

噪声环境下的汉语声调识别

顾明亮^{1,2}, 夏玉果², 王劲松¹

(1. 徐州师范大学 物理系, 江苏 徐州 221116;

2. 江苏省语言科学与神经认知工程重点实验室, 江苏 徐州 221116)

摘要:利用语音信号与噪声信号具有不同相关特性的特点,提出了一种新的加权自相关基频检测算法,该方法可以提高噪声环境下基音检测的准确性。在分类器设计方面,通过引入支持矢量机,进一步提高低信噪比下的汉语声调识别率。实验结果表明,新方法对提高噪声环境下的声调识别效果是十分有效的。

关键词:基音检测;支撑矢量机;汉语声调识别

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2007)08-0070-03

Chinese Tone Recognition under Noise Conditions

GU Ming-liang^{1,2}, XIA Yu-guo², WANG Jin-song²

(1. Physics Department, Xuzhou Normal University, Xuzhou 221116, China;

2. Jiangsu Key Lab. of Language Science and Neural Cognition Eng., Xuzhou 221116, China)

Abstract: According to the fact that speech and noise have different correlation character, presents a novel pitch extraction algorithm based on the weighed correlation. This approach can raise the correction of pitch extraction for noise speech. On the design of classifier, by introducing support vector machine, Chinese tone recognition rate is improved under low signal noise rate conditions. The experimental results show that these improvement methods is very useful.

Key words: pitch extraction; support vector machine; Chinese tone recognition

0 引言

声调作为汉语信息处理的重要特征,在汉语语音识别、汉语语音合成、汉语方言辨识中具有广泛的应用。理想实验环境下的汉语孤立词声调识别,已经取得了很好的成果^[1,2],但噪声环境下的汉语声调识别效果还不尽如人意,这里既有基频检测的问题,也有声调特征的选取问题,还有分类器设计方面的问题。由于汉语声调识别主要依赖于基频的轮廓特征,无疑基频检测是人们关注的焦点,因此,目前已有各种各样的容错性基频检测算法,其中较为公认的容错算法是各种改进的自相关算法^[3,4]。因为噪声信号通常不具有相关性,而语音信号尤其是浊音信号具有较强的相关性,但此种方法仍然存在半频和倍频现象,特别是信噪比较低时更为明显。文中在文献[3]的基础上,进一

步利用线性变换提升自相关函数在噪声环境下的峰值特性,使噪声环境下的基音频率提取更为准确。在声调特征方面,则主要采用当前主流的特征提取方法,即以各帧的基频、能量及其一阶差分作为其特征参数。在分类器设计方面,已提出了隐马尔可夫模型(HMM)方法^[1]、人工神经网络(ANN)方法^[5]和模糊分类^[6]等方法。文中利用支持矢量机(SVM)这一新的模式分类器来提高非线性决策能力和抗干扰能力。

1 改进的基音检测算法

声调与语音中的基频密切相关,基音检测一直是声调识别中最主要的一环。文献[4]结果表明,在传统的基音提取算法中,自相关法和平均幅度差法具有最好的抗干扰能力,是噪声条件下提取基音的最佳选择。随后各种改进的算法也相继提出,文中在加权自相关方法^[3]的基础上做了进一步的改进。

设分帧后的语音信号为 $x(n)$, 帧长为 N , 延迟为 τ , 则语音信号的短时自相关函数(ACF)为:

$$\varphi(\tau) = \sum_{n=0}^{N-1} s(n)s(n+\tau) \quad (1)$$

收稿日期:2006-11-20

基金项目:江苏省“十五”社科基金项目(K3-013);江苏省高校自然科学基金(99KJB510002)

作者简介:顾明亮(1963-),男,江苏无锡人,博士,副教授,硕士生导师,研究方向为数字语音信号处理、神经网络理论与应用、模式识别、机器学习等。

若带噪语音信号 $s(n) = x(n) + \omega(n)$, $x(n)$ 为干净的语音, $\omega(n)$ 为高斯白噪声。由公式(1)可得:

$$\begin{aligned} \varphi(\tau) &= \sum_{n=0}^{N-1} (x(n) + \omega(n))(x(n+\tau) + \omega(n+\tau)) \\ &= \sum_{n=0}^{N-1} (x(n)x(n+\tau) + x(n)\omega(n+\tau) + \omega(n)x(n+\tau) + \omega(n)\omega(n+\tau)) \\ &= \varphi_{xx}(\tau) + 2\varphi_{xw}(\tau) + \varphi_{ww}(\tau) \end{aligned} \quad (2)$$

