参 考 文 献

[1] McAulay R. J. and Quatieri T. F., Mid-Rate Coding Based on a Sinusoidal Representation of Speech, ICASSP, 1985, 945.

[2] McAulay R. J. and Quatieri T. F., *IEEE ASSP-34-4*(1986), 744—756.

[3] Quatieri T. F. and McAulay R. J., *IEEE ASSP-34-6*(1986), 1449—1464.

[4] Atal B. S. and Remde J. R., A New Model of LPC Exitation for Producing Natural-Sounding Speech at Low Bit Rates, in Proc. Int. Conf. Acoust., Speech, Signal Processing, 1982, 614.

[5] Alemeida L. B. and Silva F. M., Variable-Frequency Synthesis: An Improved Harmonic Coding Scheme, in Proc. Int. Conf. Acoust., Speech, Signal Processing, 1984, 27.5.1—27.5.4.

一种汉语单音节基音提取与声调识别方法

赵春霞　徐近霈

（哈尔滨工业大学）

1988 年 12 月 26 日收到

本文给出一种以时域检测获取基音候选，以动态规划提取全局优化的基音轮廓，以多级逼近截取有效调型段的基音检测器，并利用基音特征参数进行声调识别的方法。

本系统可以在不作话者训练条件下，简单、快速、准确地进行基音检测和声调识别．系统对男、女话者各 1252 个不同单音节的实验结果表明，声调正识率分别为 98.9% 和 99.4%.

一、引　言

汉语声调的变化，起着构词辩意的作用．声调识别主要有三个困难：其一是基音检测的精度难以满足要求；其二是由于声调载于韵母段内，并主要寄附于韵腹上，从而很难准确判定有效调型段的头尾；其三是在调型参数选取中，如何解决调内一致性及调间可分性问题．

本文给出一种以时域检测获取基音候选，以动态规划提取全局优化的基音轮廓，以多级逼近截取有效调型段的基音检测器，并利用基音特征参数进行声调识别的方法．

二、实验系统及信号预处理

本系统是在 IBM-PC 微型机上实现的．语音信号在经过带通滤波放大后，以 10kHz 的速率进行采样．因调型段载于韵母段内，故首先在第一级语音段的粗切中，利用平均幅值和过

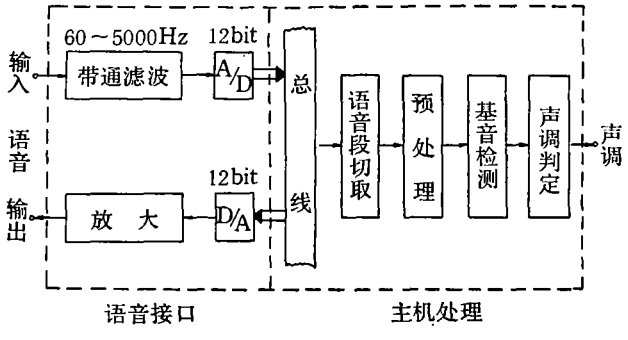

图 1　系统框图

零率，通过两次头尾判定，主要将韵母部分保留下来．随后对语音信号进行平滑滤波，以减少尖脉冲干扰的影响．图 1 为本系统框图．

三、基音检测器

基音检测是声调判定的基础，其所得基音轮廓是区分声调的主要依据． 本基音检测器，采用对时域信号正负波进行峰值筛选的方法检测基音周期，用动态规划方法在检测到的候选周期中寻求最佳基音轮廓．为提高检测精度，扩大适用范围，检测器以语音段中部周期性较强的平稳段为检测起点，检测过程是分别向两侧端点进行，其间以前一帧的检测周期 FP_{i-1} 作为当前第 i 帧基音周期 FP_i 的预测参考值．图 2 为基音检测器框图．

