

基于 Matlab 的一种基音周期检测算法

宋建华, 朱 勇, 徐 军, 张圣贤

(黑龙江大学电子工程学院, 哈尔滨 150080)

摘 要: 在对自相关基音检测算法进行分析的基础上, 对此算法进行了深入探讨, 针对以往研究中存在的不足加以改进, 考虑到检测准确度和检测速率两方面的因素, 引入了三电平削波模块, 设计了基于 Matlab 的估计方法, 然后通过对一段具体的语音进行处理, 得到了比较准确的浊音语音的基音周期。实验证明, 该方法简单、有效。

关键词: Matlab; 基音检测; 自相关函数; 三电平削波

An algorithm for pitch detection based on Matlab

SONG Jian-hua, ZHU Yong, XU Jun, ZHANG Sheng-xian

(School of Electronic Engineering, Heilongjiang University, Harbin 150080, China)

Abstract: In this paper, some further studies have been done on the basis of analyzing the autocorrelation algorithm. For the shortcomings of the research in the past, a tri-level clipping module is adopted to improve the detection which takes the accuracy and the detection rate into account. This paper designed the estimation method based on Matlab, and then through processing a specific voice, obtained more accurate voice pitch period of the voiced sound. The experiments show that the method is simple and effective.

Key words: Matlab; pitch detection; autocorrelation function; tri-level clipping

0 引言

基音周期检测也称为基频检测(Pitch Detection), 它的目标是找出和声带振动频率完全一致的基音周期变化轨迹曲线, 或者是尽量相吻合的轨迹曲线。基音周期检测在语音信号的各个处理领域中, 如语音分析与合成、有调语音的辨意、低速率语音压缩编码、说话人识别等都是至关重要的, 它的准确性及实时性对系统起着非常关键的作用, 影响着整个系统的性能。

本文在对自相关基音检测算法基本原理进行分析的基础上, 对此算法进行了深入的探讨, 针对以往研究中存在的问题加以改进, 给出了一种方便、快捷的检测方案。综合考虑了检测准确度和检测速率两方面的因素, 引入了三电平削波模块, 然后通过对一段具体的语音信号进行滤波、采样、分帧、求短时自相关函数, 较准确地得到浊音语音信号的基音周期。

1 基于自相关函数的基音周期检测原理

采用自相关函数基音周期检测流程如图 1 所

示, 共包括六个模块: 带通滤波、采样、分帧、短时能量分析、自相关计算、基音周期估计。

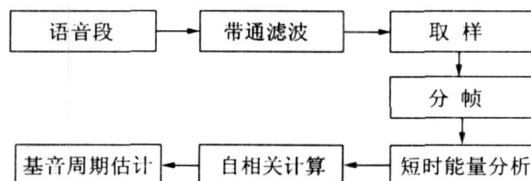


图 1 基于自相关函数的基音周期检测流程

1.1 带通滤波

由于在某些浊音中, 第一共振峰频率可能会等于或低于基音频率。此时, 如果其幅度很高, 它就可能产生一个峰值, 而当该峰值又可以同基音频率的峰值相比拟时, 从而会给基音值检测带来误差。为了克服这个困难, 最常用的方法是用一带宽为 60~900Hz 的带通滤波器对语音信号进

收稿日期: 2008-06-25

作者简介: 宋建华(1978-), 男, 讲师, 主要研究方向为语音信号处理、程控交换网理论。

行滤波,并利用滤波后的自相关函数来进行基音估计。将滤波器低端截止频率设为 60Hz,这是因为可以抑制 50Hz 电源工频的干扰;高端截止频率设为 900Hz 既可除去大部分共振峰影响,又可以当基音最高频率为 450Hz 时仍能保留其一二二次谐波。

1.2 采样和分帧

这里的“采样”是指从语音信号中选取一段样本,一般取样点数为帧长的整数倍。短时分析将语音流分为一段一段来处理,每一段就被称为“帧”。分帧时需对语音信号进行加窗操作,窗口长度的选择非常重要,窗长过短会使得分析窗内没有包含足够的数据点来进行周期判断,且短时能量变化剧烈;窗长过长,短时能量是一段长时间的平均,不但不能反映语音信号基频的细节变化部分,而且使得计算量增大。窗口长度至少要大于基音周期的两倍。一般窗长选为 120~240 点(8000Hz 采样,15ms~30ms),分帧模块主要完成将取样模块中获得的语音样值点分为若干个语音帧。

