

最小方差无失真响应浊音谱建模方法研究^{*}马震¹ 陈延萍^{2†}

(1 滨州学院计算机科学系 山东滨州 256600)

(2 山东移动滨州分公司 山东滨州 256600)

摘要 讨论了在语音处理中广泛使用的线性预测模型,分析了线性预测模型在建模语音谱包络方面存在缺点,并分析了线性预测模型过高估计浊音谐波频率处能量以及随着阶数增加线性预测谱逐渐恶化的原因。在此基础上,提出了最小方差无失真响应(MVDR)建模方法,并讨论了通过 Toeplitz 矩阵的 Cholesky 分解确定 MVDR 滤波器系数的方法,通过与线性预测方法进行比较,发现最小方差无失真响应滤波器能提供一个更好的原始语音包络。

关键词 线性预测, 最小方差无失真响应, 语音谱包络建模

Modeling of sonant based on minimum variance
distortionless response spectrumMA Zhen¹ CHEN Yan-Ping²

(1 Department of Computer Science, Binzhou University, Binzhou, Shandong 256600)

(2 Shandong Mobile Binzhou branch, Binzhou, Shandong 256600)

Abstract This paper discusses the LP model which is widely used in speech processing, analyzes its shortcomings in modeling speech spectrum envelop and the reason why LP model over-estimates the power at sonant harmonic frequency and that why LP spectrum deteriorates with the increasing order. The Minimum Variance Distortionless Response Modeling method is then introduced, and how to determine its matrix filter coefficient by means of Toeplitz matrix's Cholesky decomposition is discussed. Comparing with the LP modeling method, the Minimum Variance Distortionless Response Filter is found to provide better envelop of original speech.

Key words Linear prediction, Minimum variance distortionless response, Speech spectrum envelop modeling

2007-01-12 收稿; 2008-03-31 定稿

^{*} 山东省信息产业厅资助(2005R00012)

作者简介: 马震(1980-), 男, 山东滨州市人, 硕士, 研究方向: 语音编码、语音合成、嵌入式系统开发。

陈延萍(1980-), 女, 工程师、学士。

[†] 通信联系人 E-mail: mazhen_chen@263.net

1 引言

语音建模是频域和时域语音压缩系统和自动语音识别系统的基础,全极点模型是建模短时语音谱和获得谱包络的一种流行的参数方法,也是近年来研究的热点。

线性预测(LP)是一种语音全极点建模的常用方法。基于 LP 的全极点建模具有很多优点,所以在语音处理领域得到广泛的应用。虽然常用,可是这种方法还是有它的局限性。基于 LP 的方法使得全极点滤波器对清音和低基音的浊音都工作得很好。然而,El-Jaroudi 和 Makhoul 注意到 LP 模型更适合低基音的说话者,并且随着基音频率减少到 0,LP 的性能会提高。但对于中高基音频率的浊音,基于 LP 的全极点模型不能很好地建模包络^[1]。Oudot 认为 LP 谱包络总是过分强调浊音共振峰处的能量^[2]。另外, Sherman 和 Lou 也注意到 LP 更适合找到噪声中的正弦波频率位置,而不是正弦波的能量^[3]。

本文中,第二部分讨论了 LP 模型的局限性;第三部分在此基础上,提出了在数组处理中很流行的最小方差无失真响应(Minimum

Variance Distortionless Response, MVDR)方法来建模语音谱,并给出了通过 Cholesky 矩阵分解以获得 MVDR 滤波器系数的方法;最后,第四部分对这两种全极点建模方法的结果进行了比较和评价。

2 LP 方法对浊音谱建模的局限性

对同一语音帧,随着 LP 模型阶数的增加,其谱包络在共振峰处也随之尖锐。从图 1 可以看出,对同一帧语音的 LP 谱包络建模,当阶数从 10 变为 18 的时候,LP 谱包络在谐波频率处变得尖锐,特别是在高频范围的谐波频率处,这种尖锐更加明显,而且可以看出 LP 滤波器的极点也更加靠近单位圆。如果把 LP 方法用于语音处理,上述的情况在合成语音中体现为有金属声等问题。

