

语音识别方法在水下目标识别中的应用

曾 渊, 李钢虎, 赵亚楠, 苗 雨

(西北工业大学航海学院, 西安 710072)

摘要: 水下目标识别是潜艇在海战中, 先敌发现并有效进行水声对抗的关键技术。然而, 如何根据声纳接收到的舰船辐射噪声对三类目标进行分类识别是长期困扰人们的问题。研究了四种语音识别中常用的方法——线性预测系数(LPC), 线性预测倒谱系数(LPCC), 美尔倒谱系数(MFCC)和最小均方无失真响应(MVDR), 在水下目标识别中的应用效果, 并比较了这四种方法在无噪声情况下的识别概率, 以及在不同信噪比下的识别概率, 并通过比较找到在无噪声和有噪声情况下的最佳方法。实验表明, 在无噪声的情况下, MFCC 方法总体识别率最高, 第一类目标 MFCC 方法的识别率最高, 第二类目标 MFCC 和 MVDR 方法识别率相似, 好于其他两者, 第三类目标 MVDR 方法识别率最高。在加入噪声的情况下, MVDR 方法对三类目标的识别和抗噪声性能明显好于其余三者。

关键词: 语音识别; 线性预测系数; 线性预测倒谱系数; 美尔倒谱系数; 最小均方无失真响应

中图分类号: TB566

文献标识码: A

文章编号: 1000-3630(2012)-05-0530-05

DOI 编码: 10.3969/j.issn1000-3630.2012.05.017

Application of speech recognition methods to underwater target classification

ZENG Yuan, LI Gang-hu, ZHAO Ya-nan, MIAO Yu

(Institute of Acoustic Engineering, Northwestern Polytechnical University, Xi'an, 710072)

Abstract: In submarine battle, underwater target recognition is a key technology used for early finding enemy and then taking effective acoustic countermeasure to defeat the enemy. However, a problem that puzzles people for a long time is how to classify and identify three kinds of ship targets based on the received ship radiation noise by sonar system. This paper studies the effects of applying four frequently-used speech recognition methods on underwater target classification. The four methods are linear prediction coefficient (LPC), linear prediction cepstrum coefficient (LPCC), Mel frequency cepstrum coefficient (MFCC) and Minimum Variance Distortionless Response (MVDR). The paper verifies the results by comparing the rates of recognition for different target samples under different SNRs, and finds out the best method in the condition of having or no noise. Experiments show that, without noise, the overall recognition rate of MFCC is highest; MFCC method has the highest recognition rate to the first kind of targets; MFCC and MVDR methods have similar recognition rate to the second kind of targets and better than the other two; and MVDR method has the highest recognition rate to the third kind of target. In the case of adding noise, the recognition and anti-noise performances of MVDR method to the three kinds of targets are significantly better than the other three.

Key words: speech recognition; linear prediction coefficient; linear prediction cepstrum coefficient; mel frequency cepstrum coefficient; minimum variance distortionless response

0 引 言

水下目标识别是水声装备发展的三项关键技术(探测、定位、识别)之一, 是探测系统智能化的重要标志, 一直是声纳信息处理理论中亟待解决的难题之一。水下目标识别主要包括被动声纳目标识别和主动声纳目标识别两种, 前者是利用舰船辐射的噪声信号来识别目标, 而后者则是通过回声信号

来识别目标。被动声纳目标信号不仅含有目标物体的辐射噪声, 而且还包括海洋背景噪声以及本舰噪声等干扰噪声, 分析相当困难, 因此被动声纳目标识别一直倍受世界各国重视。本文主要是针对被动声纳目标进行识别研究。

本文是用语音特征参数来描述舰船辐射噪声。理论上, 人对语音的识别与声纳兵对舰船辐射噪声的识别机理是相同的, 因此, 对语音的特征分析方法也可用于舰船辐射噪声的特征的描述。本文通过语音识别中常用的四种特征——LPC, LPCC, MFCC, MVDR 作为描述目标的特征, 进行水下目标的识别分类^[1]。

收稿日期: 2012-03-20; 修回日期: 2012-05-02

作者简介: 曾渊(1985—), 河南南阳人, 硕士研究生, 研究方向为水声信号处理和目标识别。

通讯作者: 曾渊, E-mail: zengyuan850403@163.com

1 语音识别方法原理

1.1 线性预测系数^[2]

