基于听觉中枢模型的水下噪声音色表达与特性分析*

杨立学† 陈克安 伍莹

(西北工业大学环境工程系, 西安 710072)

(2013 年 4 月 25 日收到; 2013 年 6 月 19 日收到修改稿)

为建立水下噪声音色特征的定量表达以用于目标识别, 本文将主观评价实验获得的 4 个本质音色维度得分与声音的听觉中枢响应建立联系, 得到音色的偏最小二乘回归模型, 并基于回归系数对每个维度进行物理分析. 为验证该方法的有效性, 本文提取大量音色描述符作为自变量进行对比, 结果表明听觉中枢模型预测能力有一定优势. 同时发现, 前 3 个本质音色维度可分别由高频能量比例、谱平坦程度和时域连续性描述, 而第 4 维度则无法与任何声学特征建立联系.

关键词: 本质音色, 听觉中枢模型, 偏最小二乘回归, 音色描述符

PACS: 43.66.+y, 43.30.+m, 43.64.Bt **DOI:** 10.7498/aps.62.194302

1 引 言

水下目标自动识别在现代海战中至关重要. 为改善传统特征提取方法, 模仿人耳听觉原理提取听觉特征日益受到重视. 在人听音辨物过程中, 声音的听觉属性起着决定性的作用. 声音听觉属性包括响度、音调和音色三种, 其中音色作为一种多维感知属性, 是听觉判断目标差异的主要依据 [1]. 所以, 听觉特征的提取关键在于实现音色的定量表达.

传统的音色建模方法主要有两种. 一种是心理声学参数, 它给出了特定主观感受 (如粗糙度和波动强度) 与不同信号参数 (如调制频率和包络形状) 之间的关系 [2], 然而这种方法不能概括不同声音音色之间的全部差异, 只能从局部着手; 另一种基于音色空间的方法较为通用, 已在乐器纯音音色研究中得到应用. 该方法一般通过成对比较法 (paired comparison, PC) 和多维尺度分析 (multidimensional scaling, MDS) 将音色差异建模为低维空间内的距离, 然后基于信号分析方法提取一些信号参数特征 (如谱质心、上升时间和谱通量等 [3]) 来表达和解释每个维度. 由于上述两种方法获得的众多可计算参数各自反映了声音音色的特定部分, 因而有研究者将它们通俗地称为音色描述符 [4].

按音色的定义, 水下噪声也具有音色 [1], 其音色建模研究一般采用第二种方法, 但在描述物理意义时则常用语义法. 音色的语义描述法可分为自然音色 (natural timbre, NT) 和本质音色 (essential timbre, ET) 两类, 自然音色属性由一系列人们易于理解和表述的形容词组成, 如刺耳的、平滑的等; 本质音色是表述音色空间的正交维度, 可以由一系列自然音色组合而成, 但一般不能用简洁、便于表述的语言形式体现. 一般情况下, 可利用相关分析验证本质音色空间和 MDS 空间的等价关系, 进而利用本质音色的语义解释音色空间的每个维度. 基于此种方法, 2004 年 Geoffrey 指出, 起伏性和像动物声可以用来区分水下噪声源 [5]; 2009 年, 王娜等人将水下噪声音色空间的 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 [6]. 然而, 要将音色特征用于目标识别, 本质音色的定量表达是必要的. 目前, 国内外研究者在此方面的尝试较少, 原因是水下噪声参数复杂, 难以将特定维度描述为单一有效的计算特征. 2010 年, 王娜等人通过多元回归将前面描述的 5 个本质音色维度与大量音色描述符建立联系 [7], 并应用于水下噪声目标识别, 然而由于每个维度公式均包含多类具有不同物

* 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

理含义的描述符, 很难形成统一描述. 总之, 水下噪声音色建模需要寻找一种易实现的通用方法.

音色建模的首要问题是选择合适的信号分析形式. 由于人对音色的感知是以声音在听觉系统的处理为前端的, 因而可选择听觉计算模型作为分析工具. 典型的听觉模型一般包括听觉外周和听觉中枢两个阶段, 不同模型在听觉外周阶段均包含耳蜗处理机理, 而它们的差异在于进一步的听觉中枢处理, 因为该阶段受限于应用目的和听觉生理学的发展. 研究表明, 人脑感受音色的一个重要区域是听觉中枢的杏仁体 (amygdale), 虽然目前人们尚未建立对应的听觉表达模型, 但与之直接相关的初级听皮层 (AI) 模型取得了一定的进展, 典型代表是 Shamma 提出的多尺度谱时调制模型 [8]. 目前, 该模型已成功用于语音清晰度评估、语音与非语音的辨别、声音不愉悦度建模和车辆噪声识别等应用中 [9−12].

本文利用 Shamma 提出的模型建立水下噪声本质音色与声音在初级听皮层表达之间的关系. 具体来讲, 就是将其转化为一个多元回归问题, 其中听觉中枢模型输出的每个分量构成自变量, 样本的本质音色得分构成因变量, 而最终的回归系数特征将用于特定维度的物理解释. 为验证该方法的有效性, 本文提取大量音色描述符作为自变量, 利用相同的回归方法与听觉中枢模型进行预测能力的比较. 进一步, 通过相关分析获得与每个本质音色最相关的音色描述符, 结合其物理含义对基于听觉中枢模型的物理解释进行对比验证.