相同的阶数 18,不同频率语音帧的 LP 建模情况如图 2 所示,可以看出随着频率的增高,LP 谱包络不能很好的匹配语音谱,轮廓也变的越来越尖锐,特别是在对语音最重要的第一个和第二个谐波频率处这种变尖锐的问题尤为明显。

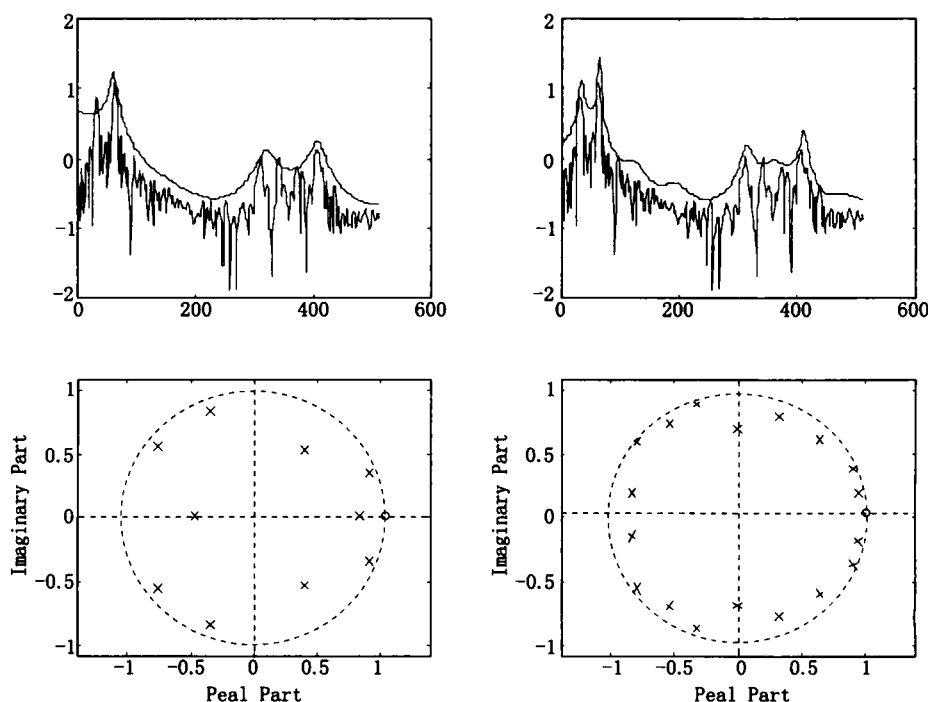


图 1 同一帧语音的 10 阶 LP 谱(左)和 18 阶 LP 谱(右)