用过去的 p 个时刻的信号采样值组合, 以最小的预测误差预测下一时刻的信号采样值, 定义为信号的 p 阶线性预测, 预测值可表示为

$$\hat{s}(n) = -\sum_{i=1}^p a_i s(n-i) \quad (1)$$

预测误差为

$$e(n) = s(n) - \hat{s}(n) = \sum_{i=0}^p a_i s(n-i) \quad (2)$$

式中, $a_0=1$ 。满足上述预测误差平方和最小的一组系数 a_i ($i=0, 1, 2, \dots, p$) 称为信号的线性预测系数 (Linear Prediction Coefficients, LPC)。

1.2 线性预测倒谱系数^[2]

本文采用的 LPCC 方法是依据全极点模型对 LPC 参数进行递推, 形成 LPC 倒谱, 其形式如下:

$$c_1 = a_1, \quad c_n = \sum_{k=1}^{n-1} \left[1 - \frac{k}{n} \right] a_k c_{n-k} + a_n \quad (1 < n \leq p) \quad (3)$$

式中: a_1, \dots, a_p 为 p 阶 LPC 特征向量, c_1, \dots, c_p 为 p 阶 LPCC 特征向量。

1.3 美尔倒谱系数^[2]

LPC 倒谱是按实际频率尺度直接对信号的对数谱做逆变换得到倒谱的一种形式。如果对信号的频谱按照 Mel 频率尺度进行表示, Mel 频率尺度的值对应于实际频率的对数分布关系, 在 1000 Hz 以下, 中心频率大致为线性划分; 在 1000 Hz 以上则为对数划分。Mel 中心频率划分见表 1。这种频率尺度更适合于表达人耳的听觉特点, 在频率的划分上应该在低频上具有较高的分辨率, 在高频上具有较低的分辨率。然后进行逆变换, 得到倒谱, 这一倒谱称为 Mel 倒谱。

将 LPCC 按符合人耳听觉特性的 Mel 尺度进行非线性变换, 可以求出 Mel 倒谱系数(MFCC):

$$M_{c_0}(n) = c_n + \alpha \times M_{c_0}(n+1)$$

$$M_{c_1}(n) = (1 - \alpha^2) \times M_{c_0}(n+1) + \alpha \times M_{c_1}(n+1)$$

$$M_{c_k}(n) = M_{c_{k-1}}(n+1) + \alpha \times [M_{c_k}(n+1) - M_{c_k}(n)] \quad k > 1 \quad (4)$$

式中: M_{c_k} 表示 Mel 倒谱系数, α 取经验值 0.35, n 为迭代次数, k 为阶数, 取 $n=k$ 。迭代从高到低, 即 n 从大取到 0。最后求得 Mel 倒谱系数 $M_{c_0}(0), M_{c_1}(0), \dots, M_{c_{order}}(0)$ 。

1.4 最小均方无失真响应^[3]

在 MVDR 谱估计方法中, 首先使用特殊设计

表 1 Mel 频率尺度的中心频率划分
Table 1 Center frequency division of Mel-frequency scale

序号	频率	序号	频率
1	195	8	1036
2	312	9	1386
3	429	10	1855
4	546	11	2400
5	664	12	3280
6	800	13	4296
7	917		

的 FIR 滤波器 $\{h_i\}$ 来对输入语音信号进行滤波, 在滤波器输出端计算输出信号的能量从而决定在频率 ω_i 处的能量。 M 阶的 FIR 滤波器 $\{h_i\}$ 设计要满足:

$$H_i(e^{j\omega}) = \sum_{k=0}^M h_k e^{-j\omega k} = 1 \quad (5)$$

这个条件被称为无失真条件。也就是说, 在关心的频率 ω_i 处的频率响应具有单位增益。

其次, 滤波器 $\{h_i\}$ 在满足无失真条件的情况下, 要最小化它的输出能量, 也就是^[3]:

$$\min_{h_i} h_i^H R_{M+1} h_i, \quad \text{subject to} \\ v^H(\omega) h_i = 1 \quad (6)$$

这里 R_{M+1} 是输入信号的 $(M+1) \times (M+1)$ 维的 Toeplitz 自相关矩阵。它的约束最优问题的解是:

$$h_i = \frac{R_{M+1}^{-1} v(\omega)}{v^H(\omega) R_{M+1}^{-1} v(\omega)} \quad (7)$$

