[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G10L 17/00

[12] 发明专利申请公开说明书

[21] 申请号 200410000055.9

[43] 公开日 2004 年 12 月 22 日

[11] 公开号 CN 1556522A

[22] 申请日 2004.1.6

[21] 申请号 200410000055.9

[71] 申请人 中国人民解放军保密委员会技术安全研究所

地址 100076 北京市 83 号信箱 351 分箱

[72] 发明人 邢方亮 潘云生 张新民 吴永军 李 森 李宏伟 雷 霆

[74] 专利代理机构 北京中创阳光知识产权代理有限责任公司

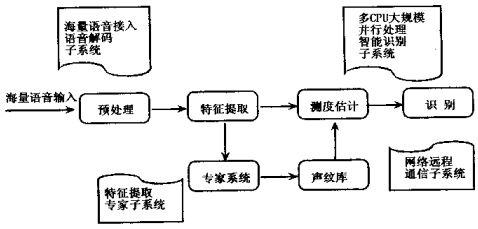
代理人 尹振启

权利要求书 2 页 说明书 10 页 附图 7 页

[54] 发明名称 电话信道说话人声纹识别系统

[57] 摘要

本发明公开了一种电话信道说话人声纹识别系统，包括；一语音接入、语音解码、预处理子系统，该子系统负责将话音数据传入系统存储系统，并进行各种话音编码格式的转换；一神经网络声纹提取专家子系统，该子系统针对目标人物利用系统提供的专家系统从已知的话音数据中提取与说话人相关的声纹参数；一网络远程通信子系统，该子系统连接声纹提取系统和声纹识别系统，为二者的通信提供保障；以及一多 CPU 大规模并行处理智能识别子系统，该子系统实现海量话音的实时处理，采用了多 CPU、多进程、多机联网并行计算。 本发明不仅具有重要的军事价值，也可用于公安侦察、司法鉴定、海关和民航控守以及银行身份认证，具有广泛的推广、应用前景。



ISSN1008-4274

1、一种电话信道说话人声纹识别系统，该系统包括：

一语音接入、语音解码、预处理子系统，该子系统负责将业务一线海量的话音数据传入系统的话音存储系统，并进行各种话音编码格式的转换；

一神经网络声纹提取专家子系统，该子系统针对目标人物利用系统提供的专家系统从已知的话音数据中提取与说话人相关的声纹参数；

一网络远程通信子系统，该子系统连接声纹提取系统和声纹识别系统，为二者的通信提供保障，同时由于该系统采用网络并行计算，计算过程所有的数据通信过程也共用该网络远程通信子系统；以及

一多 CPU 大规模并行处理智能识别子系统，该子系统实现海量话音的实时处理，采用了多 CPU、多进程、多机联网并行计算。

2、根据权利要求 1 所述的电话信道说话人声纹识别系统，其特征在于，还包括一声纹浏览器，用于对学习好的声纹进行详细的观察、挑选、拷贝、复制，从而达到优化声纹参数。

3、根据权利要求 1 所述的电话信道说话人声纹识别系统，其特征在于，该神经网络声纹提取专家子系统可对声纹参数进行优化。

4、根据根据权利要求 3 所述的电话信道说话人声纹识别系统，其特征在于，该神经网络声纹提取专家子系统进一步包括：

一导入模块，用于导入神经网络权值矩阵；

一计算模块，计算神经网络权值微分；

一判断模块，判断相邻权值微分符号是否大于等于 0；

一累加模块，若判断模块判断出相邻权值微分符号大于等于 0，则该累加模块对神经网络权值微分进行累加；以及

一输出模块，输出希望得到的前 n 名最佳声纹。

5、根据根据权利要求 1 所述的电话信道说话人声纹识别系统，其特征在于，该多 CPU 大规模并行处理智能识别子系统进一步包括：

一过滤模块，用于对音频帧进行能量过滤；

一判断模块，判断过虑后的能量是否符号要求，若不符号要求在扔掉；以及

一识别模块，采用 LPCC 一阶矩、及二阶矩的综合匹配方法识别目标，若匹配成功则正确识别了音频数据。

6、根据权利要求 1 所述的电话信道说话人声纹识别系统，其特征在于，该神经网络声纹提取专家子系统和该多 CPU 大规模并行处理智能识别子系统采用了一自组织竞争神经网络。

7、根据权利要求 7 所述的电话信道说话人声纹识别系统，其特征在于，该多 CPU 大规模并行处理智能识别子系统还采用了一语言信号声管模型。

8、根据权利要求 7 所述的电话信道说话人声纹识别系统，其特征在于，该自组织竞争神经网络根据输入空间样本的分布情况对输入向量进行聚类，以

使获胜神经元的网络权值收敛于说话人的语音特征。

9、 根据权利要求 1 所述的电话信道说话人声纹识别系统，其特征在于，该系统使用了开放式语音编码接口以实现全面兼容话音接收设备。

5

电话信道说话人声纹识别系统

技术领域

本发明涉及一种声纹识别系统，尤其是涉及一种电话信道说话人声纹识别系统。

背景技术

声纹识别技术，在话音信号处理领域具有重要意义。近年来，海量信息的智能化处理研究，对于数据业务已取得了显著进展；而对于话音业务的处理，还停留在信令引导下的模式，对人的依赖程度很大，无论是在效率上还是在效益上都远不能满足实际工作的需要。因此，开发大容量、高精度，高速度的语音信号自动识别处理系统，就成为当前话务自动化处理工作中亟待解决的课题之一。