其中 $\varphi_{xx}(\tau)$ 表示干净语音 $x(n)$ 的自相关函数, $\varphi_{ww}(\tau)$ 表示噪声信号 $\omega(n)$ 的自相关函数, $\varphi_{xw}(\tau)$ 表示语音 $x(n)$ 和噪声 $\omega(n)$ 的互相关函数。由于 $x(n)$ 和 $\omega(n)$ 没有相关性,所以 $\varphi_{xw}(\tau) = 0$, 噪声本身也没有相关性,所以 $\varphi_{ww}(\tau) = 0$ ($\tau = 0$ 除外)。因此, (2) 式可写成:

$$\begin{cases} \varphi(\tau) = \varphi_{xx}(\tau) + \varphi_{ww}(\tau) & \text{当 } \tau = 0 \text{ 时} \\ \varphi(\tau) = \varphi_{xx}(\tau) & \text{当 } \tau \neq 0 \text{ 时} \end{cases} \quad (3)$$

由此可见,短时自相关函数(ACF)具有一定的抗噪能力。当语音信号具有准周期的浊音信号时(设周期为 T),则 ACF 将在 T 的整数倍处出现峰值。通常第一个峰值点的值是除原点外的最大值,它与原点的时间间隔定义为基音周期,其倒数即为基频。由于噪声干扰,第一个峰值点处的值不一定是最大的,因此,就造成倍频和分频现象。

短时平均幅度差函数(AMDF)与 ACF 类似,同样具有与语音信号相同的周期特性,它定义为:

$$\varphi(\tau) = \sum_{n=0}^{N-1} |x(n) - x(n+\tau)| \quad (4)$$

对于浊音信号, $\varphi(\tau)$ 在基音周期整数倍的时间点上呈现谷点。它的第一个谷点的值一般也是全局最低谷点($\varphi(0)$ 除外),可以用来计算基音周期。但当浊音信号的周期性和平稳性不太好或噪声干扰时,第一周期谷点与全局最低谷点不重合,也会产生半频和倍频的错误,这种情况在噪声的基音检测中更为严重。

为了克服用 AMDF 作基音检测所出现的这种缺点,这里对 AMDF 作线性变换,即用一直线减去 AMDF,从而得到一个修正的 AMDF:

$$\gamma(\tau) = R_{\max} \left(\frac{N-\tau}{N-r_{\max}} \right) - \varphi(\tau) \quad \tau = 0, 1, 2, \dots, N \quad (5)$$

这里 $R_{\max} = \max\{\varphi(\tau)\}$, r_{\max} 是取 $\varphi(\tau)$ 最大值时所对应的位置。这条直线的斜率由 AMDF 的最大峰值点 (n_{\max}, R_{\max}) 和 $(N, 0)$ 端点共同确定。变换的目的是改变 AMDF 各个谷点的深度,并实现 AMDF 的峰点和

谷点的反转。

图 1 给出了一帧语音进行线性变换后的实验结果,原始的 AMDF 中的全局最低谷点在第三个基音周期处(不计起始谷值点),而正确的基音周期应在第一个周期,这样就造成半频现象;在修正的 AMDF 中的最高点与第一周期的最高点重合(不计起始峰值点),从而克服了半频现象。