最后, 要对 h_i 输出信号进行能量估计。无失真条件保证 MVDR 无失真, 滤波器 $h_i(n)$ 使得由频率 ω_i 组成的输入信号可以无失真地通过, 而输出能量的最小化保证在信号中剩下的频率分量可以最优地抑制。一般而言, MVDR 谱要比相应的 LP 谱平滑, 其原因主要是由于两者之间有如下关系^[4,5]

$$\frac{1}{P_{MV}^{(M)}(\omega)} = \sum_{k=0}^M \frac{1}{P_{LP}^{(k)}(\omega)} \quad (8)$$

其中: $P_{MV}^{(M)}(\omega)$ 为 M 阶 MVDR 谱, $P_{LP}^{(k)}(\omega)$ 为 k 阶 LP 谱。该关系表明: M 阶的 MVDR 谱实际上是自 0 到 M 阶的 LP 谱的谐波均值。这种平均结果使得 M 阶的 MVDR 谱比 M 阶的 LP 谱更加平滑。本文采用的 MVDR 参数是依据全极点模型对 LPC 参数进行递推, 递推关系见式(9)。

$$u(k) = \begin{cases} \frac{1}{P_e} \sum_{i=0}^{M-k} (M+1-k-2i) a_i a_{i+k}^*, & k=0, \dots, M \\ u^*(-k), & k=-M, \dots, -1 \end{cases} \quad (9)$$

其中, a_i 为 LP 谱系数, $u(k)$ 为 MVDR 谱系数。

四种特征提取方法都画出了 50 帧的数据, 每帧样本长度为 20ms, 它们的特征如图 1~4 所示。

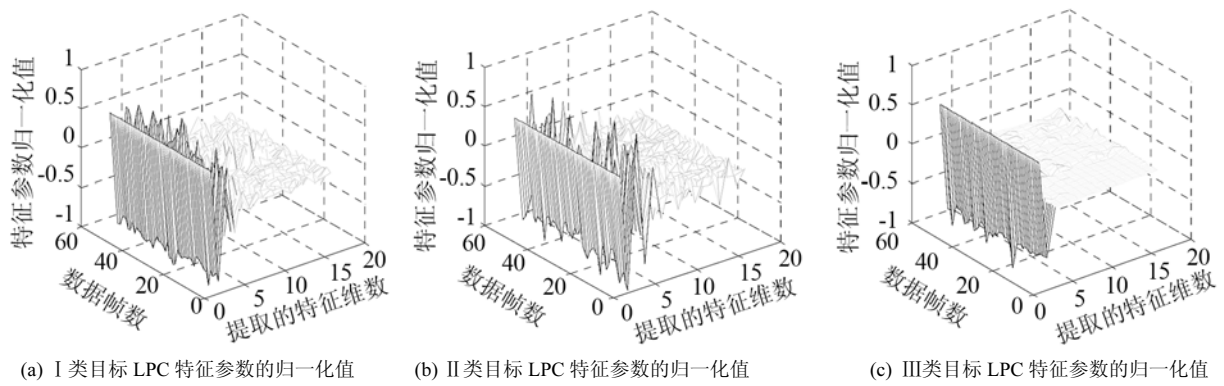


图 1 三类目标 LPC 方法提取特征参数的归一化值

Fig.1 Normalized values of the characteristic parameters extracted by LPC method for three types of targets