目前，国际上的声纹识别研究主要建立在声学特征统计规律基础上，其识别性能难以达到实用要求。为了解决这一国际难题，我们大胆提出了用自组织竞争神经网络，自适应提取说话人声纹特征的思想，其独到之处是将人的发声特征与听觉仿生学原理作为有机整体进行系统建模。在研制过程中，我们还首次发现，人的声纹分布在不同的能量层次上，且在每个层次上呈束状分布。基于上述重要发现，分别提炼出每一束声纹所涵盖的特征参数并汇总成簇，全方位刻画出说话人语音特征，从而在多维向量空间锁定目标。为了攻克海量话音处理的技术瓶颈，我们又创造性地构建了单机多 CPU、多进程并行；联合千兆网多机、多层并行的先进并行计算体系，实现了海量话音的实时处理。系统的开放式语音编码接口，全面兼容现有电话编码标准，还可不断扩展到新的电话接收装备。

发明内容

针对上面的描述，本发明的一个目的就是提出了一种电话信道说话人识别系统，该系统包括：一种电话信道说话人声纹识别系统，该系统包括：一语音接入、语音解码、预处理子系统，该子系统负责将业务一线海量的话音数据传入系统的话音存储系统，并进行各种话音编码格式的转换；一神经网络声纹提取专家子系统，该子系统针对目标人物利用系统提供的专家系统从已知的话音数据中提取与说话人相关的声纹参数；一网络远程通信子系统，该子系统连接声纹提取系统和声纹识别系统，为二者的通信提供保障，同时由于该系统采用网络并行计算，计算过程所有的数据通信过程也共用该网络远程通信子系统；以及一多 CPU 大规模并行处理智能识别子系统，该子系

统实现海量话音的实时处理，采用了多 CPU、多进程、多机联网并行计算。

根据本发明的另一方面，进一步包括一声纹浏览器，用于对学习好的声纹进行详细的观察、挑选、拷贝、复制，从而达到优化声纹参数。

根据本发明的又一方面，其中该神经网络声纹提取专家子系统可对声纹参数进行优化。

根据本发明的又一方面，其中该神经网络声纹提取专家子系统进一步包括：一导入模块，用于导入神经网络权值矩阵；一计算模块，计算神经网络权值微分；一判断模块，判断相邻权值微分符号是否大于等于 0；一累加模块，若判断模块判断出相邻权值微分符号大于等于 0，则该累加模块对神经网络权值微分进行累加；以及一输出模块，输出希望得到的前 n 名最佳声纹。

根据本发明的又一方面，其中该多 CPU 大规模并行处理智能识别子系统进一步包括：一过滤模块，用于对音频帧进行能量过滤；一判断模块，判断过虑后的能量是否符号要求，若不符号要求在扔掉；以及一识别模块，采用 LPCC 一阶矩、及二阶矩的综合匹配方法识别目标，若匹配成功则正确识别了音频数据。

根据本发明的又一方面，其中该神经网络声纹提取专家子系统和该多 CPU 大规模并行处理智能识别子系统采用了一自组织竞争神经网络。

根据本发明的又一方面，其中该多 CPU 大规模并行处理智能识别子系统还采用了一语言信号声管模型。

根据本发明的又一方面，其中该自组织竞争神经网络根据输入空间样本的分布情况对输入向量进行聚类，以使获胜神经元的网络权值收敛于说话人的语音特征。

根据本发明的又一方面，其中该系统使用了开放式语音编码接口以实现全面兼容话音接收设备。

附图说明

图 1 给出了该系统的原理结构图；
图 2 给出了系统硬件体系结构图；
图 3 给出了专家可以介入声纹分析示意图；
图 4 给出了声纹的可视化编辑示意图；
图 5 给出了非实时话音声纹识别示意图；
图 6 给出了实时话音声纹识别示意图；
图 7 给出了本系统所构造的自组织竞争神经网络的数学模型；
图 8 给出了 SOFM 随机网格拓扑关系；
图 9 给出了语音信号声管模型；
图 10 给出了多声纹智能识别流程图；
图 11 给出了声纹参数优化流程图；
图 12 给出了话音接口流程图。

具体实施方式

对下面我们参考附图，对本发明的实施例进行详细的说明。

首先说明一下本发明的技术指标：

- ✧ 识别准确率 PCM 编码的话音信号平均识别率≥95%
  压缩编码的话音信号平均识别率≥90%
- ✧ 识别误识率 PCM 编码的的话音信号平均误识率≤10%
  压缩编码的话音信号平均误识率≤15%
- ✧ 识别速率 ≥1GB/小时（控守 50 个目标）（可扩）
- ✧ 声纹训练： 训练样本话音长度≥30 秒
  1～16 条声纹自适应提取
  声纹模型可视化编辑
- ✧ 识别方式： （1）开集；
  （2）文本无关；
  （3）实时侦控并报警；
  （4）非实时海量存储数据筛选；
- ✧ 识别精度选择： 粗选、精选、确认三种
- ✧ 数据输入接口： 千兆网、160Mbyte/s SICS 总线
- ✧ 声道 单(包括合路话音)、双声道音频
- ✧ 信道 有线话音信道和无线话音信道
- ✧ 语音编码格式： （表 1）

