

扬声器质量判别中音色特征的选择及实验研究

殷贞强¹, 尹雪飞¹, 陈克安²

(1. 西北工业大学电子信息学院, 西安 710129; 2. 西北工业大学航海学院, 西安 710129)

摘要: 传统的扬声器出厂质量检测方法有两种: 试听员听音检测和扬声器频域特性检测。听音检测方法没有量化标准, 不同试听员测试结果有所差异, 而且对试听员的听力有一定损害。频域特性检测方法只能检测该扬声器的物理特性, 而对音色的检测作用有限, 常常遇到频域曲线一致, 而音色相差甚远的情况。充分考虑了音色对扬声器品质的影响, 提取了能够反映声音音色属性的特征, 利用自动分类方法对扬声器质量进行判别, 同时比较了不同特征之间的运算量。所提方法充分考虑了人耳听觉特性, 并有效地提高了识别率。

关键词: 扬声器; 音色; 自动判别; 分类

中图分类号: TN713

文献标识码: A

文章编号: 1000-3630(2012)-05-0506-05

DOI 编码: 10.3969/j.issn1000-3630.2012.05.012

Automatic recognition and experimental study of loudspeaker's quality based on characteristics of timbre

YIN Zhen-qiang¹, YIN Xue-fei¹, CHEN Ke-an²

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China;

2. School of Marine Technology, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract: In the quality inspection process of speakers, the traditional methods involve listening inspection and frequency domain detection. However, listening inspection does not have quantitative criteria, which leads to different results by different inspectors. In addition, it is harmful to the workers' hearing. Detecting the characteristics in frequency domain can only detect some physical characteristics and rarely consider the role of timbre. Sometimes two loudspeakers have the same frequency curve with different timbre. In this paper we fully account for the effect of timber on the loudspeakers' quality, extract the characteristics that reflect the representation of timbre, and use automatic classification method to determine the quality of loudspeaker, meanwhile comparing the amount of computation for these characteristics. This method fully takes into account the characteristics of human ear that suit more to people's listening habits, and meantime increases the recognition rate.

Key words: loudspeaker; timber; automatic recognition; classification

0 引言

扬声器在日常生活中越来越普及, 人们对扬声器声品质的要求也越来越高。在扬声器出厂质量检测中, 传统的方法有两种: 一种是人工试听, 另一种是观察扬声器的频域特性。这两种方法都有不足, 扬声器的质量问题有多种, 最常见的是无声、有杂音、音色差等。传统方法对无声和有杂音扬声器的识别率较高, 其中人工试听的方法对具有音色问题的扬声器有一定的识别率, 但是没有一个是量化的标准, 不同的试听员所检测的结果也存在差异;

检测频域特性的方法主要是观察扬声器的幅频特性和相频特性, 该方法描述了扬声器在频率轴上, 随着频率的变化其响应幅值和相位的变化情况, 但是却不能很好地反映音色的差异, 经常出现频域特性一致而音色相差甚远的情况。因此, 这两种方法均不能很好地识别有音色问题的扬声器。

音色是人在听觉上区别有同样响度和音调的两个声音所不同的属性^[1], 音色、响度和音调被公认为人感知声音的“三要素”, 音色感知是听觉感知的主要内容, 音色属性在目标辨识中发挥了极其重要的作用, 是辨识目标的主要依据。音色感知经历了长期的研究过程, 人们认为: 一个多维空间的特征组合可以描述音色属性, 声音的众多物理特性与音色属性密切相关。音色属性的研究已经不仅仅局限于音乐声, Lavandier^[2]对扩音器进行主观评价找到了辨别扩音器差异的音色属性及其相应的物理参数。由于音色属性定量计算存在困难, 使得目前

收稿日期: 2011-09-01; 修回日期: 2011-11-22

基金项目: 国家自然科学基金资助项目(11074202)、西北工业大学研究生创业种子基金资助项目(Z2011100)