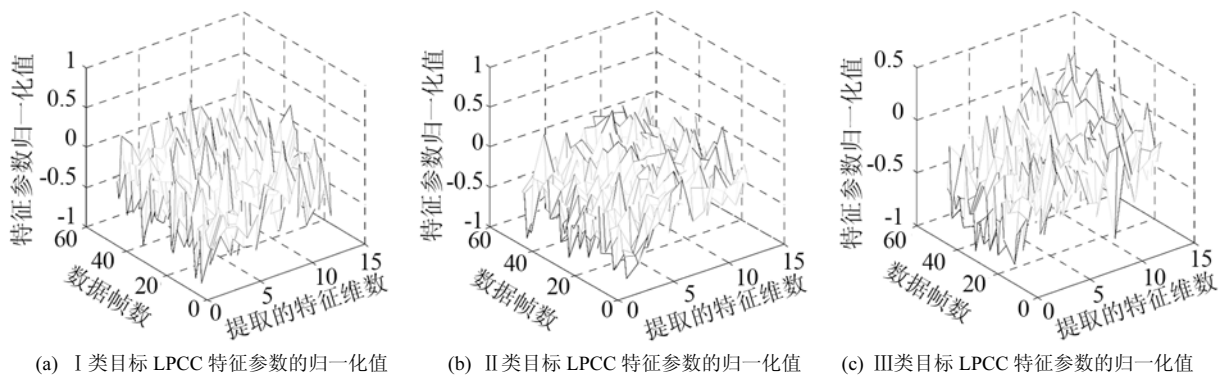


图 2 三类目标 LPCC 方法提取特征参数的归一化值

Fig.2 Normalized values of the characteristic parameters extracted by LPCC method for three types of targets

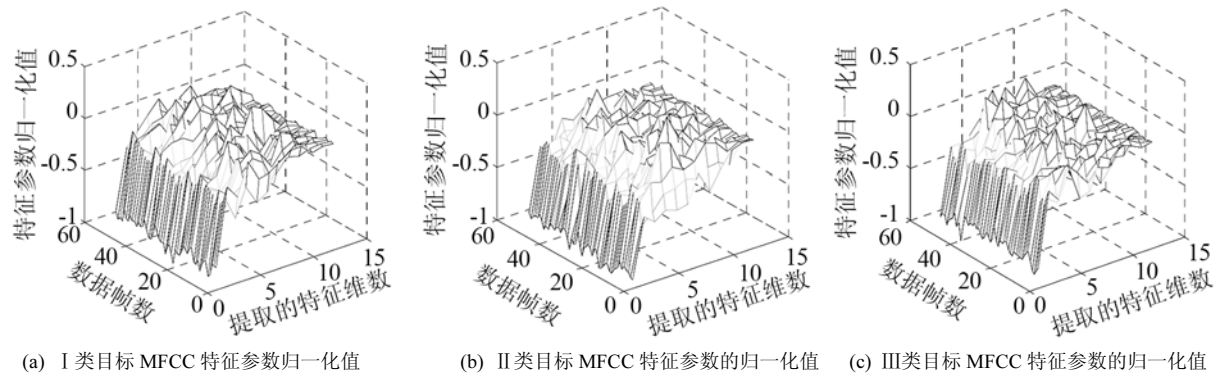


图 3 三类目标 MFCC 方法提取特征参数的归一化值

Fig.3 Normalized values of the characteristic parameters extracted by MFCC method for three types of targets

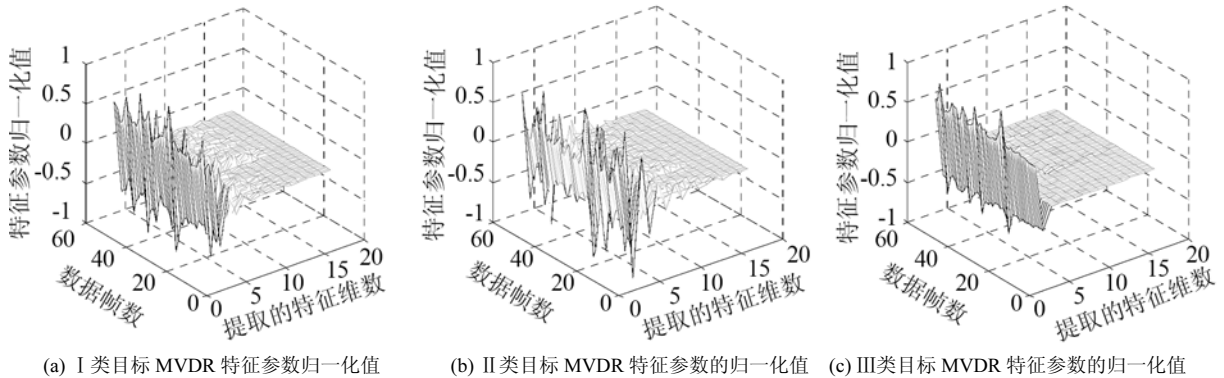


图 4 三类目标 MVDR 方法提取特征参数的归一化值

Fig.4 Normalized values of the characteristic parameters extracted by MVDR method for three types of targets