语音编码格式

音频编码类型	来源	编码参数
CCITT G.711 A-Law and u-Law	CCITT G.711	64kbs,8000Hz,8Bit
GSM 6.10	“欧洲电讯标准协会”的 6.10 标准	13kb/s
ADPCM	CCITT	32kb/s, 8000Hz,4Bit
G.723.1	基于 ITU 的 G.723 协议，通常用于 IP 电话线路	5.3/6.3kbs,8000Hz,16Bit
G.728	基于 ITU 的 G.728 协议，通常用于 IP 电话线路	16kbs,8000Hz,16Bit
G.729A G.729A	基于 ITU 的 G.729 协议，通常用于 IP 电话线路	8kbs,8000Hz,16Bit 8kbs,8000Hz,16Bit
LPC-10E	参数编码	2.4kbs,8000Hz,12Bit

下面参考图 1 和图 2，对该本发明的声纹识别系统的组成进行详细的说明。该系统主要由：语音接入、语音解码、预处理子系统；神经网络声纹提取专家子系统；网络远程通信子系统；和多 CPU 大规模并行处理智能识别子系统四部分组成。第一部分负责将业务一线海量的话音数据传入系统的话音存储系统，并进行各种话音编码格式的转换；第二部分则针对目标人物利用系统提供的专家系统从已知的话音数据中提取与说话人相关的声纹参数；第三部分连接声纹提取系统和声纹识别系统，为二者的通信提供保障，同时由于系统采用网络并行计算，计算过程所有的数据通信过程也共用这一部分；第四部分实现海量话音的实时处理，采用了多 CPU、多进程、多机联网并行计算，攻克了海量话音实时处理的难题。

参考图 2，对该系统的硬件结构进行详细的描述。如图 2 所示，由于前端的话音接收设备种类繁多，接口复杂，但又大多拥有话音存储服务器；因此为了尽量多的兼容这些设备，同时尽量小的改动这些设备，我们采用了千兆宽带网直接从话音存储服务器复制数据的办法，并同时进行语音解码，还原音频波形数据用于特征提取和识别。这样既不会影响业务人员工作，又能向他们提供需要的说话人目标。在处理海量语音数据时，一个最突出的问题就是实时性。为了满足几万甚至几十万路话音的处理工作，我们采用了多 CPU、多进程、多机并行工作模式，并开发了专门的并行算法予以实现。系统的可扩展性强，可适用于不同应用的配置需求。现系统具体配置如下，CPU: 致强 2.8GHZ×6； 内存：2GB×3； 盘阵：1TB。

该系统的工作过程主要包括两个方面：声纹智能分析和声纹高速智能识别。该声纹智能分析包括声纹智能化自适应提取和声纹的可视化编辑。该声纹高速智能识别包括非实时海量话音声纹识别和实时海量话音声纹识别。

下面参考图 3 至图 6 分别对如上所述的该系统的各个工作过程进行详细的描述。

如图 3 所示，给出了声纹智能化自适应提取。语音信号是一种典型的时变信号，然而如果把观察时间缩短到十毫秒或几十毫秒，则我们可以发现它们是近似平稳的。这是由于我们的发音器官不可能是毫无规律的快速变化，因此可以说语音信号是短时平稳的。这里的主要难度在于能不能找到、发现可以唯一标识某人或某物的这组参数，而且这组参数还要是在开集条件下（不限定识别对象），不限定文本的，这就不容易了。系统改进并构造了恰当的自组织竞争神经网络作为专家系统的核心，可以自适应的为用户提取说话人的语音特征（图 3 红线）。当然为了增加系统的鲁棒性和抗噪音能力，在系统中还保留了专家介入的接口。图 3 中，右侧红线部分是系统通过专家系统推荐出来的声纹参数。系统首次提出了神经网络分析多声纹的思想及模型，从图 3 中可以观察所有由神经网络找出来的声纹（以不同颜色表示）。上图中如果有些声纹不在系统推荐之列，但是从有经验的语音信号分析人员角度看又是十分有用的声纹，就可以用手工追加声纹的功能，将选中的绿色声纹追加到已有的声纹文件中去。通过声纹智能化自适应提取，用户可以获

得理想的说话人声纹参数，并建立属于自己的声纹库。值得一提的是，声纹提取子系统的运行并不影响识别子系统，二者可以同步进行，互不影响。

如图 4 所示，给出了声纹的可视化编辑。因为随着时间的推移，人的声纹并非完全不变；说话人语音样本学习时间越长，收集的语音相隔时间更久的话，训练的效果就越好；所以，为了更好的适应工作需要，我们还提供了方便的声纹浏览器。如图 4 所示，对学习好的声纹还可以进行详细的观察、挑选、拷贝、复制，从而达到优化声纹参数，进一步提高识别率，降低误识率的目的。