2 基于听觉中枢模型的音色建模

2010 年, 文献 [13] 利用 10 个常见的语义描述词对 50 个水下稳态噪声样本进行语义评价, 考虑到水下目标主要分为军用和民用两大类, 以及水下声环境的特点, 这些样本选自 3 个类型, 分别为水下环境声、潜艇声和民用船声. 上述声样本均来自远场接收点, 由于海洋环境的复杂性, 其声学特性与原始辐射声发生了很大程度的未知变化, 本文无法予以讨论. 样本类型及说明如表 1 所示.

实验得到的评价值通过主成分分析获得了 4 个本质音色维度 Y_1—Y_4, 其中维度 1 由震颤的、粗糙的、刺耳的、嘈杂的、沉闷的构成; 维度 2 由急促的和紧凑的构成; 维度 3 由变化的和不平缓的构成; 维度 4 则由重复的构成. 然而, 该文并未对得到的本质音色维度进行定量描述, 因而下面将基于听觉中枢模型解决此问题.

表 1 样本类型说明

声样本类型	声源描述	个数
水下环境声	水下气泡、水的流动、水下冰裂、水下雨声、水流及波浪等	16
潜艇声	潜艇不同方向形式、不同运行方式	17
民用船声	踏板船、独木舟、轮船及民船不同方向行驶和运行方式	17

2.1 听觉中枢模型

2003 年, Shamma 给出了声音在初级听皮层的听觉表达 [8], 其中初级听皮层位于大脑听觉中枢, 其响应与音色感知直接相关. 该模型包括两个阶段: 听觉外周和听觉中枢, 其中听觉外周输出是听觉中枢分析的输入. 下面给出两个阶段的听觉模型原理及实现步骤.

2.1.1 听觉外周

生理上, 该阶段首先由耳蜗基底膜将声压波转换为有序空间的膜振动, 然后再将其转换为听神经的电位活动, 最后在耳蜗核中提取声谱, 然后由音频结构通路经由中脑和丘脑投射到听皮层. 为使模型更接近实际的听觉生理特性, 本文对原模型做了一定程度的改进, 该阶段的输出表达可称为听觉谱图 (auditory spectrogram).

该阶段可分 3 个步骤实现: 1) 分析阶段: 由在频率对数轴上等间距分布的 128 个 Gammatone 带通滤波器构成, 其中心频率覆盖 6.4 个倍频程 (62.5—5 kHz); 2) 转换阶段: 利用 Meddis 提出的毛细胞模型模拟基底膜振动到听神经电位活动的转换 [14]; 3) 简化阶段: 一阶差分应用于整个通道阵列, 模拟耳蜗核神经元侧抑制作用, 紧接着是半波整流器和短时积分器.

2.1.2 听觉中枢

听觉外周阶段输出的听觉谱图经过听觉通路投射到初级听皮层, 进行进一步的听觉处理. 初级

听皮层神经元的最大特点是对特定谱时调制参数具有选择性. 所谓谱时调制, 就是时域调制和频域调制的组合, 其中时域调制反映包络沿时间轴的波动, 单位为 Hz; 频域调制则描述声谱沿频率轴的起伏, 单位为 cyc/oct. 动态纹波是一种以正弦形式被谱时调制的宽带噪声, 正 (负) 时域调制分别代表峰值向下 (上) 运动. 理论上, 任意一个听觉谱图都可以表示成一系列不同参数动态纹波的加和, 因而描述时域和频域调制快慢的度量也可称为纹波速率 (ω) 和纹波密度 (Ω).

由于初级听皮层是具有音频定位结构的, 因此它能建模为一组沿频率和时间轴抽取调制信息的滤波器组. 每个滤波器以中心频率 (F) 为中心, 并在一定纹波速率 (ω) 和纹波密度 (Ω) 范围内调谐. 由输入与每个滤波器的 STRF 直接卷积便可得到期望响应. 因此, AI 中每个滤波器的输出都是四维函数 $R(F,\omega,\Omega,t)$. 输出代表了初级听皮层神经元群对声音刺激的响应, 它与音色感知是直接相关的. 由于本文所研究样本多为水下稳态噪声, 因此听觉中枢四维响应 $R(F,\omega,\Omega,t)$ 的时间信息可通过求平均忽略, 从而得到 3 维表达 $R(F,\omega,\Omega)$.

2.2 声样本分析

图 1 至图 3 分别给出了 3 个典型声样本的听觉外周及听觉中枢表达, 其中 14 号样本为下雨声, 20 号样本为潜艇从左到右行驶的声音, 而 43 号为民船中速行驶的声音. 在每幅图中, (a) 图为对应样本的听觉谱图, 而为便于显示, 听觉中枢输出 $R(F,\omega,\Omega)$ 分别沿 Ω, ω 和 F 轴进行平均, 获得样本在 F-ω, F-Ω 和 Ω-ω 平面上的表达, 分别如 (b)—(d) 图所示, 其中红色代表兴奋的响应, 蓝色代表抑制的响应. 图 1 的下雨声具有类似于白噪声的特点, 因而较之于其他声音具有更宽的频带和时域调制范围. 图 2 的潜艇声能量主要集中在 250 Hz 以下的低频, 因而给人沉闷的主观感觉, 同时其时域调制也呈现较强的随机性; 图 3 的民船声低频处包含了 3 根明显的线谱, 且当纹波密度大于 2 cyc/oct 时可将它们分辨, 高频处为水流产生的宽带噪声; 此外, 它还在 16 Hz 附近显示出规律性的时域调制, 因而声音具有节拍感. 通过分析可以发现, 频率 F、纹波速率 ω 和纹波密度 Ω 均是影响声样本音色特性的重要参数, 因此听觉中枢输出 $R(F,\omega,\Omega)$ 可作为音色建模的基础.