作者简介: 殷贞强(1984—), 男, 河南信阳人, 硕士研究生, 研究方向为现代数字信号处理。

通信作者: 殷贞强, E-mail: yinzhengqiang@126.com

无法提取精确的音色特征。本文提取了与音色相关的信号特征, 利用分类的方法对扬声器的质量进行自动判别, 从录制的声音文件中提取出完整周期的扫频信号, 它包含了 20~5000Hz 频段的全部信息。经观察比较扬声器的频谱图, 发现有质量问题和无质量问题的扬声器在低频段有所差异, 因此主要提取 20~1000Hz 频段的信号特征, 同时也降低了计算量。采用 BP 神经网络分类器, 通过实验的方式比较了各个特征的识别率, 并对识别结果进行分析。该方法充分考虑了音色对扬声器品质的影响, 为扬声器的音色检测提供了量化标准, 同时提高了识别率, 检测结果也更符合人耳的听觉习惯。

1 与音色相关的信号特征

本文提出的扬声器质量自动判别过程包括预处理、特征提取、特征选择和分类四部分, 其流程图如图 1 所示。

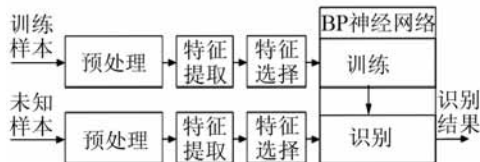


图 1 识别过程流程图
Fig.1 Flow chart of recognition

在图 1 给出的流程中, 特征提取是其关键。研究表明: 音色属性是一种多维感知属性, 声音信号的时域结构和频域结构都对其构成影响, 多个信号特征都与音色感知密切相关。本文提取了反映声音音色属性的特征, 包括时域特征和频域特征^[3-7]。

1.1 时域特征

1.1.1 时域质心

时域质心(Temporal Centroid, TC)是整个能量包络上的时间平均, 或者是给定时间信号的能量加权平均, 它是描述高频瞬态敲击声信号的重要物理量, 被广泛用于描述音色属性。计算公式如下:

$$TC = \frac{\sum_{n=1}^N nx(n)}{\sum_{n=1}^N x(n)} \quad (1)$$

其中 $x(n)$ 为时域信号, N 为信号长度。

1.1.2 零交点比率

零交点比率(Zero-Crossing Rate, ZCR)定义为通过加窗后的声音信号在时域上的过零点值。当连续的样本符号不同时就会产生零交点, ZCR 就是在短时窗口内信号改变符号的平均次数。计算公式为

$$Z_n = \frac{1}{2} \sum_m |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m) \quad (2)$$

$$\text{其中, } \text{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad (3)$$

1.2 频域特征

1.2.1 谱质心

谱质心(Spectral Centroid, SC)是声音信号的频率分布和能量分布的重要信息^[8], 它描述了声音的明亮度, 谱质心越高, 声音越明亮。计算公式为

$$SC = \frac{\int_0^{f_{\max}} f E(f) df}{\int_0^{f_{\max}} E(f) df} = \frac{\sum_{n=1}^N f(n) E(n)}{\sum_{n=1}^N E(n)} = \sum_{n=1}^N f(n) \frac{E(n)}{\sum_{n=1}^N E(n)} = \sum_{n=1}^N f(n) P(E(n)) \quad (4)$$

其中, f 为信号频率, $E(f)$ 为频域信号。 $f(n)$ 为 $x(n)$ 短时傅里叶变换后对应的频率, $E(n)$ 为离散时域信号 $x(n)$ 短时傅里叶变换后所对应频率的谱能量; N 为 DFT 的长度, $P(E(n))$ 为各点能量在总能量上的概率值。

1.2.2 谱质心带宽

谱质心带宽(Spectral Centroid Bandwidth, SCBW)是指能量集中的频带宽度, 相对谱质心来说, 它反映了声音能量集中的区域。计算公式如下:

$$SCBW = \frac{\sum_{n=1}^N |SC - f(n)| E(n)}{\sum_{n=1}^N E(n)} \quad (5)$$

1.2.3 谱通量

谱通量(Spectral Flux, SF), 它用于度量两个连续结构之间谱幅度的差异, 反映信号中各频带成分的能量之和, 即全频带能量, 与频率无关。计算公式如下:

$$SF = \int_0^{f_{\max}} E(f) df = \sum_{n=1}^N \Delta f(n) E(n) \quad (6)$$

其中 $\Delta f(n)$ 为 $x(n)$ 短时傅里叶变换后两点的频率差。

1.2.4 谱下降值

谱下降值(Spectral Roff-off, SRO), 它反映了声音能量开始下降的特定频率点, 一般该值都大于主谐波的基频值, 同样反映信号的主谐波的基频特性和主次谐波能量的分布情况。SRO 被定义为功率谱的累积幅度在 $C\%$ 以下的频率值, 经验公式如下:

$$SRO = R, \quad \sum_{n=1}^R E(n) = C \sum_{n=1}^N E(n) \quad (7)$$

其中 C 是经验系数, 一般取值为 0.85 或者 0.6, 本文取 0.85。

1.2.5 谱不规律性

谱不规律性(Spectral Irregularity, SI)它描述了普包络的形状,大幅度差值产生凹口包络,而较小差值则产生较平滑包络。计算公式如下:

$$SI = \frac{\sum_{n=1}^N (E(n) - E(n-1))^2}{\sum_{n=1}^N E(n)^2} \quad (8)$$

SI 的值越小反映声音的频谱包络越平坦,反之声音的噪声特性越明显,越不具有规律。

1.2.6 谱平整度

谱平整度(Spectral Flatness Measurement, SFM)定义为谱的几何平均与其算术平均的比值,单位为 dB,它确定了音调量:即相对于像噪声的声音,有多少声音像纯音。计算公式如下:

$$SFM = 10 \log_{10} \left(\frac{GM}{AM} \right) = 10 \log_{10} \left(\frac{\left(\prod_{n=1}^N |E(n)| \right)^{\frac{1}{N}}}{\frac{1}{N} \sum_{n=1}^N |E(n)|} \right) \quad (9)$$

1.2.7 美尔倒谱系数

Mel 倒谱是由 Davies 和 Mermelstein 提出来的,该特征参数利用了听觉原理和倒谱的解相关特性。另外, Mel 倒谱也具有对卷积性信道失真进行补偿的能力。由于这些原因, Mel 倒谱被证明是指在声音相关的识别任务中应用最成功的特征之一^[9]。美尔倒谱系数(Mel-Frequency Cepstrum Coefficients, MFCC)是很重要的音色特征,是音色感知模型中的有效模型,多用于实现音色识别和音色感知。

提取 MFCC 的计算流程如下:

(a) 预加重 $x_n = x_n - kx_{n-1}$, 其中 $k=0.95$;

(b) 加 hamming 窗, 窗长为 N ;

(c) 对时域信号进行 DFT 变换^[9];

(d) 设计一个具有 M 个带通三角滤波器的滤波器组, 频率响应如式(10)所示, 中心频率 $f(m)$ 从 $0 \sim F_s/2$ 间按 Mel 频率分布。

$$H_m(k) = \begin{cases} \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ 0 & k < f(m-1), k > f(m+1) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \end{cases} \quad (10)$$

式中 $\sum_{m=0}^{M-1} H_m(k) = 1$, Mel 滤波器的中心频率定义为

$$f(m) = \frac{N}{F_s} B^{-1} \left(B(f_h) + m \frac{B(f_h) - B(f_l)}{M+1} \right) \quad (11)$$

式(11)中: f_h 和 f_l 分别为滤波器组的最高频率和最

低频率, F_s 为采样频率, M 是滤波器组的数目, N 为 FFT 变换的点数, $B^{-1}(b) = 700 \left(e^{\frac{b}{1125}} - 1 \right)$ 。

(e) 计算每个滤波器组输出的对数能量:

$$S(m) = \ln \left(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k) \right), \quad 0 \leq m < M \quad (12)$$

(f) 经余弦变换求得 MFCC 的 M 个系数:

$$C(n) = \sum_{m=0}^{M-1} S(m) \cos(\pi n(m+0.5)/M) \quad 0 \leq n < M \quad (13)$$

按照上述计算流程提取的 MFCC 反映了信号的倒谱特征,用于识别稳态声样本, MFCC 特征考虑了人耳的听觉特性,将频谱转化为基于 Mel 频率的非线性频谱,然后转换到倒谱域上。由于充分模拟了人的听觉特性,且没有任何前提假设,在语音识别中 MFCC 参数具有识别性能和抗噪能力,因而被广泛使用。

2 实验

2.1 实验方法

2.1.1 声样本

本文所用的声样本分为两类,分类依据来自工厂职业试听员的判断,第一类为有质量问题的扬声器所发出的声音,第二类为质量好的扬声器所发出的声音,每类各有 100 个样本。实验中所用的扬声器均由同一生产厂家提供,在消声室环境下,使用 pulse 3560b 设备向扬声器发送 20~5000 Hz 的扫频信号,扫频速度为 2500 Hz/s,并用 BK 公司生产的传声器采集扬声器的发声,采样频率为 65530 Hz,每个样本的录制时间为 10s。