如图 5 所示，给出了非实时话音声纹识别。非实时话音查询（图 5 ）的意义在于可以对已经存储的海量话音进行反查，此时控守的目标人物较多，因而也需要更多的计算时间。由于是非实时工作模式，工作人员对时间要求并不苛刻，只是想知道在存储过的话音中是否存在想要的说话人对象，而该对象的号码又不在电话号码记载库中。这对发现其新的电话号码，发现不在号码搜索范围的话音具有明显的现实意义。在非实时模式下工作，系统计算峰值可达到 4GB/小时（同时控守 50 个目标），而且这一性能可以随设备扩展，呈线性增长。在筛选过程中，系统提供粗选、精选、确认三个识别精度；分别适应不同的工作情况。识别阶段系统可以根据话音与声纹的相似性，对可疑目标评分，并进行排序；分值越高者，是目标的可能性越大。

如图 6 所示，给出了实时话音声纹识别。系统还可以以实时工作模式（图 6 ）同时识别多个目标，当目标出现时进行提示。这种工作模式最大的优势就在于对焦点人物的识别不是靠电话号码，而是根据其发音特征，从而实现了以声辨人，实时发现重要人物及其经常变更的通信手段信息。这里主要的技术挑战是系统的分析、识别速度要大于或等于实际话路的接收速度。目前的我们的声纹识别系统能够在日接收话音 10 万路的条件下实时监控 5 个目标。由于系统的并行计算模式可以随硬件累加而性能呈线性增长，因而处理能力还可以按需扩展。

本发明的关键在于：如何从短时间的说话人语音信息中有效的提取具备强抗噪声能力、稳健的语音特征——即声纹；并做到声纹识别的开集性（即侦控目标的数目不受限制）和文本无关性（即识别结果与说话人的谈话内容无关）；在电话语音中各种语音的信道变化复杂，导致如何提取声纹，怎样提取才能保证识别的准确性，有了声纹如何比对，怎样处理和分离海量音频数据中的其他不相关信息，都成为本发明的难点；最后如何实现海量语音数据实时监控、处理也是不可回避的技术难题。

本发明首次应用自组织竞争人工神经网络技术自适应提取声纹特征实现了特定人声纹的高精度识别。声纹识别的研究要从两方面向人的器官及大脑的分析过程学习和借鉴。一方面，是人的发音器官如何发出不同的语音，只有利用其发音机理，在特征提取上才会有长足的进展；另一方面，是人的耳朵及其神经系统如何接收处理语音信息，并由大脑分析得到需要的结论。针对上述技术难点，本课题在声纹提取部分的技术路线是改进并构造恰当的

自组织竞争神经网络（SOFM－Self Organizing Feature Map）使之适用于电话信道中海量语音信息的声纹采集。

说话人识别的问题最终可以归结为一个模式归类的问题。为了模拟人的听觉过程，可以效仿人的生物神经元结构，构造恰当得人工神经网络，一层一层的过滤人讲话过程中的特征信息，从而使其收敛于说话人的特征。事实上在这里，人工神经网络完成的就是人在适应特定人讲话时对其特征的聚类学习过程。图 7 给出了本系统所构造的自组织竞争神经网络的数学模型。

$P^i=[p_1^i,p_2^i,\Lambda\ \Lambda\ ,p_R^i]$ 为第 i 个输入向量

网络权值矩阵：$IW^{1,1}=\begin{bmatrix} w_{1,1} & w_{1,2} & \Lambda & w_{1,R1} \\ w_{2,1} & w_{2,2} & \Lambda & w_{2,R1} \\ M & M & O & M \\ w_{S1,1} & w_{S1,2} & \Lambda & w_{S1,R1} \end{bmatrix}$

神经网络输入向量维数 R，具有 S1 个神经元。$IW^{1,1}$ 中第 j 行（$j=1,2,\Lambda\ S1$）元素就是第 j 个神经元与输入层之间的连接权值。

如果 $\|ndist\|$ 取欧拉距离，则第 j 个神经元的网络输入为：

$$n_1^j=-\sqrt{\sum_k^R\left(IW_{j,k}^{1,1}-P_k^i\right)^2}+b_j^1$$

b_j^1 为第 j 个神经元的偏差值。则:

$$n_1=\left[n_1^j\right]\qquad j=1,2,3\Lambda\ S_1$$

将上式以矩阵形式用计算机语言可以表达成：

$$n_1=-sqrt(sum(w-p)\hat{}2)+b_1$$

可以看到 n_1 实质上度量了个神经元权值向量与输入向量的贴近度。

竞争层神经元的激活函数对网络输入 n_1 作出的响应，使获胜神经元（与输入向量 Pi 最接近，从而距离绝对值最小）输出值为 1，其余神经元输出值为 0。当取 $b_1=0$，则最接近输入向量 Pi 的神经元具有绝对值最小的负值输入，从而可以赢得竞争胜利，输出 1。