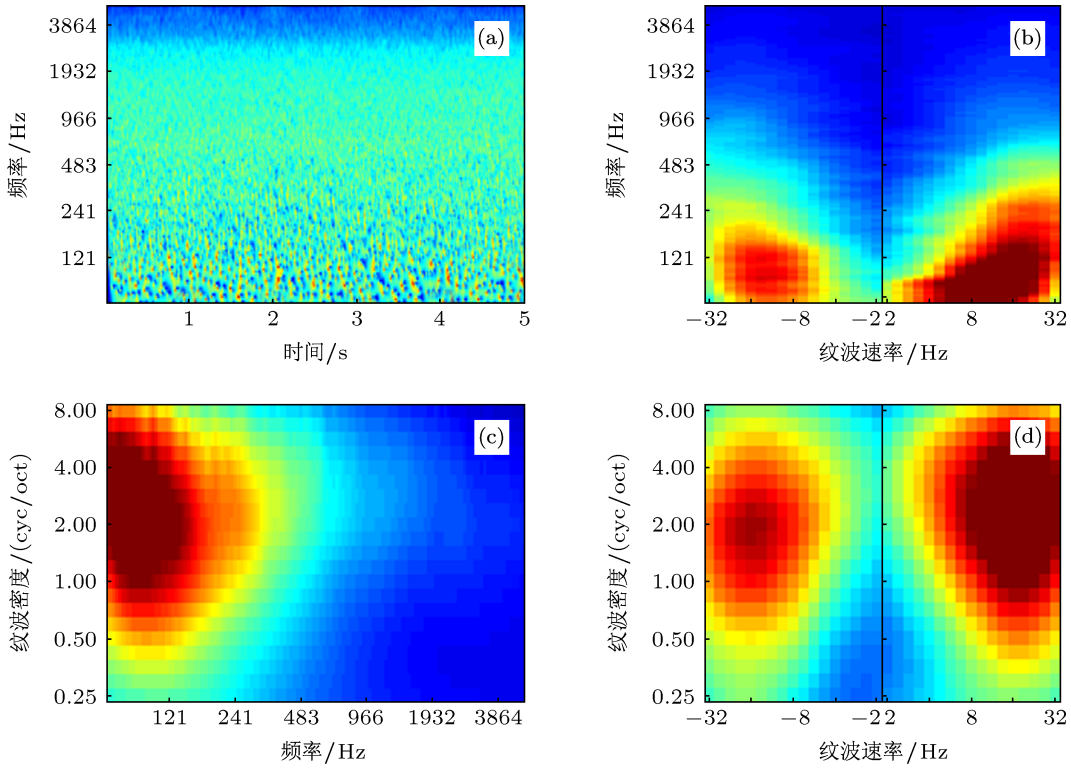

图 1 14 号样本的听觉表达

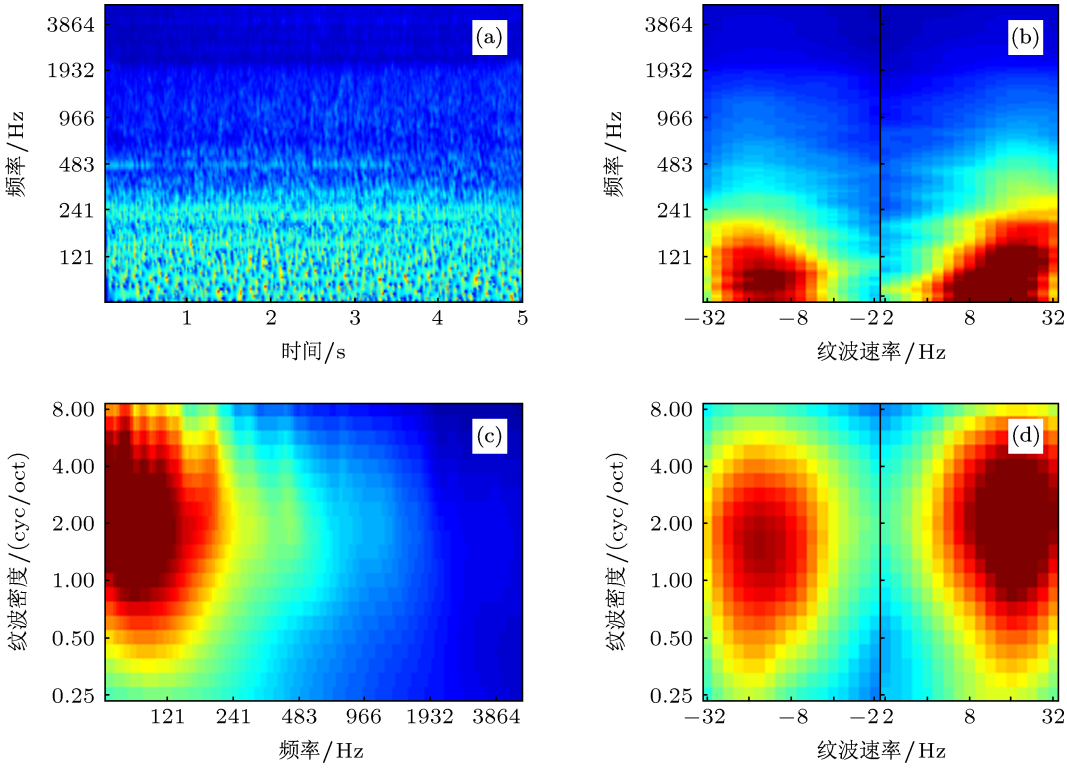

图 2 20 号样本的听觉表达

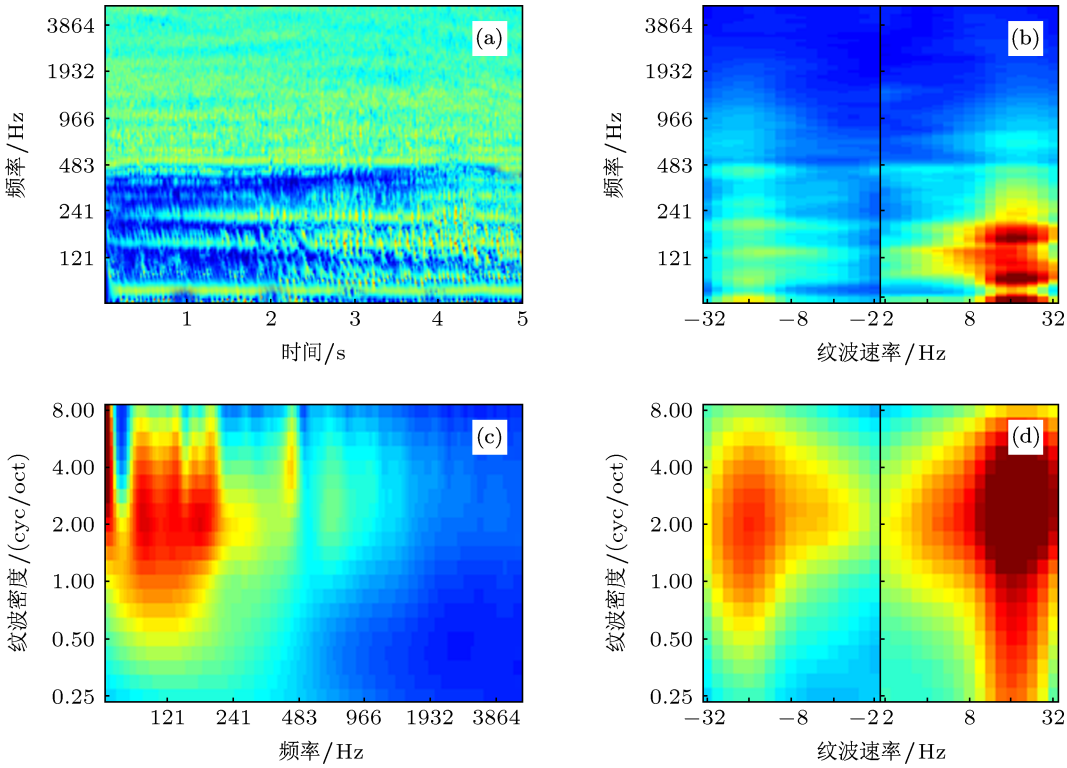

图 3 43 号样本的听觉表达

2.3 建模方法

将本质音色维度与听觉中枢响应建立联系可转化为一个多元回归问题, 对于第 i 个本质音色维度 Y_i, 有

$$Y_i = \sum_{a=1}^{N_F}\sum_{b=1}^{N_\omega}\sum_{c=1}^{N_\Omega}\beta_{abc}^{i}R(F_a,\omega_b,\Omega_c),\quad (i=1,2,3,4), \tag{1}$$

其中 $R(F_a,\omega_b,\Omega_c)$ 为听觉中枢模型输出的单个分量, N_F, N_ω 和 N_Ω 分别为 F, ω 和 Ω 的离散取值个数, β_{abc}^{i} 为每个输出分量的回归系数. 由于自变量维度 (7680) 远大于样本观测次数 (50), 且分量间存在较大相关性, 传统回归方法不能满足这一要求, 因此本文采用偏最小二乘 (partial least square, PLS) 回归作为建模工具. PLS 在建模过程中集中了主成分分析、典型相关分析和线性回归的思想, 其算法流程如下 [15]:

设 $\boldsymbol{X}$ 是一个 $(N\times K)$ 维的矩阵, 在 K 个输入/自变量上包含 N 次观测; $\boldsymbol{Y}$ 是一个（$N\times M$）的矩阵, 对应 M 个输出/因变量上的 N 次观测. PLS 方法试图在 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的行空间中寻找两个向量 $\boldsymbol{w}_1$ 和 $\boldsymbol{v}_1$, 使

$$\begin{aligned}\boldsymbol{t}_1 &= \boldsymbol{X}\boldsymbol{w}_1,\\ \boldsymbol{u}_1 &= \boldsymbol{Y}\boldsymbol{v}_1,\end{aligned} \tag{2}$$

拥有最大的协方差. 向量 $\boldsymbol{t}_1$ 和 $\boldsymbol{u}_1$ 称为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的潜变量, 可利用迭代算法抽取. 一般情况下, 一组潜变量不足以解释包含于 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 中的信息. 在减去第一潜变量的贡献后可利用相同方法提取第二潜变量, 依此类推直到达到模型所满足的精度.

此处采用单因变量的 PLS 方法对 4 个维度分别建模. 为防止过多潜变量所产生的过拟合问题, 本文采用交叉验证法: 每次从 50 个样本中随机抽取 10 个样本作为验证集, 然后利用剩余的样本拟合得到潜变量数从 1 到 10 变化时的模型, 最后利用这些模型对验证集进行预测, 并计算不同潜变量个数对应的可决系数 R^2. 重复运行 1000 次, 得到平均的可决系数, 其最大值对应于最优潜变量个数, 同时该值也将作为模型预测能力的测度.