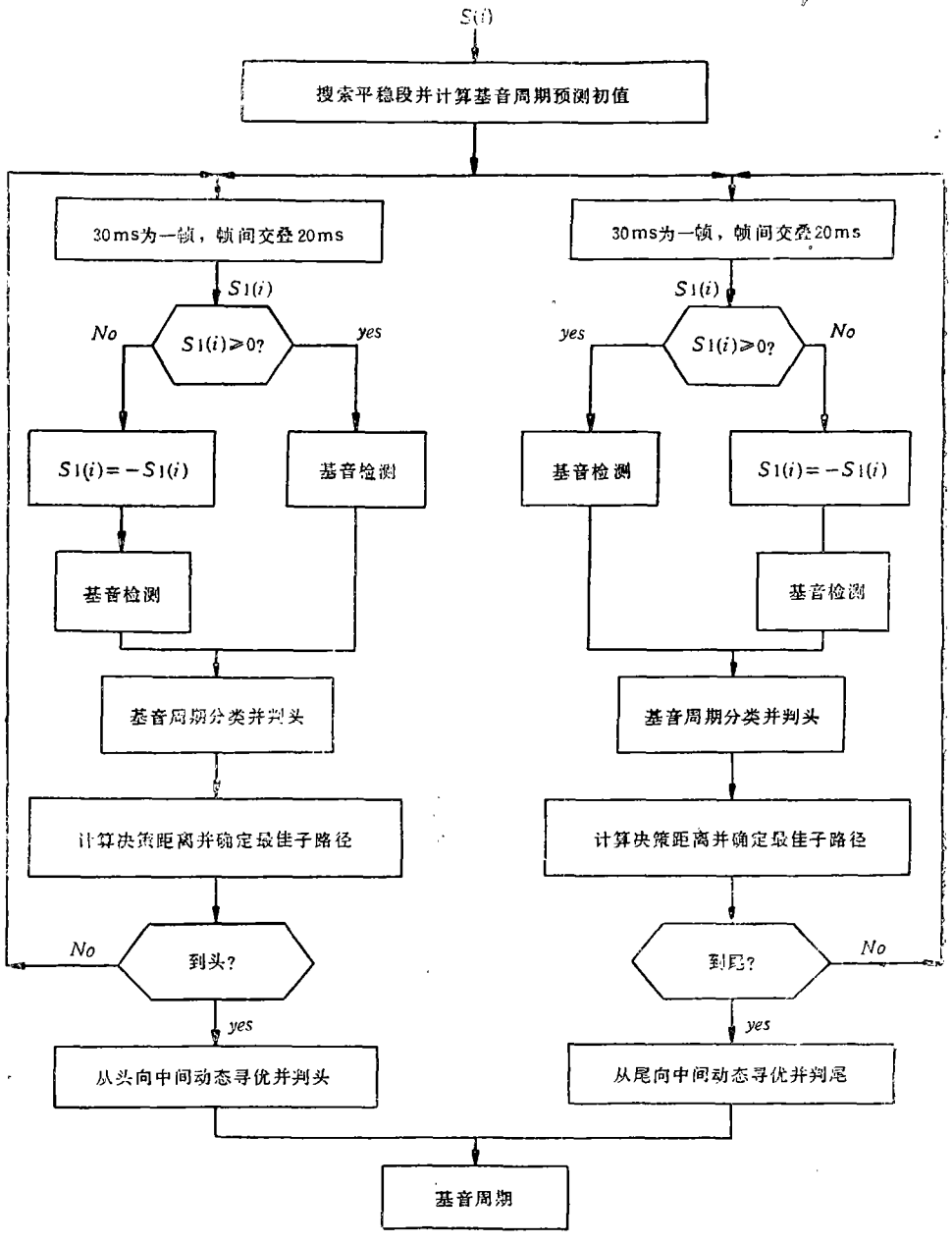

图 2　基音检测器框图

1. 平稳帧的搜索和初始 FP_0 的确定

在汉语中，韵母中部韵腹处元音的发音比较清晰、稳定，可用简单算法测得较可靠的基音估值. 所以平稳帧的搜索和初始 FP_0 的确定，是在所截语音段中间三分之一段内，利用简单的峰值筛选法进行的. 处理时以 30ms 为一帧，分别对信号的正负波进行如下处理：首先找出该帧波形中最大峰的位置，再由此峰开始分别向帧的两端，找出该帧波形中不低于最大峰 85%，且相邻峰的间隔大于一最小基音周期值 2ms (20 个采样点) 的所有峰点；当出现两个间隔不满足条件的相邻峰点时，则保留相对较高的峰，而删除较低的峰. 此后以相邻二个峰点间的采样点个数作为基音周期的估值.