1.3 短时能量分析

语音信号 $\{x(n)\}$ 的某帧信号的短时平均能量 E_n 的定义^[1]为:

$$E_n = \sum_{m=n-N+1}^n [x(m)w(n-m)]^2 \quad (1)$$

(1) 式中, $w(n)$ 为窗函数, N 为窗长, $h(n) = w^2(n)$, 则有

$$E_n = \sum_{m=n-N+1}^n x^2(m)h(n-m) = x^2(n) * h(n) \quad (2)$$

由式(2)表明,短时平均能量 E_n 相当于将语音信号的平方通过一个单位冲激响应为 $h(n)$ 的线性滤波器的输出。实验统计发现,语音中浊音段的短时平均能量远远大于清音段的短时平均能量。因此,短时平均能量 E_n 的计算给出了区分清音段与浊音段的依据,即 $E_n(\text{浊}) > E_n(\text{清})$ 。根据 E_n 由高到低的跳变可定出浊音变为清音语音的时刻, E_n 由低向高的跳变可定出清音变为浊音语音的时刻,又由于只有浊音才有基音周期,清音的基音周期为零,这样就可判别出清/浊音。

1.4 自相关函数

对于离散的语音数字信号 $x(n)$, 它的自相关函数的定义为:

$$R(k) = \sum_{m=-N}^N x(m)x(m+k) \quad (3)$$

如果信号是随机的周期的,则这时的定义为:

$$R(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{m=-N}^N x(m)x(m+k) \quad (4)$$

上述公式表示,一个信号和延迟信号 k 点后的该信号本身的相似性。自相关函数具有性质:若信号 $x(n)$ 具有周期性,那么它的自相关函数也具有周期性,并且周期与信号 $x(n)$ 的周期相同。由于清音信号不具有周期性,它的自相关函数也没有周期, $R(k)$ 会随着 k 的增大迅速衰减。浊音信号具有准周期性,它的自相关函数 $R(k)$ 具有与 $x(n)$ 相同的周期。自相关函数法基音检测正是利用的这一性质对语音信号进行基音检测的。

2 改进后的基音周期检测及其实现

通过对自相关基音检测原理的分析,考虑到检测准确度和检测速率两个方面的因素,提出了改进后算法具体实施方案,该方案共包含 8 个模块:带通滤波、采样、分帧、短时能量分析、三电平削波、自相关计算、基音周期估计、后处理,流程框图如图 2 所示。

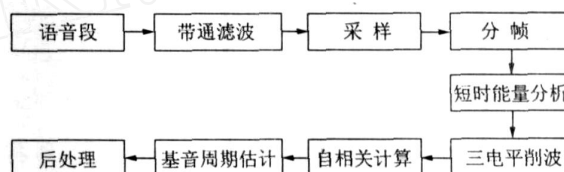


图2 改进后基音周期检测流程

如图 3 所示为原始语音信号“你好”(男音)的时域波形,图中低幅值部分是声道响应产生的谐振峰。图 4 为带通滤波降噪后的波形,很明显,谐振峰的影响被削弱了,语音波形得到了改善,而且信号能量比较集中。

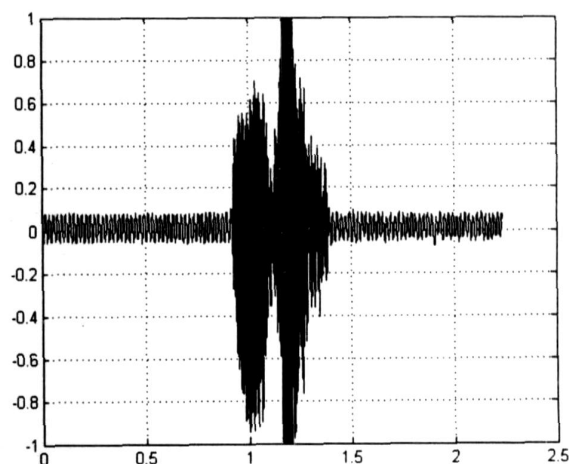


图3 “你好”原始语音波形

对滤波后信号采样,通过 Matlab 对语音“你好”的语音文件“nihao.wav”计算得知其语音长度为 2.24 秒,若采样频率为 8000Hz,共计 17920 个采样点。采样模块从采样频率为 8000Hz 的语音信号中截取 0.5 秒到 1.5 秒这一时间段,取了 8000 个样点数进行分