2.4 回归结果分析

基于 PLS 法, 水下噪声样本听觉中枢响应对前三个本质音色维度进行了较好的预测 (图 6 中的白色柱状图), 平均可决系数分别为 0.59(4), 0.50(3) 和 0.48(2), 括号内为最优潜变量数. 然而, 第四维则无法与之建立显著的联系, 无论潜变量取多少. 注意到本质音色属性重复的描述了声源规律性的发声, 该属性可能与听觉中枢响应具有某种非线性的关系, 或者可能不是描述音色差异的关键因素.

图 4 自上而下分别显示了维度 1—3 的回归模型系数在听觉中枢响应上的分布. 图中自左而右则分别是系数在 F-ω, F-Ω 和 Ω-ω 平面的投影, 其中红色区域对应与每个维度呈较大正相关的分量, 而蓝色区域则对应呈较大负相关的分量. 可以看出:

1) 位于维度 1 正方向的声音倾向于拥有较高的高频 (> 2 kHz) 能量、较大的负纹波速率 (> 8 Hz) 和较小的正纹波速率 (< 8 Hz), 由于大多数样本的时域调制集中于 ±(8—32)Hz 范围内, 因而可以认为在维度 1 上取值较高的声音一般具有更多的负调制能量, 表明声音能量随时间逐步向高频移动.

2) 纹波密度是影响维度 2 排列的主要参数, 其中小于 2cyc/oct 的区域系数较大, 其对应声音的频域波动尺度较小, 即听觉谱较平坦; 而如果声音在大于 2cyc/oct 的区域系数较大, 则表明声音有明显的线谱分量, 而此类声音在该维度取值较低.

3) 维度 3 的系数在 250 Hz 以下及 1000 Hz 以上的系数较大, 表明中频能量比例与该维度取值成反比, 同时拥有较小 (< 8 Hz) 的声音在维度 3 上取值较小, 这些声音一般包括一些脉冲或瞬态分量, 因而使包络产生较慢的波动, 而该维度另一侧的声音只存在较小程度的无规律调制, 时域连续性较强.

总之, 维度 1 和维度 3 均与声音在频率-纹波速率上的联合分布有关, 但主要影响因素则分别是高频能量比例和瞬态声引起的包络波动. 维度 2 主要受纹波密度的影响, 与谱平坦程度或有无线谱有关. 然而, 这种方法的有效性和物理解释的合理性尚不可知, 需要与其他方法进行对比验证.

3 基于音色描述符的对比验证

先前研究获得的大量音色描述符, 分别反映了不同声音结构间的差异 (信号参数特征) 或描述了特定的主观感受 (心理声学参数), 它们已作为有效特征应用到各种声目标识别任务中 [16,17]. 因而, 可以预见众多音色描述符的组合同样可描述水下噪声的音色差异. 下面利用具有典型音色描述符集合对基于听觉中枢模型的分析结果进行对比验证.