设一帧中测得的正负波各基音周期估值为 $T_i, i=1,\cdots,n$，则其平均周期值 $\bar{T}$ 与方差 σ^2 可由下式求得

$$\bar{T}=\frac{\sum_{i=1}^{n}T_i}{n}$$

$$\sigma^2=\frac{\sum_{i=1}^{n}(T_i-\bar{T})^2}{n-1}=\frac{\sum_{i=1}^{n}T_i^2}{n-1}-\bar{T}^2$$

当 σ^2 越小，说明所得估值 T_i 的离散性越小，一致性越好. 故我们以处理中 σ^2 最小的一帧为平稳帧，并以此作为基音检测的起点，而该帧的平均周期 $\bar{T}$ 为 FP_0，将作为检测起点基音周期 FP_1 的预测估值.

2. 时域峰值筛选的基音检测方法

检测时，以 30ms 为一帧，相邻帧间交叠 20 ms. 因信号正负波的周期具有一致性，且可互相校正误检，故为提高检测精度，对正负波均加以检测.

在各帧内，将信号波形中最大峰作为基音脉冲的起始位置，以该峰峰值的 25%，作为此帧信号的中心削波电平. 然后以 20 个采样点为一段，找出段内高于削波电平的最大峰，作为一可能的周期峰点的候选. 这样既可减少共振峰结构的影响，又可在不损失精度的前提下，减少运算量.

在得到两个候选峰(包括最大峰)后，每再得到一个候选峰，便与前两个候选峰一起进行筛选. 如图 3，设 P_a, P_b, P_c 为相邻的三个候选峰，其中 P_c 为最新得到的候选峰，P_b' 为 P_a、P_c 峰点间连线，在 P_b 峰处的高度. 由于这一递推筛选过程在最初时是以最大峰为 P_a 的，故筛选中以 P_a 为基准，仅对 P_b、P_c 进行筛选.

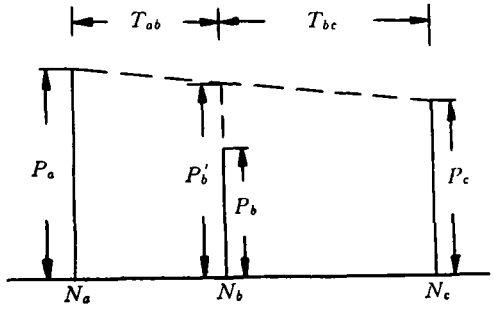

图 3 基音脉冲筛选示意图

如对第 i 帧进行处理，其筛选规则如下

(1) 按峰值筛选：删去过小的 P_b.

如 $P_b<0.65P_b'$，则将 P_b 峰删除.

(2) 按周期筛选：考虑基音周期变化的平滑性.

(a) 如 $T_{bc}\geqslant \mathrm{MIN}T_i$，将 P_b、P_c 峰均保留. 其中最小可能基音周期 $\mathrm{MIN}T_i$ 根据实验被限定在[20,55]区间之内，其值由下式确定

$$\mathrm{MIN}T_i=\begin{cases}20 & if\ FP_{i-1}-\Delta_i\leqslant 20\\ FP_{i-1}-\Delta_i & \text{其它}\\ 55 & if\ FP_{i-1}-\Delta_i\geqslant 55\end{cases}$$

$$\Delta_i=0.25\ FP_{i-1}$$

(b) 否则：当 $T_{ab}>FP_{i-1}-\Delta_i$ 时，保留 P_b 峰，删除 P_c 峰；当 $FP_{i-1}-\Delta_i<T_{ac}<FP_{i-1}+\Delta_i$，或当 $\frac{P_c}{P_b}>1.3$ 时，保留 P_c 峰，删除 P_b 峰.