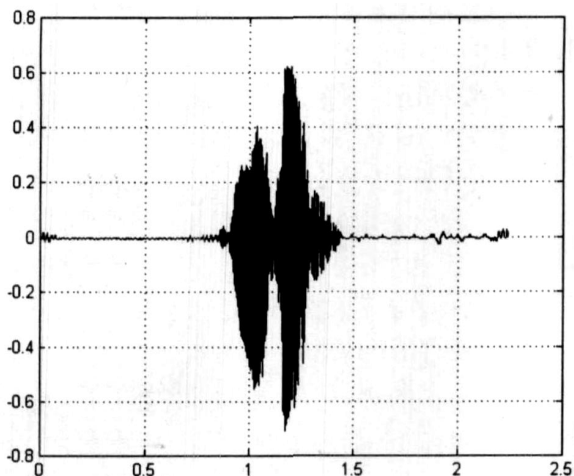


图4 滤波降噪后的波形

析。本算法中取帧长 25ms,即每帧为 200 个样点值,共分成 15 个帧。

2.1 短时能量的 Matlab 计算

求短时能量的主程序:

```
N = 200; %窗长为 200
```

```
x = wavread( 'nihao. wav', [4001 : 12000] ); %将语音信号值赋给变量 x
```

```
t = x(:,1);
```

```
h = hamming(N); %使用海明窗
```

```
En = conv(h, t, *t); %做卷积运算求短时能量
```

```
figure
```

```
plot(En)
```

```
grid
```

运行该程序后得到各采样点的短时能量如图 5 所示。

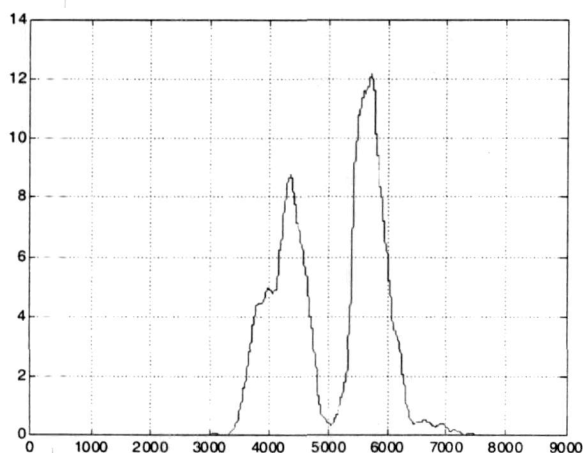


图5 语音短时能量

从图 5 中可以看出 E_n 值大的对应于浊音,而 E_n 值小的对应于清音,由此可以大致判断浊音变为清音或清音变为浊音的时刻。

2.2 三电平削波

2.2.1 中心削波法

自相关函数估计基音周期,运算量很大,其原因是乘法运算非常费时。然而从估计基音周期的角度看,短时自相关函数所包含的信息许多是多余的,真正能反映基音周期性的只是少数几个峰,而其余的大多数峰都是由于声道的谐振特性引起的。为此,可以用中心削波法^[3]来突现反映基音周期的信息,同时压缩与共无关的信息。

中心削波实质上是对语音信号作非线性处理,其形式为:

$$y(n) = C[x(n)] = \begin{cases} x(n) - C_L, & x(n) > C_L \\ 0, & |x(n)| \leq C_L \\ x(n) + C_L, & x(n) < -C_L \end{cases} \quad (5)$$

式(5)中, C_L 为削波电平, $x(n)$ 是原始语音信号, $y(n)$ 是削波后的语音信号。经中心削波处理后,由声道谐振特性产生的大部分阻尼振荡被削去,保留下来的是包含基音周期的信息。

2.2.2 三电平削波法

进一步分析可知,用自相关函数提取基音周期时,关心的是时间,也就是自相关函数峰值出现的位置,而峰值本身无关紧要。这样就可以采用三电平削波法^[4],其输入输出关系如式(6)所示。

$$y(n) = C[x(n)] = \begin{cases} 1, & x(n) > C_L \\ 0, & |x(n)| \leq C_L \\ -1, & x(n) < -C_L \end{cases} \quad (6)$$