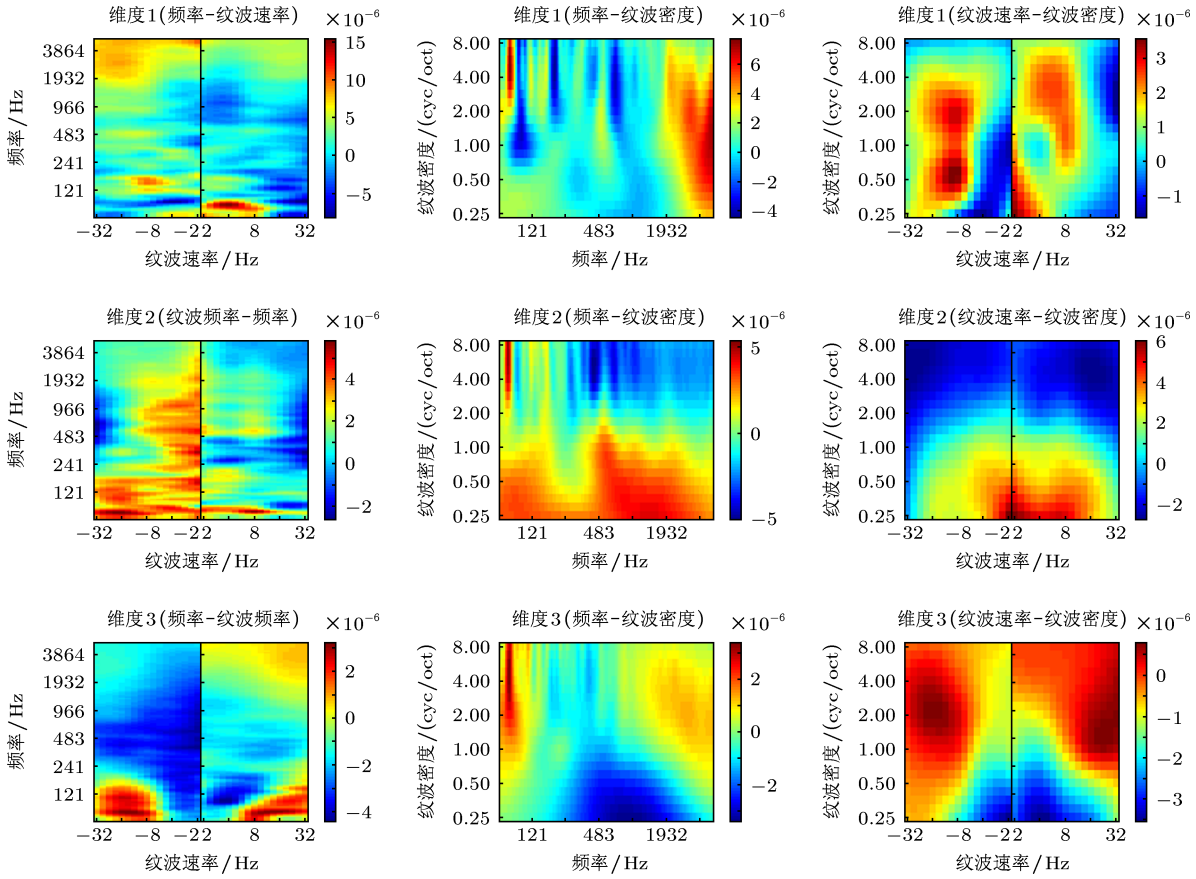

图 4 维度 1— 维度 3 的回归模型系数分布

3.1 音色描述符的提取

本文在音色描述符提取时利用了工具箱 MIRToolbox[18], 选用该工具箱的原因是其应用对象音频片段与水下噪声类似, 均具有长期平稳性. MIRToolbox 将常见描述符分为动态性、谱形状、旋律和清晰度四类 (图 5), 具有较强的全面性. 对于许多特征, 最终取值依赖于分析窗的长度. 本文经过实验决定使用长度为 23 ms 的窗, 且窗与窗之间有 50%的重叠. 每个特征进行逐帧平均后构成了预测变量, 用于建模水下噪声音色.

3.2 本质音色维度与音色描述符之间的关系

本节分析本质音色维度与音色描述符之间的关系, 主要包括两个方面

1) 基于 PLS 回归和上述交叉验证方法建立音色描述符对 4 个本质音色维度预测的回归模型, 对应平均可决系数分别为 0.48(3), 0.57(2), 0.35(3) 和 0.12(2) (图 6 中的黑色柱状图), 其中括号内为对应的潜变量数. 可见, 音色描述符对前 3 个维度很好地进行了预测, 但对第 4 维的描述同样较差.

2) 为描述单个音色描述符的相对贡献, 可利用其与本质音色维度的相关系数表征. 表 2 给出了与 4 个维度最相关的 5 个音色描述符.

在维度 1 较高一侧的声音倾向于拥有更多的高频能量 (谱下降值、谱质心、明亮度) 和较宽带的谱能量分布 (谱熵). 上升斜率大小与声音的瞬态性有关, 较大的上升斜率对应较强的瞬态性, 从而使声音拥有宽带的频域分布, 该特征描述了与谱熵类似的特性 ($r = 0.9041$, $p < 0.001$). 由于水下噪声能量多分布在低频, 因而较宽带的谱分布同样对应较大的高频能量比例.

维度 2 与粗糙度呈显著的负相关关系. 粗糙度, 也称感官不调和度 (sensory dissonance) 描述了由较窄的泛音间隔引起的嗡嗡的听觉感受. 根据 Vassilakis 提出的谱分析模型 [18], 声音的总粗糙度可计算为频谱中所有泛音对的粗糙度之和, 而每个泛音对的粗糙度正比于泛音幅度, 并随泛音间隔先增大后减小. 对于水下噪声, 由于低频线谱幅度明

显高于其他频率分量, 从而主导了粗糙度的大小, 因此无线谱噪声一般粗糙度较小, 但在维度 2 上的取值却较高. 由此可见, 维度 2 与声音谱平坦程度呈正比.

上升时间反映了包络的平缓程度, 取值较大的声音一般具有较慢的时域包络波动, 而波动强度则以 4 Hz 为中心频率呈现出与时域包络调制速率的带通特性 [2]. 由于维度 3 与这两个特征之间均呈一定的负相关关系, 因此位于维度 3 负方向的声音一般拥有较慢的时域调制, 即包含更多的脉冲或瞬态声. 此外, 该维度与 mfcc5—mfcc7 呈一定的正相关关系, 它们描述了声音的一些谱形状特征, 但其确切物理含义无法确定.

维度 4 显示出与传统特征的相关性较弱, 与其最相关的谱变化仍未达到较高的显著程度. 通过观察语义评价数据, 可发现评价值在不同样本间的可分性较小, 因而可以推断重复的可能不是描述这些样本音色差异的显著因素. 所以, 接下来的分析将不再对其进行针对性的讨论.

3.3 讨论

根据以上分析, 听觉中枢模型和音色描述符均有效地建模了前 3 个本质音色维度, 其中听觉中枢模型预测能力稍优 (图 6), 而两者均无法对维度 4 进行有效的描述. 虽然两者在结果上有所差异, 但在物理分析上存在一定的一致性.