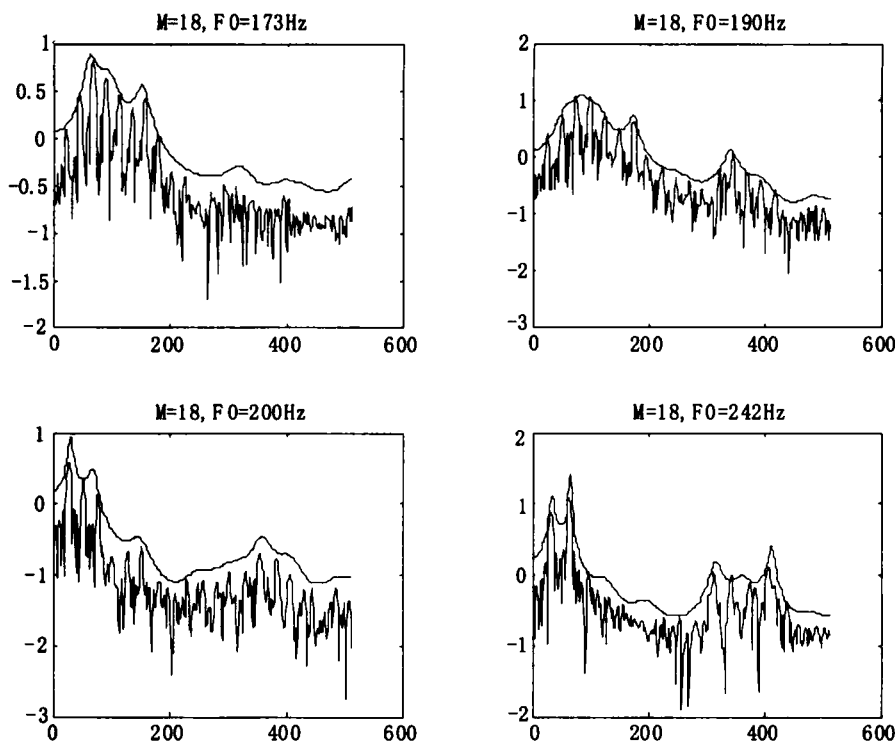


图2 不同频率的语音帧的18阶LP谱
(左上: $F_0 = 173\text{Hz}$; 右上: $F_0 = 190\text{Hz}$; 左下: $F_0 = 200\text{Hz}$; 右下: $F_0 = 242\text{Hz}$)

3 MVDR 建模

上面讨论了 LPC 存在的一些局限性,为了解决基于 LP 的谱包络估计方法的缺点,这里使用 MVDR (Minimum Variance Distortionless Response) 谱方法来建模语音。MVDR 方法是在数组处理中很流行的一种方法。在语音建模方面,相对于 LP 方法过分强调解析谐波位置,而使得各个谐波频率处有过高的能量,MVDR 建模方法更着重于精确的估计语音谱的能量,使得 MVDR 谱包络能和语音谱更匹配。

3.1 MVDR 方法概述

在 MVDR 谱估计方法中,首先使用特殊设计的 FIR 滤波器 $h_l(n)$ 来对信号进行滤波,在滤波器输出端计算输出信号的能量从而决定在频率 ω_l 处的能量。在 M 阶的 FIR 滤波器 $h_l(n)$ 设计中要满足:

$$H_l(e^{j\omega_l}) = \sum_{k=0}^M h_l(k) e^{-j\omega_l k} = 1 \quad (1)$$

这个条件被称为无失真条件。可以写作 v^H

$(\omega_l) h_l = 1$, 这里 $v(\omega) = [1, e^{j\omega}, e^{j2\omega}, \dots, e^{jM\omega}]^T$, 而 $h_l = [h_0, h_1, \dots, h_M]^T$ 。也就是,在关心的频率 ω_l 处的频率响应具有单位增益。

其次,无失真滤波器 $h_l(n)$ 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

$$\min_{h_l} h_l^H R_{M+1} h_l \text{ 并满足 } v^H(\omega_l) h_l = 1 \quad (2)$$

这里 R_{M+1} 是输入信号的 $(M+1) \times (M+1)$ 维的 Toeplitz 自相关矩阵。这个限制最优问题的解是^[4,5]:

$$h_l = \frac{R_{M+1}^{-1} v(\omega_l)}{v^H(\omega_l) R_{M+1}^{-1} v(\omega_l)} \quad (3)$$

最后,要对 $h_l(n)$ 输出信号进行能量估计。无失真滤波器在 ω_l 处的冲击响应用 $h_l(n)$ 来表示,信号能量谱为 $S_{ww}(\omega)$, 则其 MVDR 谱估计为:

$$P_{MV}(\omega_l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 S_{ww}(e^{j\omega}) d\omega \quad (4)$$

这里 $P_{MP}(\omega)$ 表示 MVDR 功率谱。无失真条件保证 MVDR 无失真滤波器 $h_l(n)$ 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。这个相互制约的最优限制是 MVDR 方法的关键方面,这个特点也