三电平削波是一种谱平整法,即对输入的语音信号进行非线性变换。实验表明,语音信号不削波、中心削波和三电平削波的信号波形及其自相关函数经详细比较,其性能方面只有微小的差别。削波后的信号只有 1, 0, -1 三种取值,之后的自相关计算把繁复的乘法运算变为简单的组合逻辑。

2.2.3 削波电平的选取及其实现

削波电平由语音信号的峰值幅度来确定,它等于语音段最大幅度的一个固定百分数。这个门限的选择是重要的,由于在整个语音段的持续时间内,信号幅度可能有较大变化,因此,采用较高的削波电平可能会把一些有用的波形削去,然而削波电平太低,又达不到应用的效果。所以在不损失基音信息的情况下应尽可能选得高些,以达到较好的效果。在 Matlab 中用 max 函数获取该语音浊音段的波形最大幅值大为 0.71,这里削波电平选为 0.48。经 Matlab 编程削波处理后的语音波形如图 6 所示。

2.3 自相关计算

经三电平削波后,自相关函数的计算极为简单。

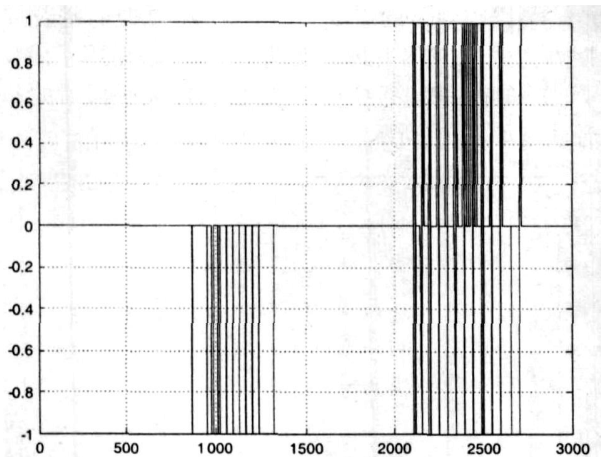


图6 削波处理后的语音波形

用 $y(n)$ 表示削波器的输出,则短时自相关函数如式(7)所示:

$$R(k) = \sum_{m=0}^{N-1-k} [y(n+m)w(m)][y(n+m+k)w(m+k)] \quad (7)$$

实现自相关计算的程序如下:

```
R=zeros(1,200); %生成一零向量用于保存结果
for k=1:200; %延迟 200 次
    sum=0;
    for m=1:200-k; %自相关函数计算
        sum=sum+x(m)*x(m+k);
    end
    R(k)=sum;
end
figure;
plot(R)
grid;
```

该浊音段共分成 15 个帧,分别进行自相关函数的计算,其中第 12 帧,如图 7 上图两黑线之间部分,下图是该帧的波形。图 8 是该帧自相关函数波形。

2.4 基音周期估计

从图 8 可以看出,第一个峰值的位置约出现在 50 的滞后点上。因为浊音语音的自相关函数具有一定的周期性,在相隔一定的取样后,自相关函数达到最大值。浊音语音的周期可用自相关函数中的第一个峰值的位置来估算。实验表明,滤波处理后的信号只含有第一共振峰以下的基波和谐波分量,明显改善了检测效果。由此估计出这段浊音的基音频率是 $8000/50 = 160\text{Hz}$ 或 $(1/160)\text{s} = 6.25\text{ms}$ 。