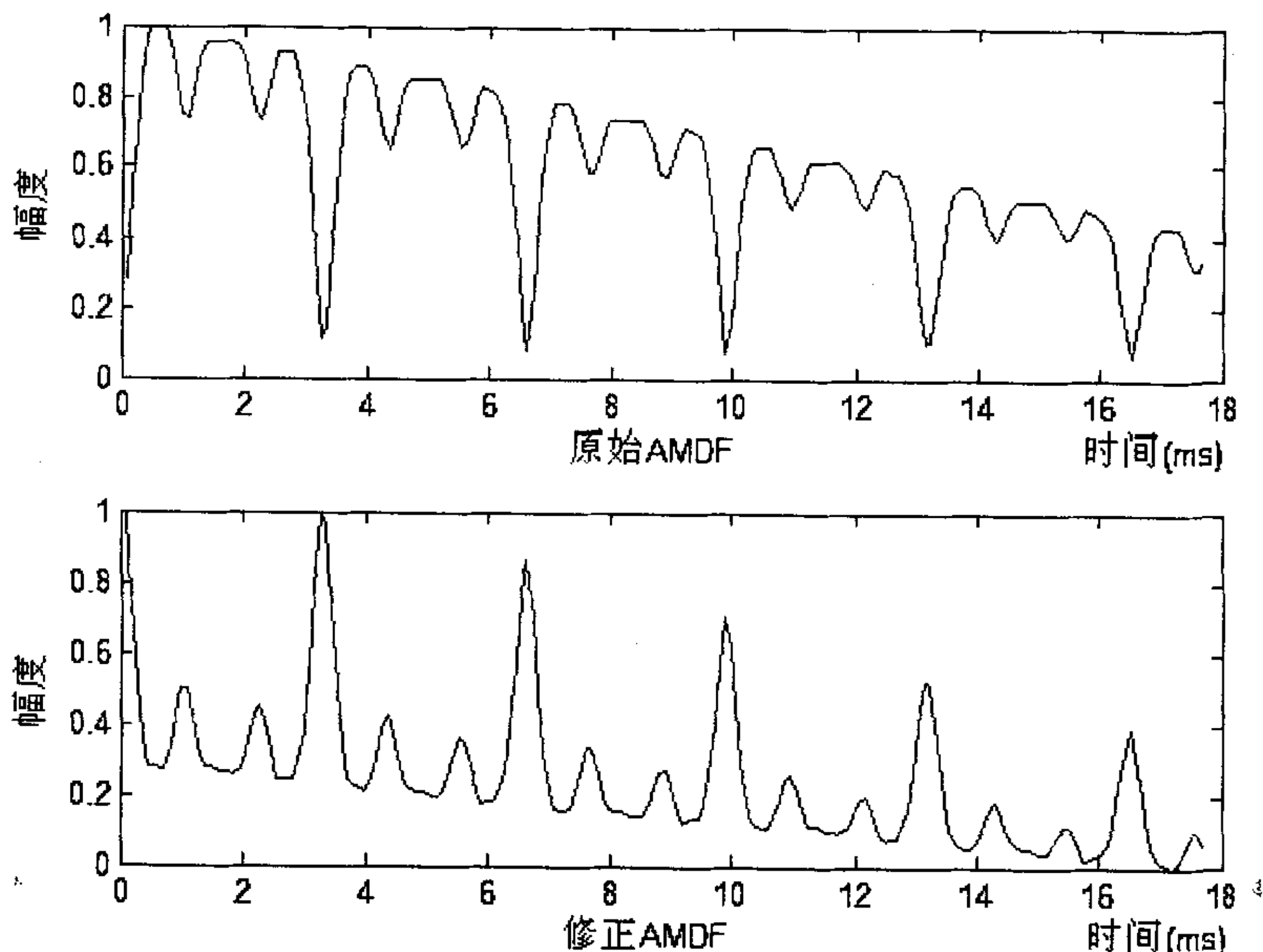


图 1 AMDF 与修正 AMDF 函数比较

由于上面分析可知,短时 AMDF 和修正的短时 AMDF 具有相同的周期性,为了提高噪声环境下的基音检测的准确性,提升短时 ACF 的峰值点,构造一个新的基频检测算法,即改进的基音检测算法,它是短时 ACF 和修正的 AMDF 的乘积:

$$\lambda(\tau) = \varphi(\tau)\gamma(\tau) \quad (6)$$

2 支持矢量机分类器设计

支持矢量机(Support Vector Machine, SVM)是 V. N. Vapnik 在长期研究统计学习理论的基础上,在 20 世纪 90 年代中期提出的一种新的模式分类方法^[7],目前该方法已在模式识别(文字识别、人脸识别等)、数据挖掘和非线性系统控制等诸多领域得到广泛应用^[8]。它的主要特点是:

(1)通过寻找那些对分类有较好区分能力的支持向量,构造分类空隙最大的有较好推广性能和较高分类准确率的最优超平面,如图 2 中黑色点表示支持向量;

(2)通过核函数变换的方法,它能够将低维空间非线性分类问题转换为高维空间线性可分的问题,同时很好地解决了高维空间中的计算复杂性问题;

(3)SVM 构造的模型具有很好的预测性能,不存在过学习问题。

在二类模式的分类问题,支持矢量机的最优超平面可用图 2 表示,它的结构如图 3 所示,其中,最优超平面的分类函数可以表示为:

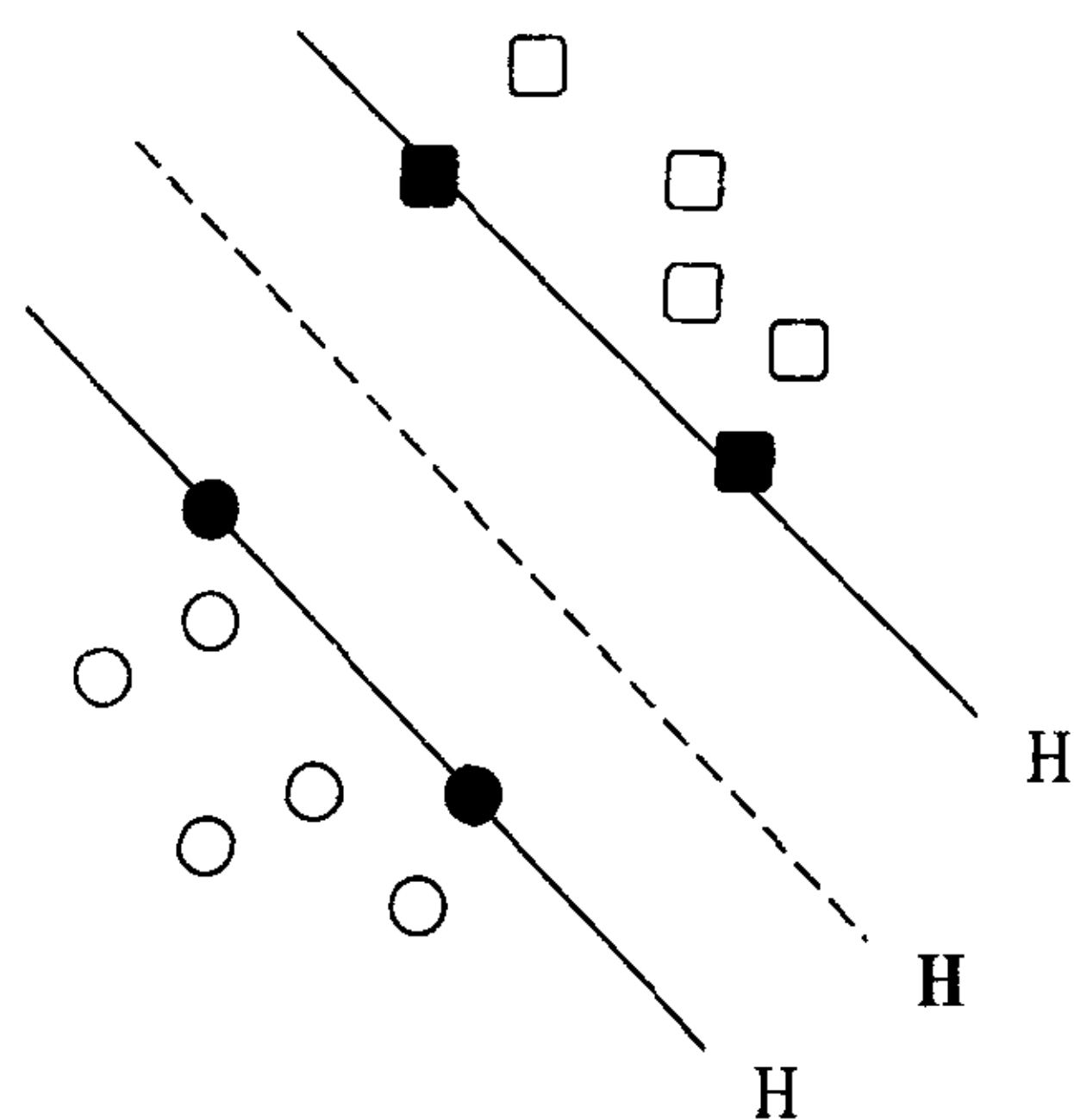


图 2 最优超平面示意图

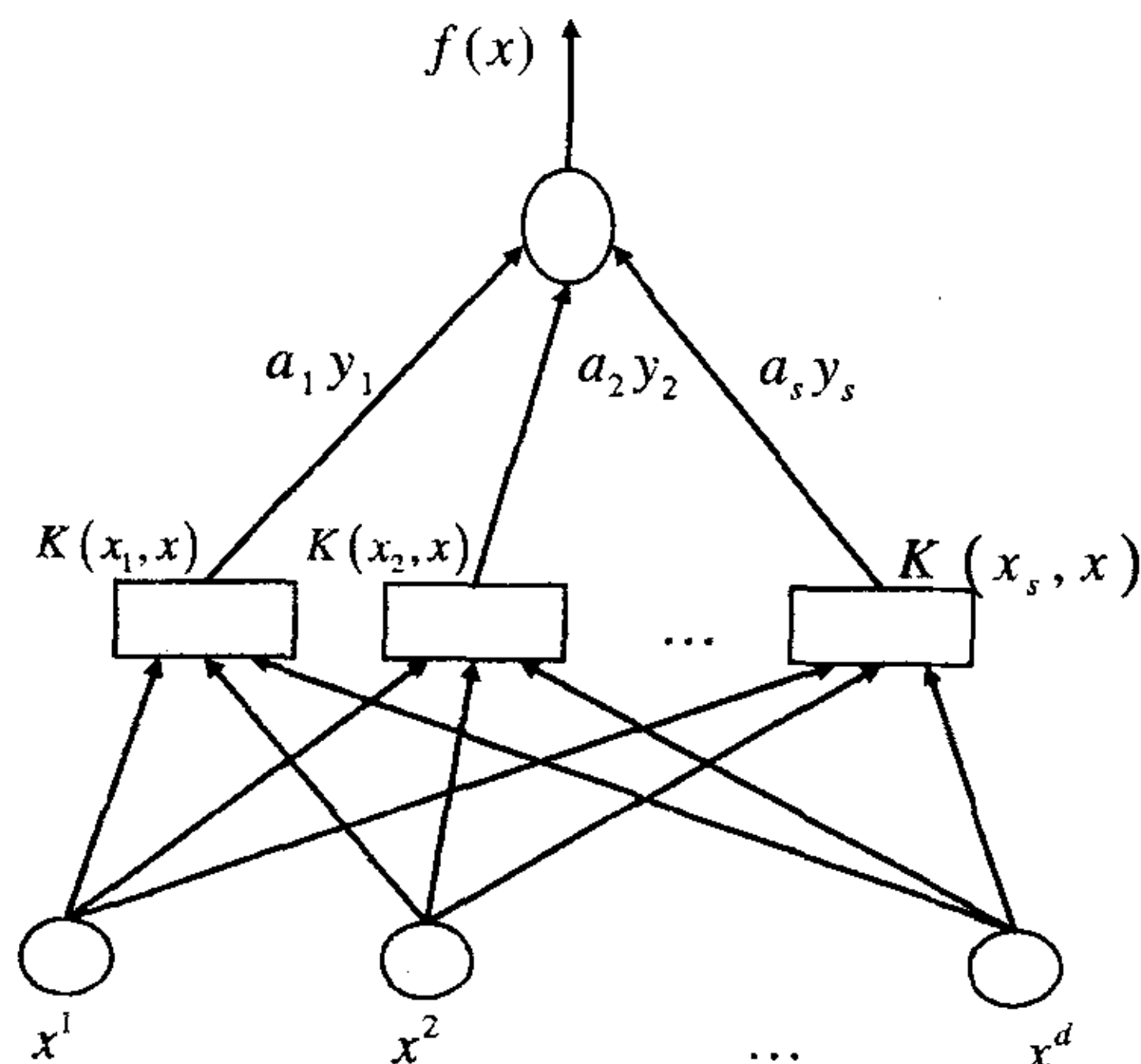


图 3 支持矢量机结构示意图

$$f(x) = \text{sgn}\left(\sum_{i=1}^s a_i y_i K(x_i \cdot x) + b\right) \quad (7)$$

其中 $K(x_i \cdot x)$ 表示内积函数。选择不同形式的内积函数,支持矢量机就会形成不同的算法,常用内积函数有以下三种:多项式核函数,径向基核函数和 sigmoid 型核函数。文中选择了第二种形式的内积函数。

对于多类模式的分类问题,目前主要有“一对一”、“一对其余”和决策导向无环图三种多类分类策略。考虑到“一对一”分类方法精度高、容错性好,而待区分的类别数(四类)较少的实际,采用“一对一”多类分类支持矢量机。它由 Friedmann 首先提出,其原理是:首先利用已知的 K 类训练样本训练 $k(k-1)/2$ 个二类支持矢量机,然后,用这 $k(k-1)/2$ 个二类支持矢量机对测试样本进行分类,最后挑选判入次数最多的类别作为测试样本的分类结果。

3 实验与分析

为了检验特征的可行性和支持矢量机的分类效

果,首先参照 863 语音数据库录制要求,建立了一个专门进行声调训练和测试的语音数据库,然后对特征的齐次化方式、特征的组合方式、核函数对 SVM 分类效果的影响和不同分类器对识别效果的影响等进行了一系列对比实验。必须指出的是,在特征参数计算前需要对语音信号进行必要的预处理,主要包括滤波、分帧、加窗、清浊音区分等多个环节。