就克服了 LP 方法只能解析谐波位置而不能很好的估计谐波频率处包络能量的问题,也让它能够更好的建模语音。同一语音帧的 LP 和 MVDR 谱如图 3 所示。

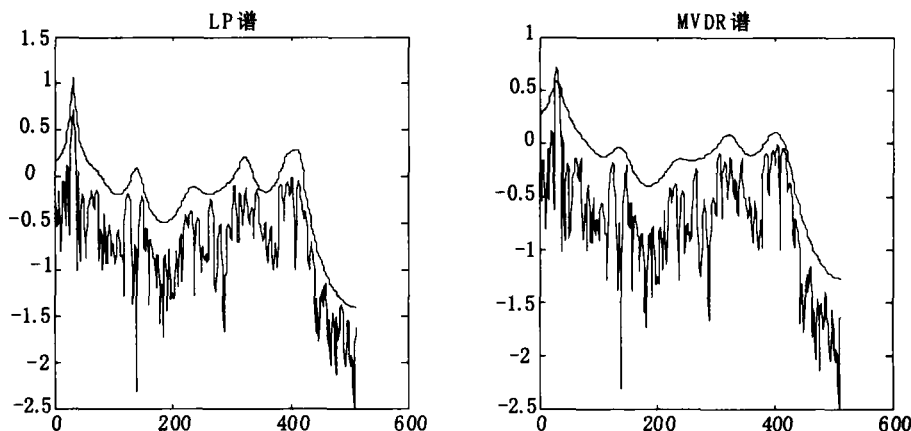


图3 同一语音的 MVDR(右)和 LP(左)谱

3.2 MVDR 滤波器及其参数化方法

从上面的讨论,可以看出 MVDR 方法中必须为每个 ω_i 设计一个单独的 $h_i(n)$,但是,可以像 LP 方法一样计算一个 MVDR 滤波器的参数表示。

M 阶 MVDR 谱可以写为:

$$P_{MV}^M(\omega) = \frac{1}{\sum_{k=-M}^M u(k) e^{-j\omega k}} = \frac{1}{|B(e^{j\omega})|^2} \quad (5)$$

于是可以使用谱因子分解来得到全极点模型的传输函数,确定一个最小相位的稳定滤波器。

$1/B(z)$ 可以通过 Toeplitz 矩阵的 Cholesky 分解^[6]来确定,Cholesky 分解可以通过迭代计算来实现,计算量适合实时处理。首先使用谱系数来构建一个 $(N+1) \times (N+1)$ 的 Toeplitz 矩阵,这个矩阵的第一列由 MVDR 系数 $\mu(0), \dots, \mu(N)$ 构成,矩阵初始化为:

$$T_0 = [\mu(0)] \quad (6)$$

依次为:

$$T_1 = \begin{bmatrix} \mu(0) & \mu(1) \\ \mu(1) & \mu(0) \end{bmatrix}, T_2 = \begin{bmatrix} \mu(0) & \mu(1) & \mu(2) \\ \mu(1) & \mu(0) & \mu(1) \\ \mu(2) & \mu(1) & \mu(0) \end{bmatrix}, \dots \quad (7)$$

对于 $N > M$,将零填入此列, T_N 变为带状的, M 为 MVDR 滤波器的阶数。

给定一个矩阵 T_N ,它的 Cholesky 分解返回一个 $(N+1) \times (N+1)$ 单位下三角矩阵 L_N , L_N 的对角线元素可以构成一个单位对角阵,和一个 $(N+1) \times (N+1)$ 对角矩阵 r_N ,满足方程

$$T_N = L_N r_N L_N^T \quad (8)$$

通过 Cholesky 分解可以使用近似公式(9)来逼近 MVDR 滤波器的传输函数:

$$B(z) \approx \sqrt{r_N} \sum_{k=0}^M L_{N(N-k)} z^{-k} \quad (9)$$

Cholesky 分解执行如下:用 r_k 表示 r_N 的对角线元素, T_{ij} 和 L_{ij} 分别表示 T_N 和 L_N 的元素,从计算 r_0 和 L_N 的第一列开始:

$$r_0 = T_{00} = \mu(0) \quad (10)$$

$$L_{00} = 1 \quad (11)$$

$$L_{i0} = \frac{T_{i0}}{r_0} = \frac{\mu(i)}{\mu(0)} \quad (i=1, \dots, N) \quad (12)$$