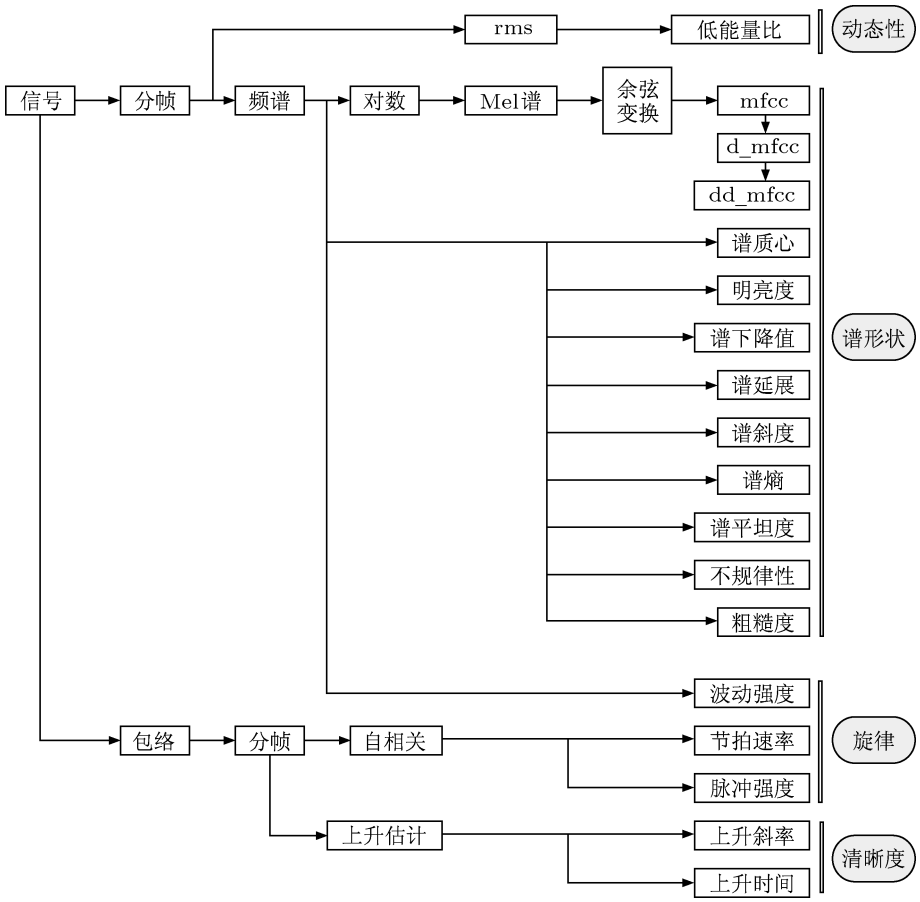

图 5 音色描述符提取框架

表 2 特定音色描述符与感知维度的关系

维度 1	维度 2	维度 3	维度 4
上升斜率 0.6762***	粗糙度 −0.7986***	上升时间 −0.5449***	谱变化 −0.4286**
谱下降值 0.6671***	d_mfcc9 −0.6340***	波动强度 −0.5426***	谱不规律性 −0.3630**
谱质心 0.6671***	零交变率 −0.5916***	mfcc6 0.5353***	mfcc12 −0.3492*
谱熵 0.5899***	谱熵 −0.5618***	mfcc5 0.4683***	mfcc11 −0.3465*
明亮度 0.5800***	d_mfcc10 −0.5210***	mfcc7 0.5353**	d_mfcc4 03294*

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

维度 1 与声音高频能量所占比例呈正相关. 通过观察发现, 维度 1 正方向的声音一般为水流声、下雨声和气泡声等宽带噪声, 而位于另一端的声音一般为低频的潜艇声或深水隆隆声, 从而印证了这一点. 此外, 听觉中枢模型结果表明负纹波频率的能量也对维度 1 有正的影响, 表明频率随时间的增加, 产生这种现象的原因可能是舰船的加速运行或靠近, 而音色描述符则没有对此现象进行描述.

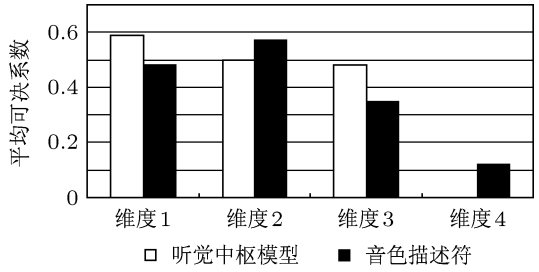

图 6 两种方法的预测能力对比

维度 2 与谱平坦程度或有无线谱有关. 位于维度 2 较低一侧的为民船声, 它们在低频拥有紧密排列的线谱, 从而在较大纹波密度处的能量较高, 且粗糙度较大; 而位于另一侧的声音为类似于水流声的平坦噪声, 它们无明显的频域起伏或峰值间隔, 从而在较小纹波密度处的能量较高, 且粗糙度较小. 注意到粗糙度在该声音集合中的取值主要受线谱有无的影响, 其描述的物理意义与其原本所描述的粗糙的主观感受是无关的.

维度 3 与声音包络的时域调制有关, 还可解释为时域连续性. 位于维度 3 较低一侧的声音为破冰声和沉船发出的吱吱声, 它们包含多个具有一定时间间隔的瞬态声, 从而存在较大程度的低频调制; 位于另一端的声音为水流声和潜艇声, 它们较为连续且嘈杂, 时域调制范围较宽. 除此之外, 基于听觉中枢模型的分析显示声音在中间频率范围 (250—1000 Hz) 的能量比例与维度 3 呈反比, 音色描述符 mfcc5—mfcc7 虽描述了一定的频域特征, 但两者并未有显著的相关关系.

总的来说, 两者在解释本质音色维度时具有较好的一致性. 然而, 听觉中枢模型较之音色描述符具有几个方面的优势: 1) 听觉中枢模型通过融入多尺度的谱时调制分析, 对声音进行了更为完备的描述, 而一般的音色描述符在提取过程中会进行时域或频域的平均, 从而丢失了部分信息; 2) 听觉中枢模型形象具体, 物理意义明确, 而音色描述符虽描述了一定的音色特征, 但在样本集变化时决定其确切的物理意义也是变化的, 如粗糙度; 3) 听觉中枢模型融入的一些生理学特性对于音色感知是关键的, 例如毛细胞的压缩非线性, 通过实验可发现, 如果在去掉该特性, 则听觉中枢模型的预测能力将大幅度下降, 不再具有优势 (具体结果文中没有给出); 4) 听觉中枢模型具有可逆性, 即可从音色表达出发合成具有特定音色的声音, 而大多数音色描述符则不具备这一性质.