2.1.2 预处理

从录制的声音文件中提取出一段完整的扫频信号,将该信号的频率降采样为 22050 Hz,并采用 hamming 窗进行分帧处理,帧长 N 为 512,并规定 50% 的帧重叠。经分析发现扬声器的质量问题多发生在低频部分,所以重点选择低频部分进行处理。

2.1.3 特征提取与选择

提取上述听觉特征:时域质心(1 维)、零交点比率(1 维)、谱质心(1 维)、谱质心带宽(1 维)、谱通量(1 维)、谱下降值(1 维)、谱不规律性(1 维)、谱平整度(1 维)、美尔倒谱系数(12 维),并将这些特征随机组合成描述样本的 20 维特征向量。

由于提取的每个特征其取值范围不同,所以必须对特征集进行规范化处理。规范化可以帮助防止具有较大初始值域的特征与具有较小初始值域的

特征相比权重过大的问题。本文使用了零-均值规范化(z-score 规范化)^[10]。

由于 z-score 规范化中, 特征 A 的值基于 A 的平均值和标准差进行规范化。 A 的值 v 被规范化为 v' , 计算公式为

$$v' = \frac{v - A}{\sigma_A} \quad (14)$$

其中, A 和 σ_A 分别为特征 A 的平均值和标准差。

特征选择^[11]在模式识别领域中扮演着极其重要的角色。一方面, 在样本有限的情况下, 用大量特征来设计分类器, 无论是从计算开销还是从分类器性能来看都不合时宜; 另一方面, 特征和分类器性能之间并不存在线性关系, 当特征数量超过一定限度时, 会导致分类器性能变坏。因此, 进行正确有效的特征选择成为模式识别中必须要解决的问题, 在海量数据条件下尤为重要。我们用一个 M 位的 0 或 1 构成字符串表示一种特征组合, 其中数字 1 表示所对应的特征被选中, 而数字 0 表示所对应的特征未被选中。比如 [1 0 0 0 1 1] 是一种特征组合, 它表明原始特征集包含 6 个特征, 其中第 1, 5, 6 个特征被选中。

本文选择 k-最近邻分类器的识别率作为合适度函数, 采用序列前向选择方法。该方法也称为集合增加法, 它是一种自下而上的搜索方法。先把所需要的特征集合初始化为一个空集, 每次向特征集合中增加一个特征, 当所需要的特征集合达到要求时所得到的特征集合作为算法运行的结果。

2.1.4 分类

本文使用的分类器采用 BP 神经网络分类器^[12]。其设计主要考虑四个方面的因素: (1) 输入层节点数, 一般由提取的特征空间的维数来决定; (2) 隐层层数和隐层节点数的选择, 通常情况下, 选用一层或两层隐层即可; (3) 学习参数和动力因子的选择, 通常情况下学习参数都小于 0.5, 动力因子的选择范围为 [0,1]; (4) 各层神经元的激活函数的选取, 通常隐层选用非线性可导函数, 输出层去线性激活函数。

由于声样本只有 2 类, 类别较少, 所以采用 3 层 BP 神经网络。采用前馈 BP 神经网络分类器, 输入层的输入神经元个数与特征矢量的维数相等, 隐层的神经元个数为 20, 输出层神经元个数等于要识别的目标类别数(为 2), 各层的激励函数分布为 S 型激活函数 tansig、S 型激活函数 tansig 和线性激活函数 purelin。该 BP 网络的训练参数为: 最大循环次数 $max-epoch=100$, 期望误差最小值 $err-goal=$

0.01, 修正权值的学习率 $lr=0.01$ 。

2.2 实验结果

本次实验中第一类为有质量问题的扬声器, 第二类为质量好的扬声器, 扬声器质量好坏的判定由试听员来完成的。3 个试听员都在工厂工作一年以上, 并进行了反复验证, 因此可以认为扬声器的分类是可信的。每类各取 100 个样本随机选择各类的 50 个样本作为训练样本, 输入到三层 Bp 神经网络分类器进行训练, 剩余 50 个样本作为测试样本, 进行 50 次识别, 将识别结果平均得到最终的识别率。提取上述特征, 分别使用各个特征来判别扬声器的质量, 经特征选择方法选择的特征是 MFCC 第 7 维参数、SC 和 SRO 共三维特征参数, 其识别率如图 2 所示, 其中用特征选择方法选择的特征用 CF 表示。对于同样的声样本, 听觉感知特征中的频域特征的有效性是有所不同的。其中 MFCC 对扬声器的识别最为有效。利用特征选择方法所提取特征的识别率比 MFCC 特征的识别率略高一点, 但是其运算量要大很多。