每次计算得到一列:

$$r_j = T_{jj} - \sum_{k=0}^{j-1} L_{jk}^2 r_k \quad j=1, \dots, N \quad (13)$$

$$L_{jj} = 1$$

$$L_{ij} = \frac{T_{ij} \sum_{k=0}^{j-1} L_{ik} r_k L_{jk}}{r_i} \quad i = j+1, \dots, N \quad (14)$$

当然,对于 $|i-j| \leq M$ 有 $T_{ij} = \mu(i-j)$, 其它情况 T_{ij} 为零,应用于式(12),可以看到

$$L_{ij} = 0, i > M, j < i - M \quad (15)$$

对 $i > M$, L_N 的第 i 行的第一个元素等于零。因此,可以限制计算,在式(12)中 $i = 1, \dots, \min(M, N)$ 和式(15)中 $i = j+1, \dots, \min(j+M, N)$ 。

可以使用 Cholesky 矩阵的嵌套性质以减少计算时间。嵌套性质在于, T_{N+1} 可以通过附加一行和一列从 T_N 得到,由此得出结论仅通过附加一行到 L_N 可计算 L_{N+1} , 附加一行到 r_N 可计算 r_{N+1} 。

如此,给定 $L_{ij}, i = 0 \dots N, \max(0, i-M) \leq j \leq i$ 对 $j = \max(0, N+M-1) \dots N$, 有:

$$L_{(N+1)j} = \frac{T_{(N+1)j} - \sum_{k=0}^{j-1} L_{(N+1)k} r_k L_{jk}}{r_i} \quad (16)$$

$$r_{N+1} = T_{(N+1)(N+1)} - \sum_{k=0}^N L_{(N+1)k}^2 r_k \quad (17)$$

4 MVDR 滤波器性能分析

4.1 不同基音频率浊音的 LP 和 MVDR 谱性能比较

基音频率分别为 136Hz、190Hz、242Hz 的浊音帧(每帧包括 160 个样点)的 LP 谱和 MVDR 谱如图 4 所示。通过比较发现,对不同的基音频率的浊音, MVDR 谱都要比 LP 谱更加的平滑,很大程度上减少了共振峰频率处谱包络过分尖锐的现象,因此能更好的匹配浊音谱。

4.2 同一帧语音的不同阶数的 LP 和 MVDR 谱性能比较

一帧基音频率为 224Hz 的浊音的 20 阶和 30 阶的 LP 和 MVDR 谱如图 5 所示。通过比较发现,20 阶和 30 阶的 MVDR 谱都要比 LP 谱平滑,能更好地逼近浊音的谱;而且随着阶数的增加 LP 谱在共振峰频率处更加尖锐,而 MVDR 谱的性能则能随着阶数的增加而改善。

