFT 变换, 具有周期性节拍谱的声音会在特定频率上呈现峰值, 而其他声音经过变换得到的谱则较为嘈杂或平坦. 本文利用变换后的谱熵来描述, 其定义式为

$$SE = -\frac{\sum_{i=1}^N p(f_i) \log_2 [p(f_i)]}{\log_2 N}, \quad (6)$$

其中 N 为总频率点数, $p(f_i)$ 为频率 f_i 处能量占总能量的比例 [27]. 其中, 具有较强节拍性的声音谱熵较小, 而其他声音则较大.

4.5 特征验证

上面所提取的特征分别描述了不同目标类别所具有的特性, 然而这些特征是否对于新样本的分类具有适用性以及如何将它们用于分类, 仍是未知的. 接下来, 本文挑选了 20 个新的声音样本, 共分为 4 个类别: 水面大型舰船 (1—5)、巡逻艇 (6—10)、哺乳动物发声 (11—15) 和水下背景声 (16—20), 括号内为样本代号, 分别计算每个样本对应的 4 个听觉特征值. 这些样本同样来自文献 [12], 每个样本长度为 5 s.

由于上述特征的推导起源于空间模型, 因此首先基于层次聚类模型来观察这些特征能否对这些目标进行有效的聚类. 如图 10 所示, 聚类结果十分不理想, 并未出现预想的聚类. 究其原因我们发现, 这些特征满足二元分类特性, 即一类目标的特征值全部大于其他目标, 而在同类之间差异也很大, 此种情况下适于采用决策树分类模型.

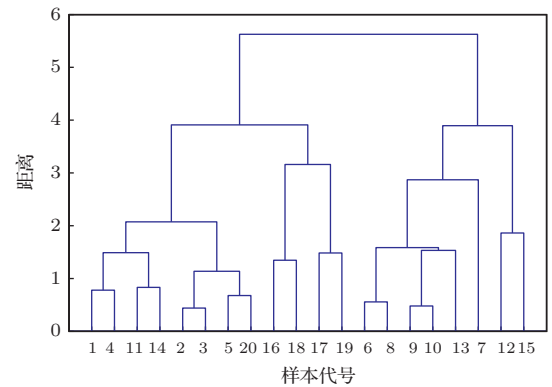


图 10 基于听觉特征的样本聚类图

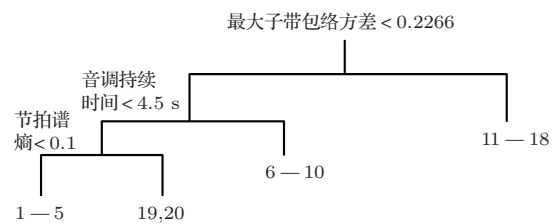


图 11 基于听觉特征的决策树

由于样本量较少, 因而不适宜采用机器学习, 这里需通过数据观察手动构造. 由上面分析可知, 最大子带包络方差能够有效区分人为目标和自然目标, 因而可作为决策树的第一个分支节点, 这样便把大多数自然目标声 (11—18 号样本) 分离, 而由于它们的具体类型是动物发声还是水下背景声对

于海战的重要性是很低的, 这里不再尝试区分它们, 因而最终的分类将与人的感知分类有所不同. 然后, 由于巡逻艇声具有持续的音调, 因而音调持续时间将它们(6—10号样本)分离. 最后, 由于大型舰船有明显节拍特性, 因而节拍谱熵也将它们清楚地分开(1—5号样本), 这样最后两个水下背景声(19, 20号样本)也得到正确的分类. 谱下降值在此处并未作为分支节点出现, 因而该特征不具备分类的有效性. 我们可以将类似的决策树模型应用到实际中, 而这里的分支节点阈值来自于较少的样本从而可靠性较差, 因此还需要通过构造大样本集进行统计学习来确定它们的取值.

5 讨论和总结

本文采用基于不相似度评价的多维尺度分析法研究了人类被试如何根据感知不相似度对水下声目标进行分类, 并利用听觉模型提取了有效分类特征. 首先, 挑选几类典型的水下声信号进行成对比较实验, 利用 CLASCAL 模型对实验所得的不相似度数据进行建模, 得到了具有 3 个潜类、包含特异性的 3 维空间模型.

然后, 本文分析了空间公共维度、特异性和潜类权重各自的特性以及它们在目标感知分类中所起的作用. 在 3 个公共维度中, D_1 有效区分了人为目标声和自然目标声, 而 D_2 和 D_3 则分别将水下背景声和巡逻艇声与其他声音明显分开; 基于层次聚类分析发现, 公共维度将水下目标大致分为 5 类: 哺乳动物、鱼虾、水下背景声、大型舰船和潜艇以及小型舰艇, 这说明人的感知分类与实际的目标分类大体一致, 而实际中水下目标的种类不止这些, 如果将更多的样本进行比较, 将会出现更多的类别及更多的公共维度. 实验中的大多数声样本均显示出较大的特异性值, 这源于声源类型或时频结构巨大的变化性以及听者对声音的不熟悉; 而特异性作为样本独特的性质对于感知分类起到了干扰的作用, 但 CLASCAL 模型能将这些因素进行有效排除, 从而证明了它相对于其他 MDS 模型的优越性. 听者在进行不相似度评价时也表现出一定的个体差异, 其中具有丰富实验经验的被试更能重视样本间的公共变化, 而且更重视能够区分人为目标和自然目标的 D_1 , 可见训练程度对于目标辨识的影响是很大的; 因此, 如果在以后的研究中能更关注这些性能优越的被试, 那么得到的结论将更具有价值.