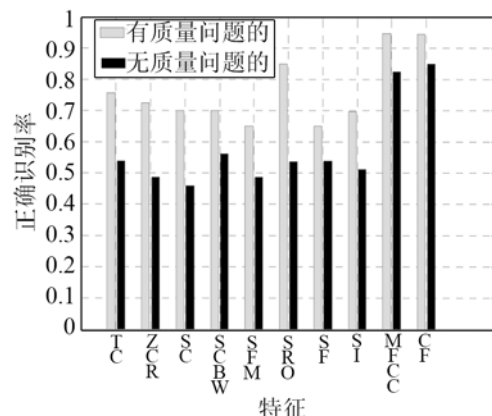


图 2 50 个样本训练, 50 个样本测试时各特征的识别率
Fig.2 The recognition rate of each characteristic for 50 training samples, and 50 testing samples

2.3 运算量分析

本文取一帧长度为 N 的信号做计算量分析, 对于长度为 N 的样本, 时域特征 TC 的运算量为 N 次实数乘法和 $2N-2$ 次实数加法。ZCR 的运算量为 $2N-2$ 次实数加法和 1 次实数乘法。由于 FFT 变化需要 $(N/2)\log_2 N$ 次复数乘法, $N\log_2 N$ 次复数加法^[13]。因为一次复数乘法相当于四次实数乘法, 两次实数加法, 一次复数加法, 相当于两次实数加法。所以频域特征 SC 的运算量为一次 FFT 运算, $3N$ 次实数乘法, $N-1$ 次实数加法, 共 $2N\log_2 N+3N$ 次实数乘法和 $3N\log_2 N+N-1$ 次实数加法。SCBW 运算量为 SC 的运算量加上 $N+1$ 次实数乘法和 $2N$ 次实

数加法共 $2N \log_2 N + 4N + 1$ 次实数乘法和 $3N \log_2 N + 3N - 1$ 次实数加法。SF 运算量为一次 FFT 变换, $2N$ 次实数乘法和 $N - 1$ 次加法运算, 共 $2N \log_2 N + 2N$ 实数乘法和 $3N \log_2 N + N - 1$ 。SRO 运算量为一次 FFT 变换, N 次实数乘法和 $N - 1$ 次实数加法, 共 $2N \log_2 N + N$ 实数乘法和 $3N \log_2 N + N - 1$ 实数加法。SI 运算量一次 FFT 变换, $3N$ 次实数乘法, $2N - 2$ 次加法共 $2N \log_2 N + 3N$ 实数乘法和 $3N \log_2 N + 2N - 2$ 。SFM 运算量为一次 FFT 变换, $2N$ 次实数乘法, N 次实数加法共 $2N \log_2 N + 2N$ 实数乘法和 $3N \log_2 N + N$ 。M 个滤波器组的 MFCC 运算量为一次 DFT 变换, $M + 2N * M + M * M$ 次乘法和 $2M$ 次加法。各种特征运算量的对比如表 1 所示。

表 1 各个特征的运算量比较(N=512,M=12)

Table 1 The amount of computation for different characteristics

特征	实数乘法	实数加法	总运算量
TC	512	1022	1534
ZCR	1	1022	1023
SC	10752	5122	15874
SCBW	11265	6146	17411
SF	10240	5122	15362
SRO	9728	5122	14850
SI	10752	5633	16385
SFM	10240	5123	15363
MFCC	23708	3848	37556

计算机中乘法运算是通过加法运算来完成的, 因此乘法运算的系统开销远大于加法运算, 从表 1 可知, 时域特征的计算量远远低于频域特征的计算量, 而频域特征中除 MFCC 外运算量相差不多。MFCC 的计算量则远远大于其它的频域特征。