3.1 语音数据库的建立

实验所用的发音语料选自《普通话水平测试实用手册》中的 100 个单音节字,这些单字包括单元音和复元音,每个音节大部分都有四声读音。录音由声卡和 Cool Edit 软件共同完成,其中语音的采样频率为 11025Hz、量化级长为 16bit。发音人是 10 位在校大学生(男生和女生各 5 人),年龄介于 18~23 岁,都来自于徐州市区。每个发音人对每个单字读 5 遍共得到 5 000 个单字的语音,发音不准时要求重录。选用其中的 3 男 3 女前三次读音共 1 800 个字音为训练集(记作 TRAIN),他们的后两次发音组成测试集 1(记作 TEST1),其他 2 人发音组成测试集 2(记作 TEST2)。

3.2 噪声环境下声调识别实验结果

对于不同的核函数,SVM 分类器的识别效果是不同的,表 1 给出了基于径向基核函数的声调识别结果,由此可见,随着信噪比的不断降低训练集内样本的识别率变化不大,但两个测试集中的样本识别率有明显降低,其中测试集 2 下降的更为明显。

表 1 噪声环境下声调识别结果

SNR(dB)	TRAIN	TEST1	TEST2
20	98.89%	95.9%	93.3%
10	98.83%	95.7%	92.9%
5	98.67%	94.1%	91.4%
0	97.83%	89.5%	87.6%
-5	96.5%	81.6%	77.2%
-10	95.5%	63.3%	60.3%

3.3 不同基频检测方法下的声调识别比较

为了验证算法的可靠性,实验比较了四种噪声环境下的基音检测算法,提取基音后输入径向基核函数为核函数的 SVM 后端分类器中,分别测试 TEST1 和 TEST2 中的语音,得到的结果如图 4、图 5 所示。从图中可以看出在 TEST1 和 TEST2 中,文中所提出的方法的识别率都要高于其他三种算法,特别在低信噪比的环境下,文中方法的优越性更加明显,从而说明该方法具有一定的抗噪性能。

4 结 论

介绍了一种新的基音检测算法,在低信噪比的情

(下转第 76 页)

生一定影响。如果 PSF 方向是一维的,即:只是 x 方向或者只是 y 方向则离散化的时候不会带来问题,如图 4 所示。如果是二维的,离散化的时候会带来一些问题,如图 5 所示,可以看到,由于离散化的原因 PSF 并非直线。而 PSF 带来的这种误差会对恢复结果产生一定影响。

(3)一些恢复算法都是以物体匀速运动为前提的,但现实中运动物体的速度不可能是完全匀速。

(4)很难具体确定模糊点数,如果差 1 个像素或者 0.5 个像素都将对恢复后的图像产生很大影响。

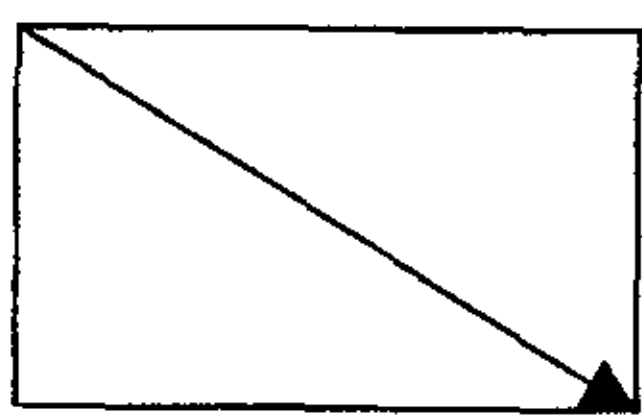


图 4 连续 PSF

图 5 离散 PSF

4 总 结

文中先从分析运动模糊图像退化模型入手,介绍比较了几种运动模糊恢复算法,通过对维纳算法具体的实验,分析仿真运动模糊图片生成中应注意的地方,频率域算法中常见的振铃效应产生的原因及其与维纳滤波算法中参数的关系。最后给出了运动模糊恢复过程中图像质量变差的几个原因。维纳滤波是一种综合

考虑了退化函数和噪声统计特征两个方面进行恢复处理的方法,是最常用的方法。

文中选用维纳滤波进行模糊图像恢复,所遇到的现象及对其产生的原因的分析也可以适用在别的恢复算法中。

参考文献:

- [1] 何 斌,马天予,王运坚,等. Visual C++ 数字图像处理 [M]. 第 2 版. 北京:人民邮电出版社,2002:509-512.
- [2] Sonka M, Hlavac V, Boyle R. 图像处理、分析与机器视觉 [M]. 第 2 版. 北京:人民邮电出版社,2003:70-71.
- [3] Likhterov, Boris, Kopeika, et al. Motion-blurred image restoration using modified inverse all-pole filters[C]// Proceedings of SPIE. [s.l.]:[s.n.],2002:56-62.
- [4] 刘政凯,瞿建雄. 数字图像恢复与重建[M]. 合肥:中国科学技术大学出版社,1989:148-182.
- [5] 陆 俊,舒志龙,阮秋琦. 基于尺度旋转的图像恢复研究[J]. 通讯学报,2000,21(7):67-71.
- [6] 蔡利栋. 传播波方程与运动模糊图像恢复[J]. 自动化学报,2003,29(3):465-471.
- [7] 潘 琪. 运动模糊仿真图像的正确生成[D]. 广州:暨南大学,2005.
- [8] 于红斌,李志能,陈抗生. 一种运动模糊图像的快速恢复算法[J]. 浙江大学学报:工学报,1999,33(5):564-568.

(上接第 72 页)