3 分类器的设计

采用高斯径向基(RBF)支持向量机作为分类器, 取参数 $\delta=3$ 。期望输出矢量与该型对应时输出结果为 1, 否则为 0。

分类正确与否的准则是: 当输出矢量与期望输出矢量完全一致时, 为正确识别, 否则为错误识别。

4 分类实验仿真与结果

用不同的方法, 对水下目标辐射噪声的音频文件进行特征提取, 每类提取的特征向量数量为 300 组, 即对应每种方法共有 900 组数据进行识别。所有方法阶数均为 15。

4.1 无噪声情况下的分类实验仿真

未向文件加入噪声的情况下, 选取不同的训练样本数量进行分类, 结果如表 2~5, 四种方法的总识别率如图 5 所示。

4.2 不同信噪比下的分类实验仿真

向音频文件分别加入信噪比为 20 dB、15 dB、10dB、5dB、0dB 的高斯噪声, 原始信噪比大概是 70dB。信号带宽和人耳听觉带宽一致, 为 0~20kHz。选取 30 个样本进行训练, 分类结果如表 6~9, 四种

表 2 LPC 方法数据分类结果

Table 2 The data classification result of LPC method

训练样本数	LPC 识别结果/%			
	I 类	II 类	III 类	总识别率
10	77.9	79.9	91.8	83.2
15	83.4	85.4	93.5	87.4
20	85.6	82.7	93.8	87.4
30	87.9	86.3	92.8	89.0
45	90.1	86.1	93.5	89.9
60	91.9	87.2	93.8	91.0
75	90.8	89.0	93.0	91.0
90	91.1	89.2	93.7	91.3

表 3 LPCC 方法数据分类结果

Table 3 The data classification result of LPCC method

训练样本数	LPC 识别结果/%			
	I 类	II 类	III 类	总识别率
10	75.6	71.0	90.1	78.9
15	77.5	75.4	89.4	80.8
20	83.7	76.8	89.9	83.5
30	87.2	78.6	91.1	85.6
45	88.8	80.5	91.6	87.0
60	89.0	81.4	91.1	87.2
75	90.4	81.9	91.1	87.8
90	91.1	82.2	92.5	88.6

表 4 MFCC 方法数据分类结果

Table 4 The data classification result of MFCC method

训练样本数	MFCC 识别结果/%			
	I 类	II 类	III 类	总识别率
10	87.2	76.0	94.6	85.9
15	90.1	78.8	96.3	88.4
20	90.7	80.6	96.6	89.3
30	92.2	81.5	96.8	90.2
45	92.4	86.4	97.1	92.0
60	93.3	87.2	97.8	92.8
75	93.8	89.6	97.6	93.7
90	94.6	89.8	98.3	94.2

表 5 MVDR 方法数据分类结果

Table 5 The data classification result of MVDR method

训练样本数	MVDR 识别结果/%			
	I 类	II 类	III 类	总识别率
10	81.3	80.6	98.4	86.7
15	82.5	81.5	99.0	87.6
20	84.2	85.0	99.2	89.4
30	86.2	87.0	99.4	91.0
45	87.2	88.2	99.6	91.7
60	88.8	89.4	99.6	92.6
75	88.9	89.6	99.4	92.6
90	89.1	90.2	99.7	93.0

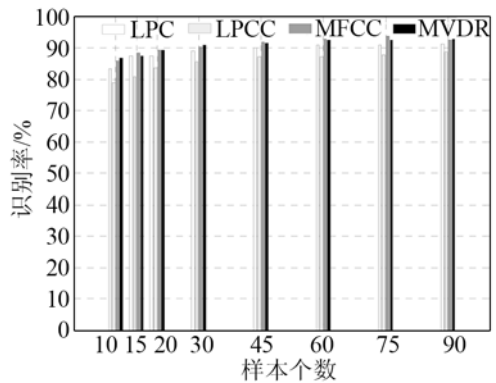


图 5 四种方法在无噪声情况下的总识别率

Fig.5 Total recognition rates of four methods under no noise circumstance

方法的总识别率如图 6 所示。

5 结论

(1) 本文中用到的四种基于语音识别方法的水下目标特征提取方法在水下目标识别中是有效的。

(2) 在无噪声和信噪比较大的情况下, MFCC 方法总体识别率最高, 第一类目标 MFCC 方法的识别率最高, 第二类目标 MFCC 和 MVDR 方法识别率相似, 好于其他两者, 第三类目标 MVDR 方法识别率最高。