2.5 后处理

无论采用哪一种基音检测算法都可能产生基音检测的错误,使求得的基音周期轨迹中一个或几个

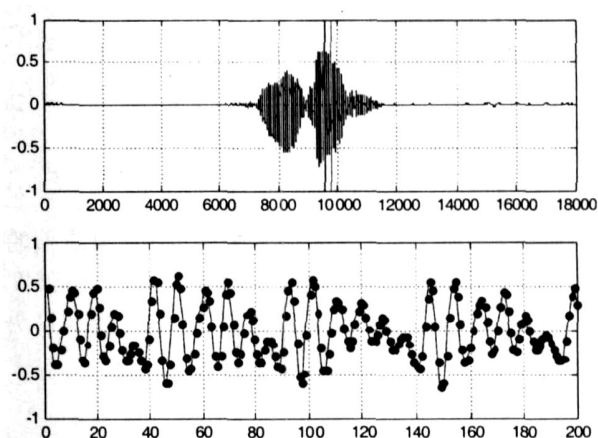


图7 第12帧波形

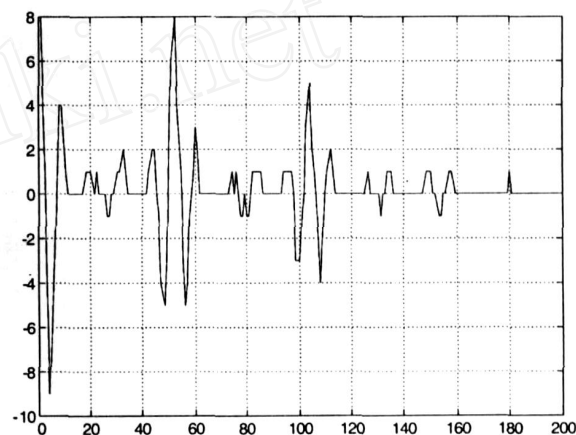


图8 第12帧自相关函数波形

基音周期的估计值偏离了正常的轨迹(通常是偏离到正常值的两倍或 1/2),并称这种偏离点为基音轨迹的“野点”。此时为了去除这些野点^[5],可以采用各种平滑算法,其中最常用的是中值平滑算法和线性平滑算法。

2.5.1 中值平滑处理

中值平滑处理的基本原理是:设 $x(n)$ 为输入信号, $y(n)$ 为中值滤波器的输出,采用一个滑动窗,输出值就是将窗外的中心移到 n 处时窗口输入样点的中值,及 n 点的左右各取 L 个样点。连同被平滑点中心共同构成一组信号采样点(共 $2L+1$ 个样值),然后将这 $(2L+1)$ 个样点按大小序列排成一排,此序列中中间者作为平滑器的输出。 L 值一般为 1 或 2,即中值平滑起的“窗口”一般套住 3 个或 5 个样点。中值平滑的优点是既可以有效地除去少量的野点,又不会破坏基音周期轨迹中两个平滑段之间的阶跃性变化。

2.5.2 线性平滑处理

线性平滑使用滑动窗进行线性滤波处理,如式(8)所示:

$$y(n) = \sum_{m=-L}^L x(n-m)w(m) \quad (8)$$

其中, $\{w(m), m = -L, -L+1, \dots, 0, 1, 2, \dots, L\}$ 为 $2L+1$ 点平滑窗, 满足式(9):

$$w(m) = 1 \quad (9)$$

线性平滑在纠正输入信号中部平滑处样点值的同时, 也使附近各点的值作了修改。所以窗的长度加大虽然可以增加平滑的效果, 但是也可能导致两个平滑段之间的阶跃模糊程度加重。

2.5.3 仿真实现

用程序:

$P = R(2:200);$

$[w, idx] = \max(P);$

求得 $k=0$ 之后的峰值的位置, 即基音周期的位置, 根据前面的计算, 15 帧的基音周期分别为 [20 41 42 42 38 39 40 43 3 5 42 48 53 55 58], 然后即可得到该段语音的基音轨迹, 如图 9 所示。

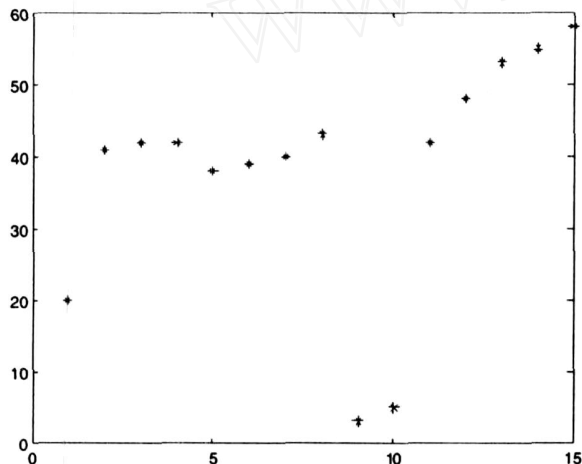


图 9 处理前的基音轨迹图

可以很明显的看到在位置 1, 9, 10 处的基音周期点为野点, 需要用平滑处理掉, 本算法使用了函数 $y = \text{medfilt1}(x)$ 进行中值滤波处理, 滤波平滑后的基音轨迹如图 10 所示。