自组织竞争神经网络学习的目的就是使获胜的神经元对经常出现的相近模式敏感进而成为这类模式的中心。本文系统就要利用这一点，使获胜神经元的网络权值收敛于说话人的语音特征；当然，真正做到这一点还要对网络的构造、训练算法做进一步的研究和改进。只有这样当这些相近的模式再次出现时，训练好的网络才能适当作出反应，找到这一类模式。在竞争学习中，学习算法主要是针对获胜神经元的。获胜神经元（输出值为 1，其权向量在当前与输入模式距离是所有神经元中最近者）的权向量（整个神经网络中的一行）可以由 Kohonen 学习法则进行调整。假定第 i 个神经元获胜，则第 i 个行权矩阵可如下进行调整；

$$iIW^{1,1}(q)=iIW^{1,1}(q-1)+\alpha(P(q)-iIW^{1,1}(q-1))$$

式中$iIW^{1,1}$代表输入层到第一层神经元的输入（Input Weight）权矩阵的第 i 行（获胜神经元所在行）权值矩阵；q 代表第 q 次学习；α为学习速率，且 $0<\alpha<1$，用于调整学习进度；P(q)为第 q 次输入模式；$iIW^{1,1}(q-1)$则为上次 i 神经元获胜时修改后的权矩阵。由上式可以看出与输入模式最相近的神经元权值被调整，以便更接近该类模式。其结果就是，当下一次该类模式再出现时该神经元就更加容易获胜，而当不相近模式出现时，它则更不易获胜，从而导致该神经元只对一类特有模式敏感，而对其它模式迟钝。当有更多的输入提供时，输入空间的模式是未知的，但可以由上式推论，只要所构造的神经网络合适，神经元足够则竞争层上的每个神经元都会对一群相近的（group）模式敏感，形成它们的中心（模式类的代表）。最终每一组相似输入的聚类都对应于一个输出为 1 的神经元，该神经元对其它聚类中的模式则不敏感输出为 0。这样竞争层的神经网络就有效地对输入模式空间进行了聚类。本文系统所设计的神经网络拓扑结构如图 8 所示。

这样，我们所构造的自组织映射神经网络就可以根据输入空间样本的分布情况对输入向量进行聚类。它既学习输入向量的分布也同时学习其拓扑关系；即，它能够无导师的学习大量样本空间的模式关系，并发现有用的模式(能代表说话人的语音特征)。这正是本系统所找到的说话人特征具有强稳健性（能够适应复杂的信道变化及低速率编码语音）和高精度识别的根本原因，也是系统识别性能取得突破的关键。

本系统的海量语音计算的实现基础是系统的多 CPU、多进程、多机联网并行计算子系统。该子系统能自适应的将不同的计算任务最佳的分配到各个 PE(处理单元)上去，使得各个 PE 都能充分发挥性能，并在同一时间把计算结果回传到主控计算机上，用于回显给用户。这一点成为系统整体设计迈向实用化的一个不可忽视的关键技术。

该子系统的硬件体系结构见图 2，逻辑上总体设计原则是：（1）在存储方面，各个 PE 通过千兆网共享外存储器（存储海量话音及计算结果），同时独享内存储器（存储声纹识别计算过程的临时数据）；（2）计算方面，网间 PE 并行处理通过千兆网通信，PE 内部同共享内存和 CPU 资源实现通信和并行处理。在软件体系结构上，主控服务器首先触发识别主进程，从服务器登陆主机，计算分成网络并行和单机并行两个层次同时进行。

在进程调度上，采用了多进程和多线程并行计算模式；将系统分成一个主进程和多个识别子进程同时运行，由主进程分配并调度子进程的识别任务。并在主进程中创建多个不同功能的线程，分别完成任务发现、任务分配调度、任务结果接收等功能。在具体实现过程中，我们还根据不同的功能要求采用不同的任务分配机制：例如：对于声纹识别实时工作模块，我们只是在语音数据刚从线路上采集下来时进行识别，所以这就决定了在瞬时时间内，我们只是针对少量话音进行识别，而此时待识别的特定人却相对较多（3－5 个），针对这种情况，主进程按识别人分配识别任务；而对于声纹识别

控守模块，程序将在积累下来的上万份话音文件中对指定的几十个人进行识别，为了更好地平衡多 CPU 的资源利用率，设计上以语音文件为准，让每个进程识别的特定人相同，而语音文件平均分配。

这种多进程和多线程并行计算模式，使得进程之间的通信及协调控制相当困难。为此在算法设计上采用了管道机制来实现进程之间的通信，采用了信号量和临界区等多种方式对同进程下线程进行协调。通过这些关键技术，我们最终将所有 CPU 资源利用率基本保持在 90%以上，而且程序非常稳定，很好地完成了声纹识别侦控过程中的海量语音处理任务，其操作界面如图 5、图 6 所示。整个并行计算系统可以根据不同的阵地容量按需配置多 CPU 计算机；更大容量的计算只需要象“搭积木”一样累加 CPU、存储器等硬件设备即可解决问题。

本发明采用语音特征相关分析与多声纹智能识别技术，突破了声纹识别开集和文本无关的技术难点。人的发音器官包括肺、气管、喉（包括声带）、咽、鼻、和口等，这些器官共同形成一个复杂的管道。其中声带的开闭使气流形成一系列脉冲，这些脉冲再经过咽、鼻、和口腔所构成的声道的变换就形成了语音信号。那么声道数学参数的不同，就导致了不同的人有不同的说话特征。全极点线性预测模型（LPC）可以对声管模型进行很好的描述，这里信号的激励源是由肺部气流的冲击引起的，声带的周期振动和不振动分别对应元音和清音。声道可以用若干段前后连接的声管进行模拟，每一段声管对应一个 LPC 模型的极点。一般 1 2～1 6 个参数就可以比较清晰的描述语音信号了。图 9 出了基于 LPC 的语音信号声管模型。