(c) 否则：临时保留 P_b、P_c 峰，待后续递推筛选时再行考查.

在一帧的基音脉冲全部检测完后，以相邻脉冲间的采样点个数，作为候选周期值，以备后续处理.

3. 基音轮廓的寻优方法

为实现动态寻优，首先需要解决候选值的

分类问题．由于各帧中基音候选的检测错误基本为如下三类：半频、倍频及随机错误；其中以前两者居多．可见候选值本身具有一定的类聚性．我们将帧内各候选基音周期值由小到大排队，并根据排队后候选值的连续性，以断续点作为分类界面，从而实现了既符合聚类情况，又不固定类别数目的分类．分类后以类内所含候选值的个数作为其频度，以其均值作为该类的周期估值．从检测效果上讲，某类的频度越高，其可信度也越高；从基音周期的平滑性角度讲，两类周期估值间的差值越小，说明此间周期过渡越平滑，该路径也就越佳．为减弱个别帧决策错误对后续帧决策的影响，在动态规划中仅取路径累加距的50%．如设第 k 帧中，共得 M_k 类候选，其中第 i 类的频度为 N_i，周期估值为 T_i，$i=1,\cdots,M_k$，则第 k 帧至 $k-1$ 帧的路径累加距 d_{ij} 取为

$$d_{ij}^{k}=\frac{|T_i^k-T_j^{k-1}|}{N_i+N_j+5}+0.5D_j^{k-1}$$

$$i=1,\cdots,M_k, j=1,\cdots,M_{k-1}$$

式中 D_j^{k-1} 为 $k-1$ 帧中 j 类的最佳路径累加距离，其初值 $D_j^0=0$，而

$$D_i^k=\min\{d_{i1}^k,d_{i2}^k,\cdots,d_{iM_{k-1}}^k\}=d_{iJ}^k$$

则 $i\to J$，为第 k 帧第 i 类至第 $k-1$ 帧的最佳子路径，并将其记录下来．

为获取预测参考值 FP_k，我们找到 k 帧中累加距最小的类别 I

$$D^k=\min\{D_1^k,D_2^k,\cdots,D_{m_k}^k\}=D_I^k$$

则取：$FP_k=T_I^k$

作为搜索下一帧基音周期的预测参考值．

当搜索至端点时，由语音段端点帧中累加距最小一类的基音周期开始，可以回溯得到所有帧的基音周期值，从而得到全局最佳的基音周期轮廓．此过程可参看图4．图中的各节点代表 D_i^k，且有 $D_1^8=\min\{D_i^k\}$