(3) 在加入噪声的情况下, 尤其是低信噪比的情况下, MVDR 方法对三类目标的识别和抗噪声性

表 6 LPC 方法数据分类结果

Table 6 The data classification result of LPC method

信噪比/dB	LPC 识别结果/%			
	I 类	II 类	III 类	总识别率
20	82.0	71.2	84.3	79.2
15	81.3	69.1	88.8	79.7
10	75.1	66.9	92.1	78.0
5	69.0	64.8	89.6	74.5
0	63.7	64.4	78.8	69.0

表 7 LPCC 方法数据分类结果

Table 7 The data classification result of LPCC method

信噪比/dB	LPCC 识别结果/%			
	I 类	II 类	III 类	总识别率
20	77.5	67.4	81.4	75.4
15	75.6	66.2	83.6	75.1
10	68.8	63.0	81.1	70.6
5	68.4	63.0	78.8	70.0
0	63.1	61.9	71.0	65.3

表 8 MFCC 方法数据分类结果

Table 8 The data classification result of MFCC method

信噪比/dB	MFCC 识别结果/%			
	I 类	II 类	III 类	总识别率
20	82.6	72.7	88.3	81.2
15	77.4	71.2	85.1	77.9
10	73.1	66.9	72.6	70.9
5	69.5	65.6	71.7	68.9
0	65.1	61.1	68.2	64.7

表 9 MVDR 方法数据分类结果

Table 9 The data classification result of MVDR method

信噪比/dB	MVDR 识别结果/%			
	I 类	II 类	III 类	总识别率
20	82.2	75.3	80.9	79.5
15	85.3	72.5	85.4	81.1
10	82.0	74.7	93.7	83.5
5	72.5	68.1	92.1	77.2
0	71.2	65.2	82.7	73.3

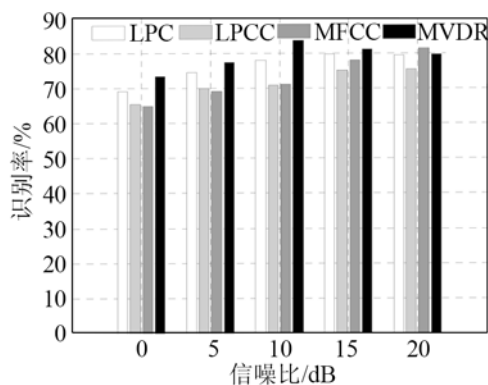


图 6 四种方法在不同信噪比下的总识别率

Fig.6 Total recognition rates of four methods for different SNRs

能明显好于其余三者。

(4) 尚需解决的问题：一是关于每种语音识别方法达到其各自的最佳识别率的阶数问题；二是对于不同的方法，为了到达其各自最佳识别率，支持向量机的参数取值问题，这两个问题有待于进一步研究。

参 考 文 献

- [1] 陆振波, 章新华, 朱进. 基于 MFCC 的舰船辐射噪声特征提取[J]. 舰船科学技术, 2004, 26(2): 50-54.
LU Zhenbo, ZHANG Xin-hua, ZHU Jin, Feature extraction of ship-radiated noise based on Mel-frequency cepstrum coefficients[J]. Ship science and technology, 2004, 26(2): 50-54.
- [2] 柳革命, 孙超, 杨益新. 两种倒谱特征提取技术在水声目标识别中的应用[J]. 西北工业大学学报, 2008, 26(3): 276-281.
LIU Geming, SUN Chao, YANG Yixin. Feature extraction of passive sonar target based on two cepstrums[J]. Journal of northwestern polytechnical university, 2008, 26(3): 276-281.
- [3] Manohar N Murthi, Bhaskar D Rao. All-pole modeling of speech based on the minimum variance distortionless response spectrum [J]. IEEE Transactions on Speech and Audio Processing, 2000, 8 (3): 221-239.
- [4] Manohar N Murthi, Bhaskar D Rao. All-pole model parameter estimation for voice speech[C]// Proc. IEEE Workshop on Speech Coding for Telecommunication. Pennsylvania, USA, 1997: 17-18.
- [5] Manohar N Murthi, Bhaskar D Rao. All-pole modeling of speech based on the minimum variance distortionless response spectrum[C]// Signals, Systems & Computers, Conference Record of the Thirty-First Asilomar Conference, Pacific Grove, California, 1997: 1061-1065.