线性预测分析的基本思想是：用过去 P 个样本点值来预测现在或未来的样点值：$\hat{s}(n)=\sum_{i=1}^{p} ai * s(n-i)$

预测误差 $\varepsilon(n)$ 为：

$$\varepsilon(n)=s(n)-\hat{s}(n)=s(n)-\sum_{i=1}^{p} ai * s(n-i)$$

这样就可以通过在某个准则下使预测误差 $\varepsilon(n)$ 达到最小值（一般采用最小均方误差）的方法来决定唯一的一组线性预测系数 $ai \quad (i=1,2,...p)$。

定义语音帧的 $s(n)$ 的自相关函数：

$$Rn(j)=\sum_{n=j}^{N-1} Sn(n)Sn(n-j) \qquad j=1,...,p$$

则可以将求解 LPC 参数的方程表达如下：

$$\begin{bmatrix} Rn(0) & Rn(1) & \Lambda & Rn(p-1) \\ Rn(1) & Rn(0) & \Lambda & Rn(p-2) \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ Rn(p-1) & Rn(p-2) & \Lambda & Rn(0) \end{bmatrix} \begin{bmatrix} a1 \\ a2 \\ \mathrm{M} \\ ap \end{bmatrix} = \begin{bmatrix} Rn(1) \\ Rn(2) \\ \mathrm{M} \\ Rn(p) \end{bmatrix}$$

可以采用莱文逊—杜宾（Levinson—Durbin）递推算法求解。在系统的识别过程中还采用了 LPCC 参数的一阶微分作为二次匹配来保证识别的准确性。声管模型与人工神经网络的结合使用，为系统的高识别率和低误识率打下了坚实的基础。

另一方面系统在识别方法上作了大胆创新，进一步提高了系统整体的识别性能。系统认为人的语音在电话音频中占主要能量成分，因此在识别上首先对音频帧进行能量过滤，选择那些最能表现说话人的音频帧进行处理。这一技术可以极大地提高系统抗噪音能力。同时，系统分别又采用 LPCC 一阶矩、及二阶矩的综合匹配方法识别目标，收到了良好的效果。其流程如图 10 所示。

因此本系统的多 CPU 大规模并行处理智能识别子系统进一步包括：一过滤模块，用于对音频帧进行能量过滤；一判断模块，判断过虑后的能量是否符号要求，若不符号要求在扔掉；一识别模块，采用 LPCC 一阶矩、及二阶矩的综合匹配方法识别目标，若匹配成功则正确识别了音频数据。

本发明采用了无导师声纹学习与人工干预相结合的方法进行声纹参数选择有效的降低了误识率。由自组织竞争神经网络构成的专家系统聚类得到的声纹参数很好的刻画了人的发音特征，但这只是发现了说话人发声特征空间的拓扑分布情况，哪些区域的分布才是稳健的特征还不知道。只有设计合适的算法才能真正解决问题。基于声管模型越曲折、特征越明显的假设，其如图 11 所示。

因此本系统的神经网络声纹提取专家子系统进一步包括：一导入模块，用于导入神经网络权值矩阵；一计算模块，计算神经网络权值微分；一判断模块，判断相邻权值微分符号是否大于等于 0；一累加模块，若判断模块判断出相邻权值微分符号大于等于 0，则该累加模块对神经网络权值微分进行累加；一输出模块，输出希望得到的前 n 名最佳声纹。

上述流程可以有效计算并向用户推荐合理的声纹曲线，同时系统还保留了专家手工添加声纹的权利。从有经验的语音信号分析人员角度看又十分有用的声纹，可以选择手工追加声纹的功能，将选中的声纹追加到已有的声纹文件中去。对选好的声纹用户还可以在声纹浏览器中详细的观察、挑选、拷贝、复制声纹，从而达到优化声纹参数，进一步提高识别率，降低误识率的目的。

本发明了使用了开放式语音编码接口实现了全面兼容话音接收设备。目前在实际工作中，无论是因特网线路还是传统的电话线路，话音都占有相当的比重，各单位针对不同线路的特点研制了多种话音设备，但是在后端存储

方面没有统一的格式和接口，给话音的进一步处理带来了一定的困难。为了使本系统能够具有丰富的兼容性和扩展能力，前端话音接口子系统在充分研究目前我单位在用的话音设备基础上，采用不同的方式对多种话音压缩编码进行快速解调，并生成统一格式的 Windows 标准 PCM 话音，方便了话音识别的过程。

对于本领域的普通技术人员来说可显而易见的得出其他优点和修改。因此，具有更广方面的本发明并不局限于这里所示出的并且所描述的具体说明及示例性实施例。因此，在不脱离由随后权利要求及其等价体所定义的一般发明构思的精神和范围的情况下，可对其作出各种修改。