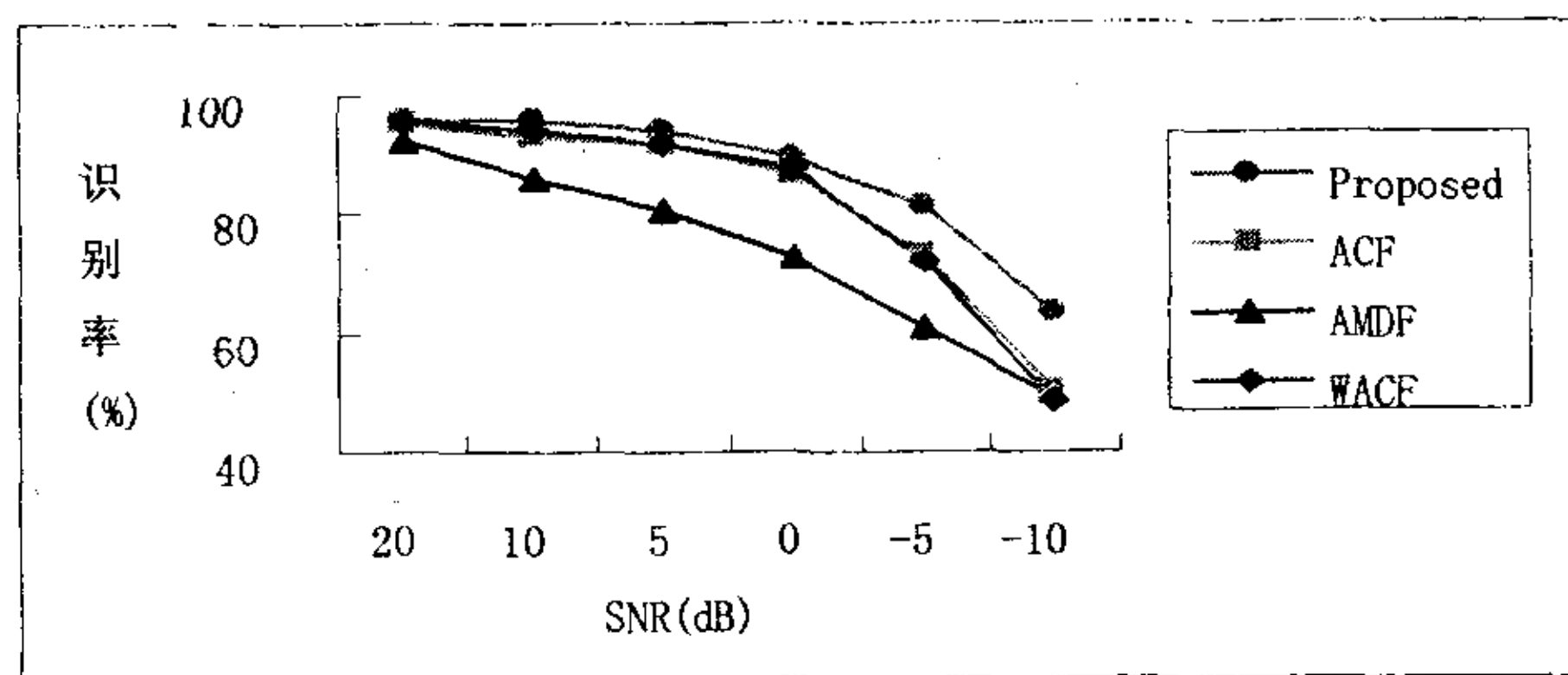


图 4 四种方法在 TEST1 上的实验结果

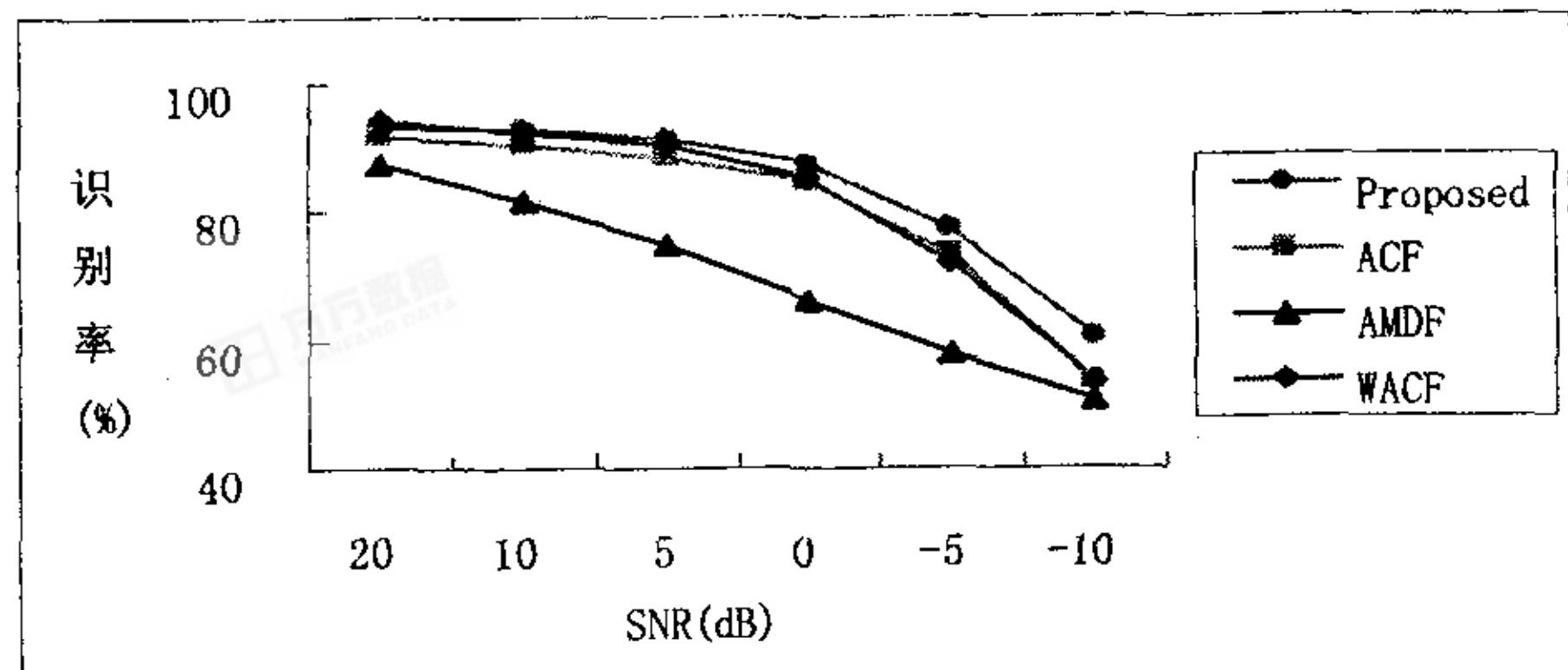


图 5 四种方法在 TEST2 上的实验结果

况下,该算法提取基音频率的准确性有提高作用,将其运用于基于支持矢量机的噪声环境下的声调识别中,实验表明相对于传统的基音提取算法,声调识别率有明显的提高。

参考文献:

- [1] Yang W, Lee J, Chang Y, et al. Hidden markov model for mandarin lexical tone recognition[J]. IEEE Trans on ASSP, 1988,36(7):988-992.
- [2] 关存太,陈永彬. 非特定人四声识别[J]. 声学学报,1993,18(5):379-385.
- [3] Shimamura T, Kobayashi H. Weighted autocorrelation for pitch extraction of noisy speech[J]. IEEE Trans on SAP, 2001,9(7):727-730.
- [4] Oh K A, Un C K. A performance comparison of pitch extraction algorithms for noisy speech[C]//IEEE Trans on ASSP. US:IEEE,1984:85-88.
- [5] S. Chen, Y. Wang, Tone recognition of continuous mandarin speech based on neural networks[J]. IEEE Trans on SAP, 1995,3(2):146-150.
- [6] 徐士林. 四声模糊识别方法[J]. 电子学报,1996,24(1):119-121.
- [7] Vapnik V N. The nature of statistical learning theory[M]. New York:Springer Verlag, 1995.
- [8] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Knowledge Discovery Data Mining, 1998,2(2):121-167.