2.4 讨论

对扬声器的判别标准是基于人工试听的结果, 所以存在一定的误差, 但我们采取多个试听员多次反复试听的方法最大限度地减少了误差。综合上述结果可以看出, 除了 MFCC 特征外单个特征的识别效果并不理想, 这主要因为有质量问题和无质量问题的扬声器所发出来的声音差别并不大, 尤其是具有音色问题的扬声器和无质量问题的扬声器差异更小。MFCC 的识别效果是单个特征识别效果中最好的, 这与 MFCC 特征本身的特点分不开, 它充分模拟了人的听觉特性, 因而也更符合人耳的听觉特性。经特征选择方法所选择的特征有较高的识别率, 同时也降低了特征的维数。但就计算量来说, MFCC 和特征选择方法的运算量要远远高于其它特征。第 2 类的识别效果总体上比第 1 类的识别效果差, 可能是由于我们评判识别是否正确的依据仍然是试听员的识别结果, 试听员可能对有些扬声器缺

陷项不太敏感, 而把稍微有质量问题的扬声器误判为质量好的扬声器, 也可能影响识别率。

3 结论

本文分析了传统扬声器质量检测方法中存在的不足, 并在此基础上提出了基于音色特征的自动判别扬声器质量的方法, 该方法充分考虑了音色对扬声器品质的影响, 比较了各个音色特征对扬声器质量判别的结果, 同时也比较了它们的计算量, 由实验结果可知, MFCC 和经选择的特征有较高的识别率, 但其计算量相对较大, 考虑了音色特征对扬声器质量的影响, 不但提高了识别率, 而且检测结果更符合人耳的听觉习惯。

参 考 文 献

- [1] 马大猷. 声学名词术语[M]. 北京: 海洋出版社, 1984.
- [2] Lavandier M, Meunier S, Herzog P. Identification of some perceptual dimensions underlying loudspeaker dissimilarities[J]. *Acoust. Soc. Am*, 2008, **123**(6): 4186-4198.
- [3] Zhang X, Ras W Z. Analysis of sound features for music timbre recognition[C]// 2007 Int. Conf. N. on Multimedia and Ubiquitous Engineering(MUE'07), 2007, 3-8.
- [4] Park T H. Towards automatic musical instrument timbre recognition[J]. Ph. D. Thesis, USA: Princeton University, 1998.
- [5] Eronen A. Automatic musical instrument recognition[J]. Master Thesis, Finland: Tampere University of Technology, 2001.
- [6] Chu S, Narayanan S, Kuo C.-C. J. Environmental sound recognition with time-frequency audio features[J]. *IEEE Trans. on Audio, Speech, and Language Process*, 2009, **17**(6): 1142-1158.
- [7] Cowling M, Sitte R. Comparison of techniques for environmental sound recognition[J]. *Pattern Recognition Letters*, 2003, **24**: 2895-2907.
- [8] 王娜, 陈克安. 分段谱质心特征在水下目标识别中的应用[J]. *兵工学报*, 2009, **30**(2): 144-149.
WANG Na, CHEN Ke'an. Application of sub-band spectral centroid features to recognizing underwater targets[J]. *Acta Armamentarii*, 2009, **30**(2): 144-149.
- [9] T.F.Q 著, 赵胜辉等译. 离散时间语音信号处理-原理与应用[M]. 北京: 电子工业出版社, 2004, 566-568.
Thomas F Quatieri. *Discrete-time speech signal processing: principles and practice*[M]. Beijing: Publishing House of Electronics Industry, 2004, 566-568.
- [10] 袁骏, 张明敏. 遗传特征选择方法在水下目标识别中的应用[J]. *声学技术*, 2006, **25**(3): 196-200.
YUAN Jun, ZHANG Mingmin. Feature selection for underwater target recognition based on genetic algorithm[J]. *Technical Acoustics*, 2006, **25**(3): 196-200.
- [11] 王娟, 慈林林. 特征选择方法综述[J]. *计算机工程与科学*, 2005, **27**(12): 68-71.
WANG Juan, CI Linlin. A survey of feature selection[J]. *Computer engineering & science*, 2005, **27**(12): 68-71.
- [12] MATLAB 中文论坛. MATLAB 神经网络 30 个案例分析[M]. 北京: 北京航空航天大学出版社, 2010, 1-10.
- [13] 程佩青. 数字信号处理[M]. 第 2 版. 北京: 清华大学出版社, 2001, 144-150.