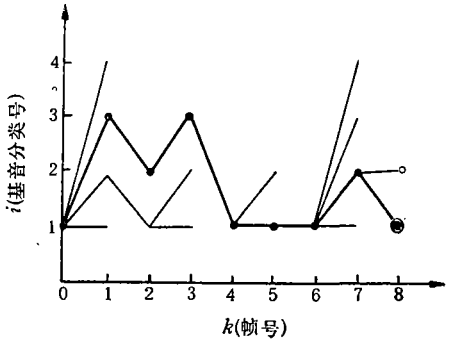

图4　动态搜索示意图

4. 基音检测中的两级头尾截取

为逼近调型段，在检测中根据基音结果的

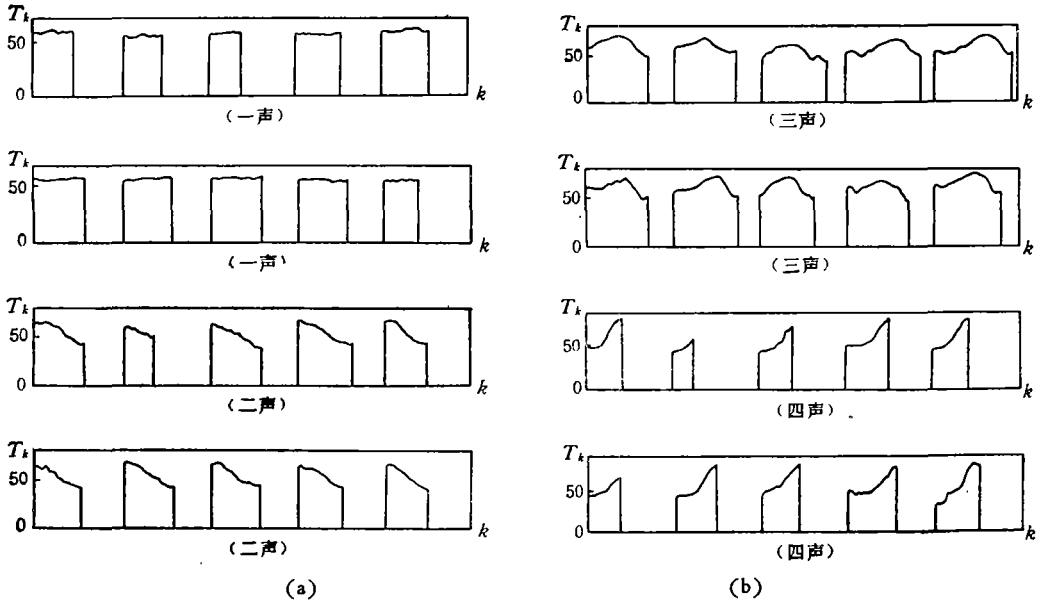

图5　基音轮廓的检测结果

平滑性，分别作了两级头尾截取.

首先在搜索接近语音段端点时，由于信号的周期性不强，检测得到的候选周期存在很大的离散性，从而分类后得到的类别数目较多.故在分类后，如候选序列个数 n 与类别数目 j 越接近，说明检测结果的一致性越差. 当

$$n/(j+1) \leqslant c_3 \quad (c_3 = \text{常数})$$

我们就认为语音段进入对声调判定无用的音段，并以此帧的前一帧作为动态寻优的回推起点.

在回推过程中又作了第二级截取. 当路径中相邻两帧间的周期差值超过给定门限 ΔT，即

$$|T^{k\pm1} - T^k| \geqslant \Delta T$$

就认为 $T^{k\pm1}$ 与 T^k 间发生断续，则将 T^k 作为新的回溯起点，而将 $T^{k\pm1}$ 至原端点的部分截去.

5. 基音检测器的定性考核结果

在图 5 中，我们给出了利用本基音检测器，对 40 个音(每种声调各 10 个音)进行检测所得到的基音轮廓线. 其中横轴为帧号 k，纵轴为相应帧的基音周期 T_k. 从图中可看出，基音轮廓较平滑，声调特性也很明显，从而为声调判定打下了良好的基础.

四、声调识别

基音轮廓反映的声调规律，从本质上讲可用图 6 所示的四条曲线来表示. 这样，采用对基音轮廓线进行一阶曲线拟合，并利用拟合参数构成的规则系统进行声调识别的方法，收到了良好的效果.

设基音周期序列为 T_i，$i = 1, \cdots, n$，拟合方程为

$$\hat{T}_i = \hat{a} + \hat{k} \cdot \hat{i},$$

利用最小二乘法即可求得

$$\hat{k} = \frac{\left[\sum_{i=1}^{n}(i-\bar{i})\cdot(T_i-\bar{T})\right]}{\sum_{i=1}^{n}(i-\bar{i})^2}$$

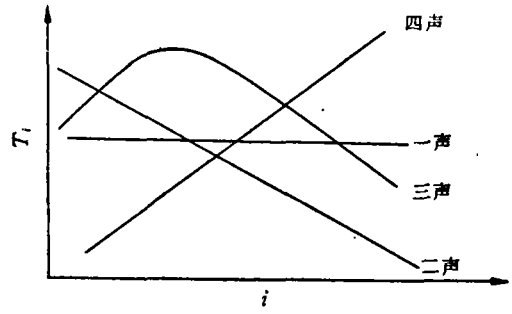

图 6　基音周期的四声轮廓

$$\hat{a} = \bar{T} - \hat{k} \cdot \bar{i}$$

其中 $\bar{T} = \sum_{i=1}^{n} T_i/n$；$\bar{i} = \sum_{i=1}^{n} i/n = n/2$.