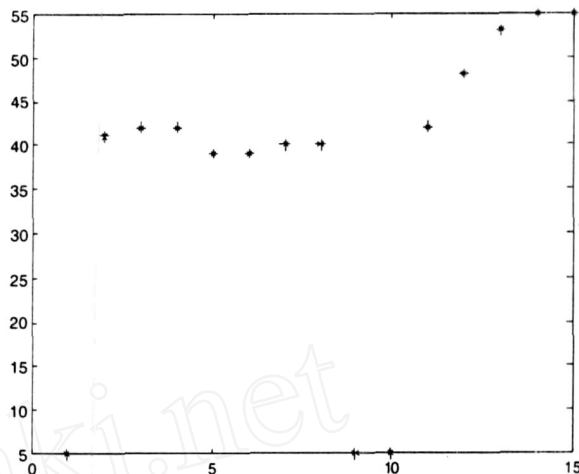


图 10 处理后的基音轨迹图

3 结束语

本文表明利用改进后自相关函数法检测到的基音周期值与目测结果基本一致。另外, 采用了三电平削波方法, 避免了自相关函数中繁杂的乘法运算, 从而提高了程序的运行速度及准确度, 本文为深入开展这方面的研究工作进行了有益的探索。

参考文献:

- [1] 韩纪庆, 张磊, 铁然. 语音信号处理[M]. 北京: 清华大学出版社, 2004.
- [2] 程佩青. 数字信号处理教程[M]. 北京: 清华大学出版社, 2007.
- [3] 段修生, 孙世宇, 王立冬. 基音检测的一种时域方法[J]. 军械工程学院学报, 1998, 10(3): 16-20.
- [4] 马祥杰, 孟相如, 张百生. 一种自相关基音检测算法[J]. 通信技术, 2003(5): 9-11.
- [5] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003.

责任编辑: 肖滨

信息天地

“星光中国芯工程”取得丰硕成果

日前, 工业和信息化部、科技部、财政部和北京市政府在人民大会堂联合主办了“星光中国芯工程”十年成果与展望报告会。工业和信息化部副部长、科技部杜占元副部长、财政部丁学东副部长、北京市苟仲文副市长、中国科协名誉主席周光召院士、国务院发展研究中心张玉台党组书记、中国科学院白春礼副院长、中国工程院潘云鹤副院长等出席会议。

报告会上宣读了中共中央政治局常委李长春同志的贺信。据了解, “星光中国芯工程”的顺利实施是坚持自主创新、以市场为导向获得的成果, 也是国家支持留学生回国创业、各级政府与企业努力促进产业发展的一次成功探索, 在机制创新方面提供了有益的经验。

姜勤俭在发言中指出, 加快集成电路产业发展, 实现核心技术的突破, 是实现信息产业由大变强的必然选择。“星

光中国芯工程”坚持自主创新, 坚持以市场为导向、以企业为主体, 是工业和信息化部组织实施的“中国芯”工程中的突出代表, 为我国集成电路产业在核心技术产业化方面起了良好的示范作用。目前产业的发展正面临着严峻的挑战, 增强自主创新能力, 是推动产业科学发展的突破口, 也是应对经济形势变化、增强产业发展后劲的有效措施。工业和信息化部作为行业主管部门, 将在营造良好环境、突破核心技术、完善产业链条、开拓应用市场等方面继续加大力度支持集成电路产业的发展, 促进信息化与工业化的融合。

在工业和信息化部、财政部、北京市政府的支持下, 邓中翰、杨晓东等一批留美归国博士于 1999 年在北京中关村创办了中星微电子公司, 启动并承担实施了“星光中国芯工程”。实施十年来, 该工程已取得 8 项核心技术突破, 获得国内外专利授权 480 项, 成功开发了两个系列的数字多媒体芯片产品并实现大批量生产, 其多媒体芯片占领了全球 60% 的市场份额, 取得良好的经济效益和社会效益, 成为我国集成电路产业自主创新的成功范例。

国务院有关部委、北京市政府、天津市政府等单位的代表, 企业界及中星微电子公司相关合作伙伴等约 200 余人出席了报告会。