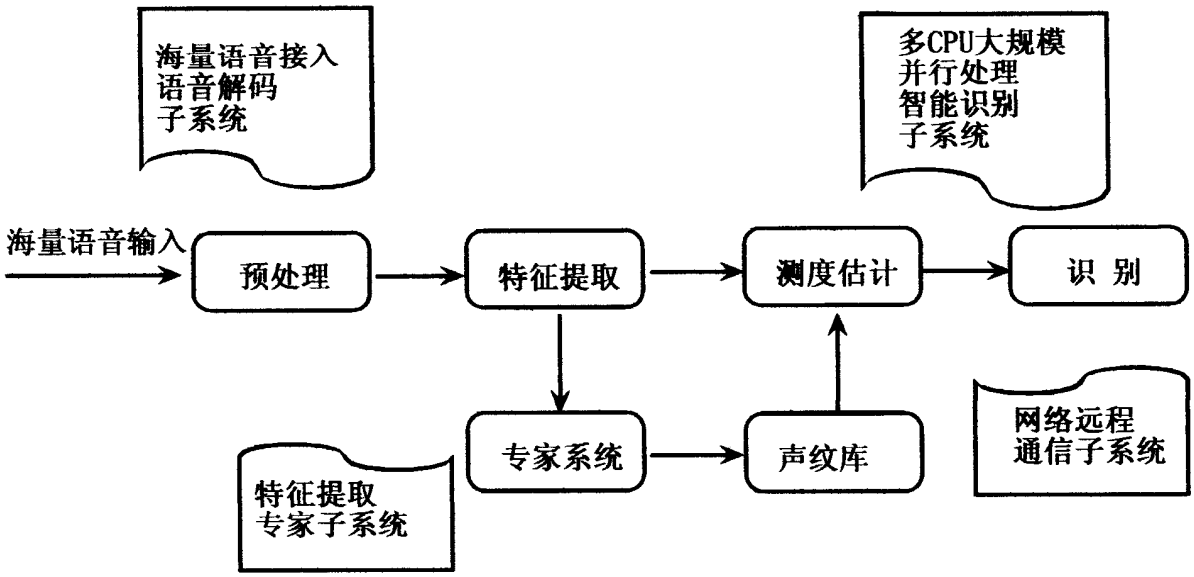

海量语音接入
语音解码
子系统
多CPU大规模
并行处理
智能识别
子系统
海量语音输入
预处理
特征提取
测度估计
识 别
网络远程
通信子系统
专家系统
声纹库
特征提取
专家子系统