为评价拟合结果的优劣，根据

$$\sum_{i=1}^{n}(T_i - \bar{T})^2 = \sum_{i=1}^{n}(\hat{T}_i - \bar{T})^2 + \sum_{i=1}^{n}(T_i - \hat{T}_i)^2$$

定义拟合优度 R^2 为

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{T}_i - \bar{T})^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2} = 1 - \frac{\sum_{i=1}^{n}(T_i - \hat{T}_i)^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2}$$

$$0 \leqslant R^2 \leqslant 1$$

我们利用对 40 个音(每种声调各 10 个音)进行检测所得到的基音结果，对其方差 σ^2，拟合直线斜率 k，拟合优度 R^2，及残差平方和 e^2，均作了分析计算，结果绘于图 7 中. 其中

$$\sigma^2 = \left[\sum_{i=1}^{n}(T_i - \bar{T})^2\right] \Big/ n = \frac{\sum_{i=1}^{n} T_i^2}{n} - \bar{T}^2$$

$$e^2 = \frac{\sum_{i=1}^{n}(T_i - \hat{T}_i)^2}{n}$$

图 7 中各图的横坐标中，1—10 的 10 个点代表一声的 10 个音，11—20 的 10 个点代表二声的 10 个音，21—30 的 10 个点代表三声的 10 个音，31—40 的 10 个点代表四声的 10 个音.