4 结 论

为模仿听觉功能提取音色特征用于目标识别, 本文通过主观评价实验获得的 4 个本质音色维度为因变量, 同时计算声音的听觉中枢模型输出分量作为自变量, 利用 PLS 回归和交叉验证方法获得最优模型, 并基于回归系数的分布对每个维度进行了物理解释; 为验证该方法的有效性, 本文还提取大量先前音色感知研究总结的音色描述符作为音色的预测变量, 从而构成对比, 结果表明听觉中枢模型稍优; 最后, 通过相关分析获取与每个维度最相关的音色描述符, 结合其物理意义与听觉中枢模型对比, 发现位于前三个维度正方向的声音分别具有较高的高频能量、较大的谱平坦程度和较强的时域连续性, 而第四维却无法与任何特征建立联系.

本文对音色空间的描述与其他文献的研究具有一定的可比性: 维度 1 和维度 3 分别与频域能量分布和时域包络特性有关, 它们在很多研究中都是描述音色差异的重要方面. 例如, 一些文献提出了谱质心来描述高低频能量比 [3], 而时域连续性也是文献 [19] 中环境声音色空间的一个重要维度. 本文对维度 2 的描述似乎是水下噪声特有的, 它区分了具有线谱的舰船噪声和其他宽带噪声, 文献 [20] 中的音调性 (tonality) 描述了相似的特性.

水下噪声音色的研究具有重要的实用价值, 本文所提出的基于听觉中枢模型音色建模方法的结果令人鼓舞, 因而值得在未来研究中借鉴使用. 然而, 本文的研究具有一定的局限性: 1) 研究范围仅限定于 3 类典型的目标 (水下环境声、潜艇声和民船声) 辐射的稳态噪声, 无法适用于全部的水下噪声种类; 2) 实验样本均为相同条件下的记录声, 未能考虑声传播距离、水深等因素的影响. 因此, 未来的研究将通过改进实验方法扩大样本的种类及数量, 使模型通用性更强, 并试图改变声记录条件, 探索音色随声传播特性及环境因素的变化规律.

[1] 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]
[2] Zwicker H E, Fastl H 1999 *Psychoacoustics: Facts and Models* (Berlin Heidelberg: Springer-Verlag Press)
[3] Donnadieu S 2007 *Analysis, Synthesis, and Perception of Musical Sounds* (Berlin Heidelberg: Springer-Verlag Press) p272–319
[4] Peeters G, Giordano B L, Susini P, Misdariis N, MaAdams S 2011 *J. Acoust. Soc. Am.* **130** 2902
[5] Collier G L 2004 *Speech Commun.* **43** 297
[6] 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]
[7] 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)]
[8] Shamma S 2003 *IETE J. Res.* **49** 1
[9] Chi T, Gao Y H, Guyton M C, Ru P, Shamma S 1999 *J. Acoust. Soc. Am.* **106** 2719
[10] Mesgarani N, Shihab S, Slaney M 2004 *IEEE International Conference on Acoustics, Speech, and Signal Processing* Montreal, Quebec, Canada, May 17–21, 2004 I–601
[11] Kumar S, Foster H M, Bailey P, Griffiths T D 2008 *J. Acoust. Soc. Am.* **124** 3810
[12] Chen K A, Wu Y, Yang L X 2011 *Appl Acoust* **30** 407 (in Chinese) [陈克安, 伍莹, 杨立学 2011 应用声学 **30** 407]
[13] Chen K A, Wang N, Wu Y, Ma M, Zhang B R 2010 *Chinese Sci. Bull.* **55** 651 (in Chinese) [陈克安, 王娜, 伍莹, 马苗, 张冰瑞 2010 科学通报 **55** 651]
[14] Meddis R, Hewitt M J, Shackleton T M 1990 *J. Acoust. Soc. Am.* **87** 1813
[15] Wang H W 1999 *Partial Least-Squares Regression: Method and Applications* (Beijing: National Defense Industry Press) (in Chinese) [王慧文 1999 偏最小二乘回归方法及其应用 (北京: 国防工业出版社)]
[16] Akitoshi I, Hiroshi Y 2006 *IEEE Asia Pacific Conference on Circuits and Systems* Singapore, December 4–7, 2006 992
[17] Fabian M, Ultsch A, Thies M, Lohken I 2006 *IEEE Trans. Audio and Speech Processing* **14** 81
[18] Vassilakis P N 2007 *Proceedings SMC'07, 4th Sound and Music Computing Conference* Lefkada, Greece, July 11–13, 2007 319
[19] Gygi B, Kidd G R, Watson C S 2007 *Percept Psycho* **69** 839
[20] 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* ADA102598, California: Human Factors Research

Timbre representation and property analysis of underwater noise based on a central auditory model*

Yang Li-Xue† Chen Ke-An Wu Ying

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

(Received 25 April 2013; revised manuscript received 19 June 2013)

Abstract

In order to establish quantitative timbre representation of underwater noise, this paper tries to build a relationship between essential timbre scores and central auditory responses to stimulus based on partial least squares regression, and use regression coefficients to interpret the physical meaning of each dimension. In order to verify the utility of this method, this paper extracts a large amount of timbre descriptors as independent variables for comparisons, and it is shown that the predictive ability of the central auditory model is better. Finally it is found from two types of timbre representations that the first three dimensions in the timbre space can be respectively interpreted as high-frequency energy ratio, spectral flatness and temporal continuity; however, dimension 4 cannot be related to any acoustical features.

Keywords: essential timbre, central auditory model, partial least squares regression, timbre descriptor

PACS: 43.66.+y, 43.30.+m, 43.64.Bt **DOI:** 10.7498/aps.62.194302

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

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