接着, 本文通过 Gammatone 滤波器组对声样本进行分析, 辨识出能够进行目标分类的听觉特征. 通过对比观察, 3 个公共维度可分别由子带包络方差的最大值、谱下降值和 20—100 Hz 范围内的音调持续时间来有效描述; 此外, 大型舰船和潜艇的节拍特性可由节拍谱经 FFT 变换得到的谱熵描述. 最后, 利用这些特征对 4 类共 20 个新样本进行分类, 考虑到特征的二元分类特性, 构造了决策树的分类模型; 在进一步的研究中, 可通过构造大的样本集, 利用机器学习的方式建立更精确的分类模型, 从而能更接近于实际应用. 需要说明的是, 本文给出的特征并不是惟一的, 还存在其他能够描述这些感知特性的测度, 实际应用中可根据稳定性、计算量等因素选用合适的方法.

在本文的研究中, 水面大型舰船和潜艇被分为一类, 说明它们之间的差异相对于其他目标是很小的. 然而, 它们却是海战中真正所关心的目标, 因此后续的研究需要继续发掘它们之间更细致的分类以及能够进行分类的听觉特征.

参考文献

- [1] Li Z H, Chi H S 2006 *Acta Acustica* **31** 449 (in Chinese) [李朝晖, 迟惠生 2006 声学学报 **31** 449]
- [2] Mill R, Brown G 2005 *Tech. Rep.* CS-05-12
- [3] Li J, Chen Y M, Yuan Y, Chen Y 1998 *Acta Acustica* **23** 394 (in Chinese) [李靖, 陈耀明, 袁毅, 陈岳 1998 声学学报 **23** 394]
- [4] Lemaitre G, Susini P, Winsberg S, McAdams S 2007 *Acta Acust. United Ac.* **93** 457
- [5] Herrera-Boyer P, Peeters G 2003 *J. New. Music. Res.* **32** 1
- [6] Howard J H 1977 *J. Acoust. Soc. Am.* **62** 149
- [7] Tucker S 2003 *Ph. D. Dissertation* (England: University of Sheffield)
- [8] 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]
- [9] Wang N, Chen K A 2010 *Acta Phys. Sin.* **59** 2873 (in Chinese) [王娜, 陈克安 2010 物理学报 **59** 2873]
- [10] Yang L X, Chen K A, Wu Y *Acta Phys. Sin.* **62** 194302 (in Chinese) [杨立学, 陈克安, 伍莹 2013 物理学报 **62** 194302]
- [11] Zhu J, Zhang X H 2003 *Ship Science and Technology* **25** 55 (in Chinese) [朱进, 章新华 2003 舰船科学技术 **25** 55]
- [12] <http://www.hnsa.org/sound>
- [13] Ma Y F, Chen K A 2009 *Appl. Acoust.* **28** 59 (in Chinese) [马元峰, 陈克安 2009 应用声学 **28** 59]
- [14] Collier G L 2004 *Speech Commun.* **43** 297

- [15] Mackie R R, Wylie C D, Ridihalgh R R, Shultz T E, Seltzer M L 1981 *Some Dimensions of Auditory Sonar Signal Perception and Their Relationship to Target Classification ADA 102598*, California: Human Factors Research
- [16] McAdams S, Winsberg S, Donnadiou S, De Soete G, Krimphoff J 1995 *Psychol Res* **58** 177
- [17] Kruskal J B 1964 *Psychometrika* **29** 1
- [18] Carrol J D, Chang J 1970 *Psychometrika* **35** 283
- [19] Winsberg S, Carrol J D 1989 *Psychometrika* **54** 217
- [20] Winsberg S, De Soete G 1993 *Psychometrika* **58** 315
- [21] Hope A C 1968 *J. R. Stat. Soc.* **30** 582
- [22] Susini P, McAdams S, Winsberg S, Perry I, Vieillard S, Rodet X 2004 *Appl. Acoust.* **65** 763
- [23] Slaney M 1998 Technical Report #1998-010 (Interval Research Corporation)
- [24] Wang N 2010 *Ph. D. Dissertation* (Xi'an: Northwestern Polytechnical University) (in Chinese) [王娜 2010 博士学位论文 (西安: 西北工业大学)]
- [25] Martin K D *Ph. D. Dissertation* (America: Cornell University)
- [26] Foote J, Cooper M, Nam U 2002 *International Conference on Music Information Retrieval* Paris, October 13–17, **2002** 265
- [27] Elliot T M, Hamilton L S, Theunissen F E 2013 *J. Acoust. Soc. Am.* **133** 389

Underwater acoustic target classification and auditory feature identification based on dissimilarity evaluation*

Yang Li-Xue[†] Chen Ke-An Zhang Bing-Rui Liang Yong

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

(Received 6 January 2014; revised manuscript received 26 March 2014)

Abstract

The purpose of this study is to explore perceptual classification of underwater acoustic targets and auditory features used by human being. First, we design a paired comparison experiment. Then we use the CLASCAL algorithm to model the dissimilarity ratings as a perceptual space, and analyze the properties in three common dimensions, specialties, 3 subjects' latent classes and their roles in target perceptual classification. Finally, based on the gammatone filterbank, we find some auditory features that can effectively underlie 3 common dimensions and beat properties, so as to use them to construct a binary decision tree to classify some new samples; thus we can provide some guidance about how to use these features in practical applications.

Keywords: underwater acoustic target classification, dissimilarity, multidimensional scaling, auditory features

PACS: 43.66.+y, 43.30.+m

DOI: 10.7498/aps.63.134304

* Project supported by the National Natural Science Foundation of China (Grant No. 11074202).

† Corresponding author. E-mail: yanglixue.2008@163.com