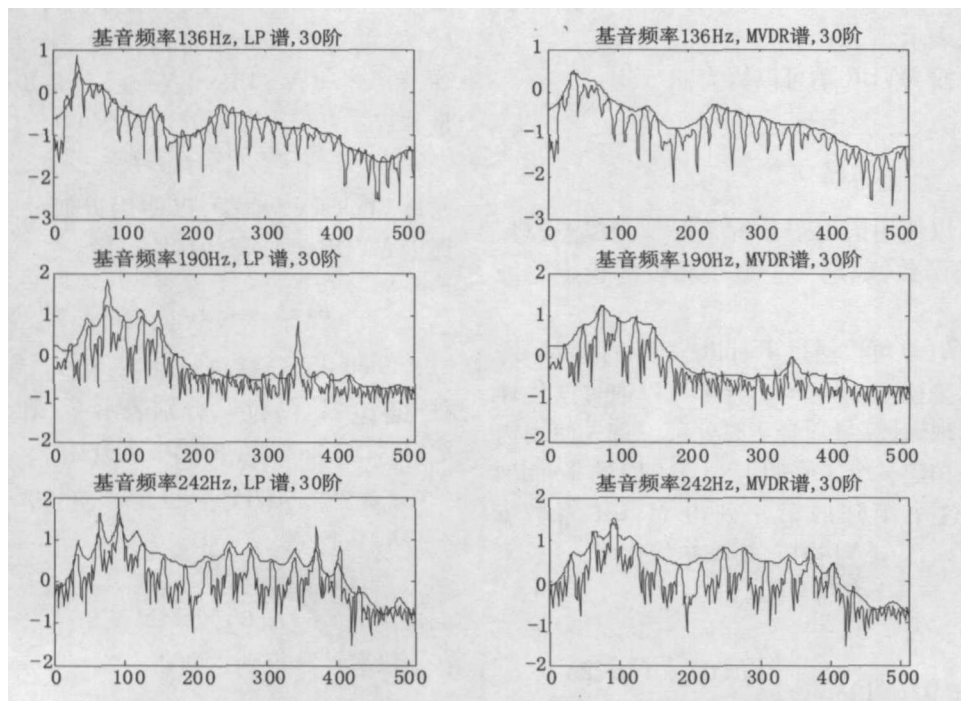


图4 不同基音频率浊音的 LP 和 MVDR 谱

(上:基音频率为 136Hz,中:基音频率为 190Hz,下:基音频率为 242Hz)