从结果中可看出，使用特征 σ^2，可以区分出一声. 利用特征 R^2，可将三声与二、四声区分

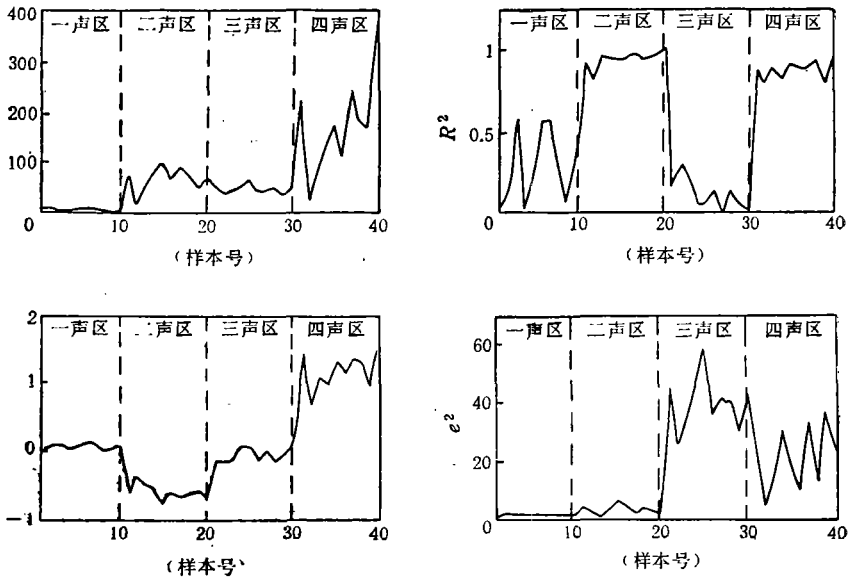

图 7　声调参数的考核结果
图中横坐标每点对应一个音

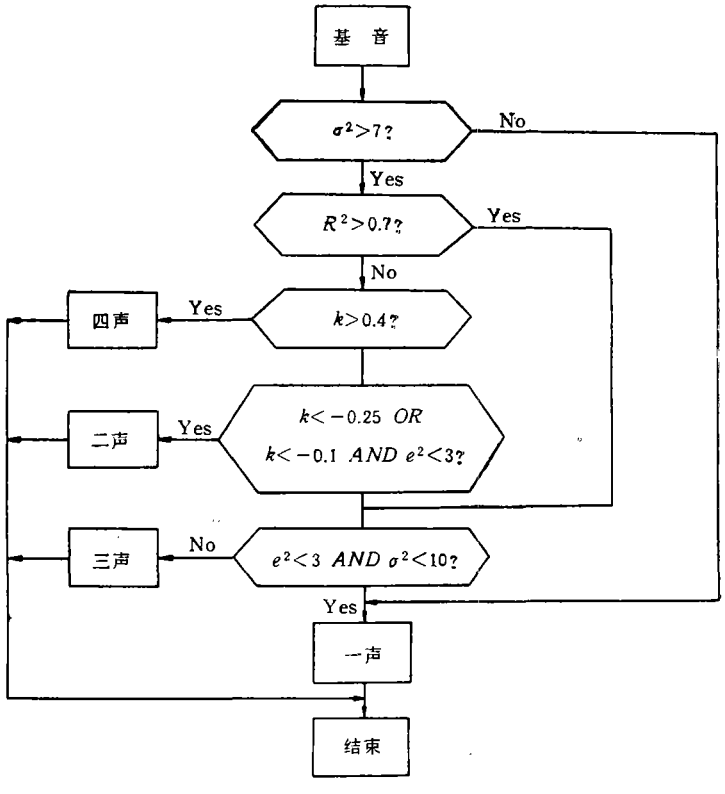

图 8　声调判定框图

开来；这是由于二、四声的基音轮廓近似于直线，而三声近似于二次曲线，使其拟合结果最差．根据拟合斜率 k，二声的斜率为负，四声的斜率为正，则又可将二、四声区分开来．此外，残差平方和 e^2 对三声也表现出一定的分离性能，这是由于三声按直线拟合，必然拟合得很

差，使其 e^2 值高于其它声调的值.

具体的声调判定过程如图 8.

五、实验结果及结论

本系统对现代汉语中出现的 1252 个不同单音节，其中含一声 336 个音，二声 253 个音，三声 315 个音，四声 348 个音，对一男一女两名话者分别录制在磁带上的语音，均作了考察.发音在保证准确的前提下，可按正常发音方式，读得较为轻松自然.

每个音识别一遍的结果，列于表 1 中. 系统对女性话者的平均正识率为 99.4%，对男性话者的平均正识率为 98.9%，总平均正识率为 99.2%.

总之，本文提出一种比较简单的时域基音检测方法，具有较宽的适用范围，不失为一种良好的基音检测器. 此外，所用声调识别的特征参数，能够较准确地反映调型特性，从而可保证声调识别的精度. 目前主要存在的问题，是头尾截取的结果尚不够理想，造成一些二、三声的混淆.

表 1 声调识别结果

发音声调	识别声调(女)					识别声调(男)				
	一	二	三	四	正识率	一	二	三	四	正识率
一	336				100%	334	1	1		99.4%
二		250	3		98.8%		249	4		98.4%
三		4	311		98.7%		6	309		98.1%
四				348	100%			2	346	99.4%

参 考 文 献

[1] Prezas D. P. and Picone Toe, ICASSP-86, Proceedings: IEEE, Vol. 1, 1986, 109—112.

[2] Cooper Leon and Cooper Mary W., 动态规划导论，国防工业出版社，1985，1—36.

[3] 张寿，于清文，计量经济学，上海交通大学出版社，1984，19—31，304—307.

100kHz—10MHz 智能超声衰减、声速综合测试仪

方 彦 军

(武汉水利电力学院)

1989 年 2 月 28 日收到

本文基于单片机技术，研制了频率范围为 100kHz—10MHz 的超声衰减、声速综合测试仪. 文中声衰减系数的测量是采用两个不同反射脉冲回波序列包络的面积之比——即“面积比值法”来求得；声速的测量是在定距离下，用时差法测出声传播时间，再换算成声速值.

一、概 述

超声衰减、声速的综合测试，在基础理论研究方面，如固体物理、分子声学、物理声学等领域上均有广泛的应用；在工程上，对材料的声学性能检验方面，如金属和非金属材料内部的探伤方面就有相当的用途. 此外，在生物医疗方