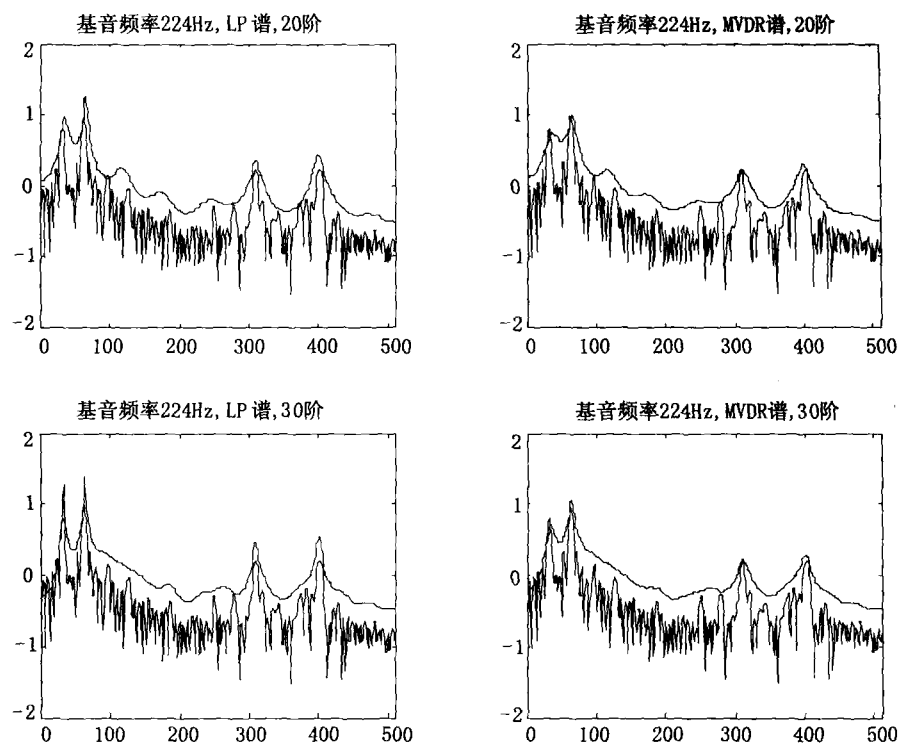


图5 基音频率为224Hz的浊音的20阶、30阶LP和MVDR谱比较

4.3 浊音数据库的LP和MVDR对数谱失真的比较

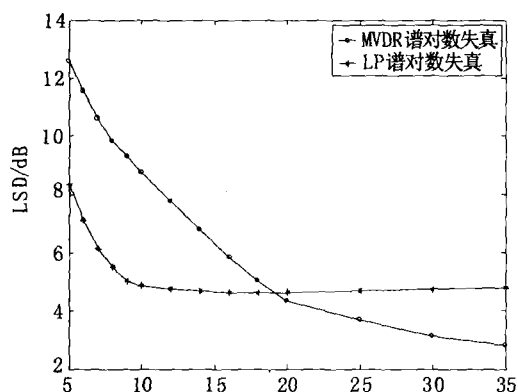


图6 LP和MVDR对数谱失真比较

这里讨论MVDR对一个浊音数据库中浊音帧的整体性能。试验中采用的数据库包括320帧浊音,每帧160样点,对整个数据库中的每一帧浊音分别求其LP和MVDR对数谱失真,然后再求平均。可以从图6看到,随着

阶数的增加,MVDR谱的对数失真是呈下降趋势的,而LP的对数失真趋于平稳甚至有增加的趋势。由此可见,虽然LP谱能够在低阶很好的匹配浊音谱,但是随着阶数的增加,特别是到12阶以上,其对浊音谱的匹配能力增加微弱,而且呈下降的趋势;另一方面,MVDR方法可以在高阶提供优于同阶的LP谱的包络,更精确地对声道模型建模。

5 结论

MVDR谱从理论上保证了可以准确估计语音功率谱的能量,而不是像LP那样能精确解析基音频率位置但过分强调谐波频率处的能量,造成尖锐的包络轮廓。而且,MVDR谱可以通过一个最小相位滤波器来表示,MVDR滤波器的传输函数可以通过Cholesky分解迭代产生,这也可以确保这种方法的计算复杂度不大,可以用于实时语音信号处理。最后本文

给出了不同基音频率的浊音、同一帧语音不同阶数以及清音的 LP 和 MVDR 谱的性能比较。本文仅讨论了 MVDR 方法对浊音谱的建模理论以及 MVDR 滤波器对浊音的工作性能,在 MVDR 方法如何对清音谱建模方面有待于进一步研究。

参 考 文 献

- [1] A. El-Jaroudi and J. Makhoul. Discrete all-pole modeling, . IEEE Trans. Signal Processing, Feb. 1991, vol. 39:411 - 423.
- [2] M. Oudot, O. Cappe, and E. Moulines. Robust estimation of the spectral envelope for harmonics + noise models. IEEE Workshop Speech Coding Telecommunications Proceedings, 1997.
- [3] P. J Sherman and K. -N. Lou. On the family of ML spectral estimates for mixed spectrum identification. IEEE Trans. Signal Processing, Mar. 1991, vol. 39:644 - 655.
- [4] S. Haykin. Adaptive Filter Theory. Englewood Cliffs, NJ: Prentice Hall, 1991:273 - 285.
- [5] S. L. Marple, Jr. Digital Spectral Analysis with Applications. Englewood Cliffs, NJ: Prentice-Hall, 1987:159 - 172.
- [6] Sayed A. H. , Kailath T. A survey of spectral factorization methods. Numerical Linear Algebra with Applications, 2001, 08: 467 - 496.

欢 迎 购 买

《应用声学》合订本

为更好地服务于各界读者,本刊装订了《应用声学》部分年度的合订本。如购买 2000 年至 2002 年合订本,优惠价为 40 元/本。2003 年至 2007 年合订本,优惠价为 50 元/本(以上价格均免邮、挂费)。数量有限,售完为止。

邮购者请在汇款单上写清详细地址、姓名、邮编及所需年卷、册数。

邮购地址:北京海淀北四环西路 21 号 《应用声学》编辑部 邮编:100190