图 1

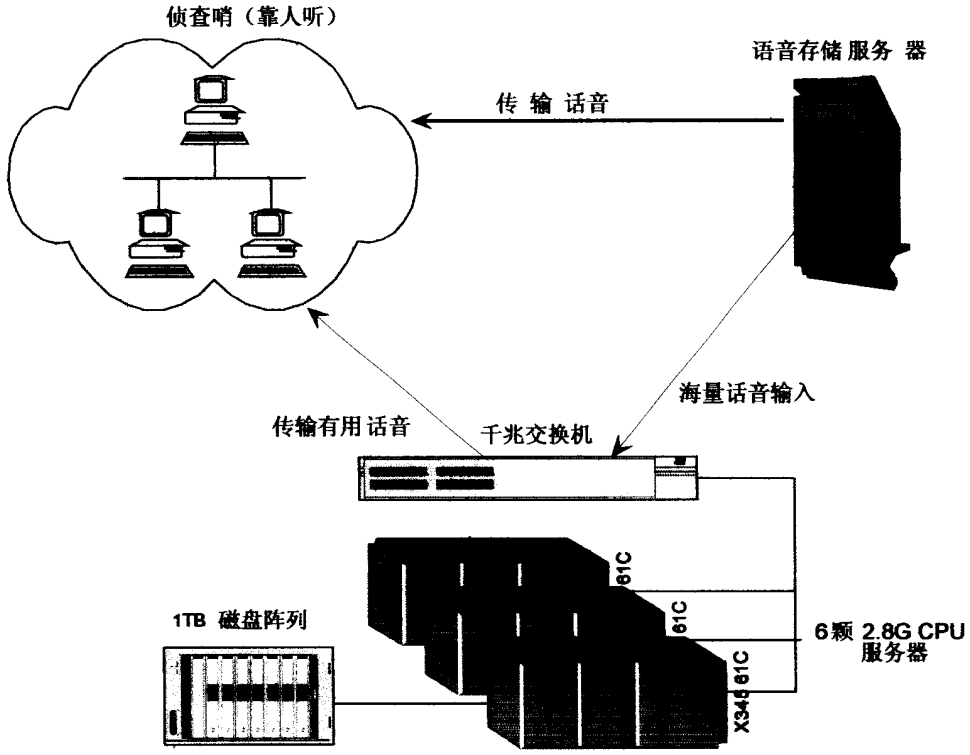

侦查哨（靠人听）
语音存储服务 器
传 输 话音
海量话音输入
传输有用 话音
千兆交换机
1TB 磁盘阵列
6颗 2.8G CPU
服务器

图 2

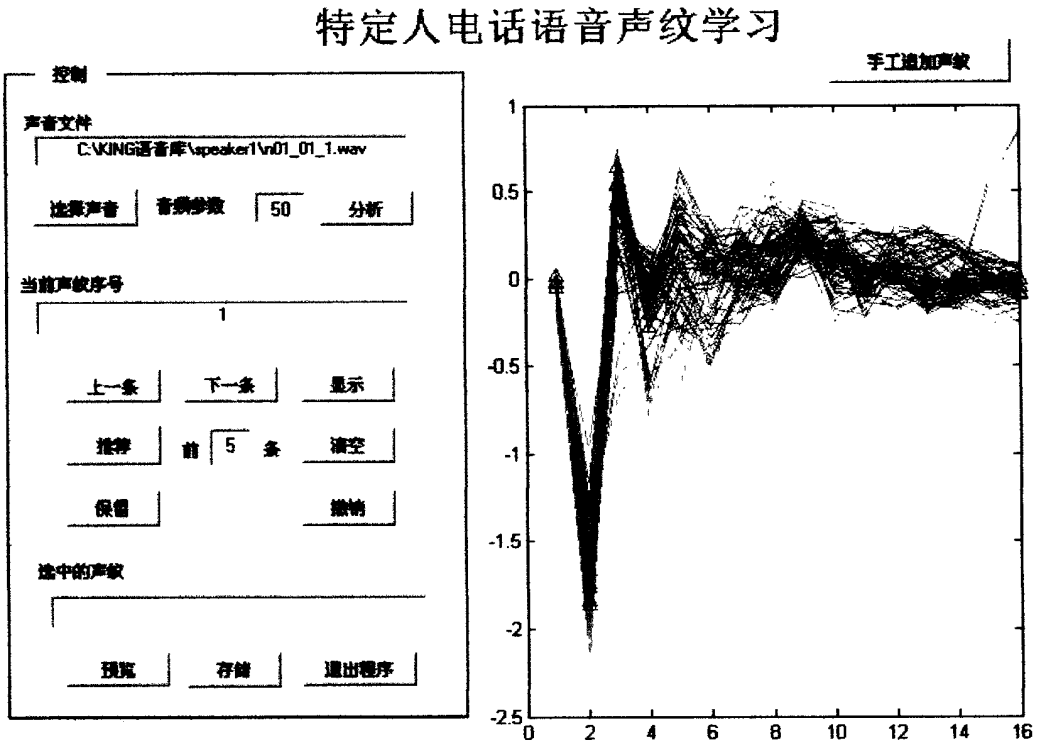

特定人电话语音声纹学习
控制
声音文件
C:\KING语音库\speaker1\n01_01_1.wav
注册声音
音频参数
50
分析
当前声纹序号
1
上一条
下一条
显示
推荐
前
5
条
清空
保留
撤销
选中的声纹
预览
存储
退出程序
手工追加声纹
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
2
4
6
8
10
12
14
16

图 3

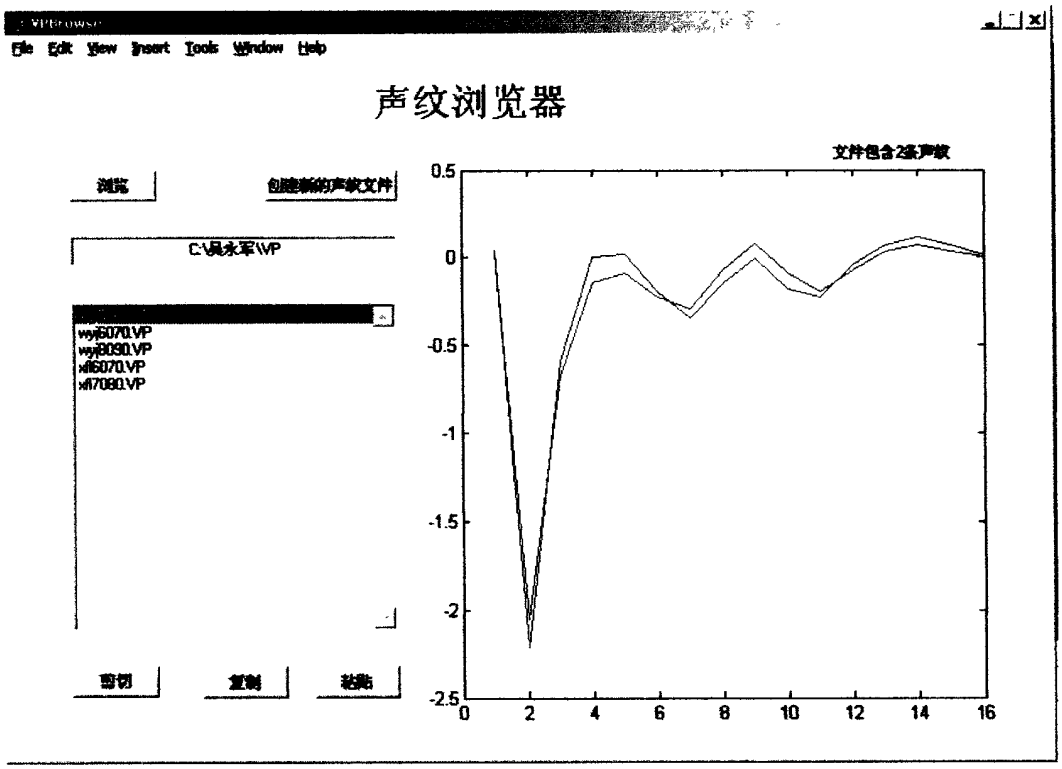

File Edit View Insert Tools Window Help
声纹浏览器
浏览
创建新的声纹文件
C:\吴永军\VP
wyj6070.VP
wyj8090.VP
xll6070.VP
xll7080.VP
剪切
复制
粘贴
文件包含2条声纹
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
2
4
6
8
10
12
14
16

图 4

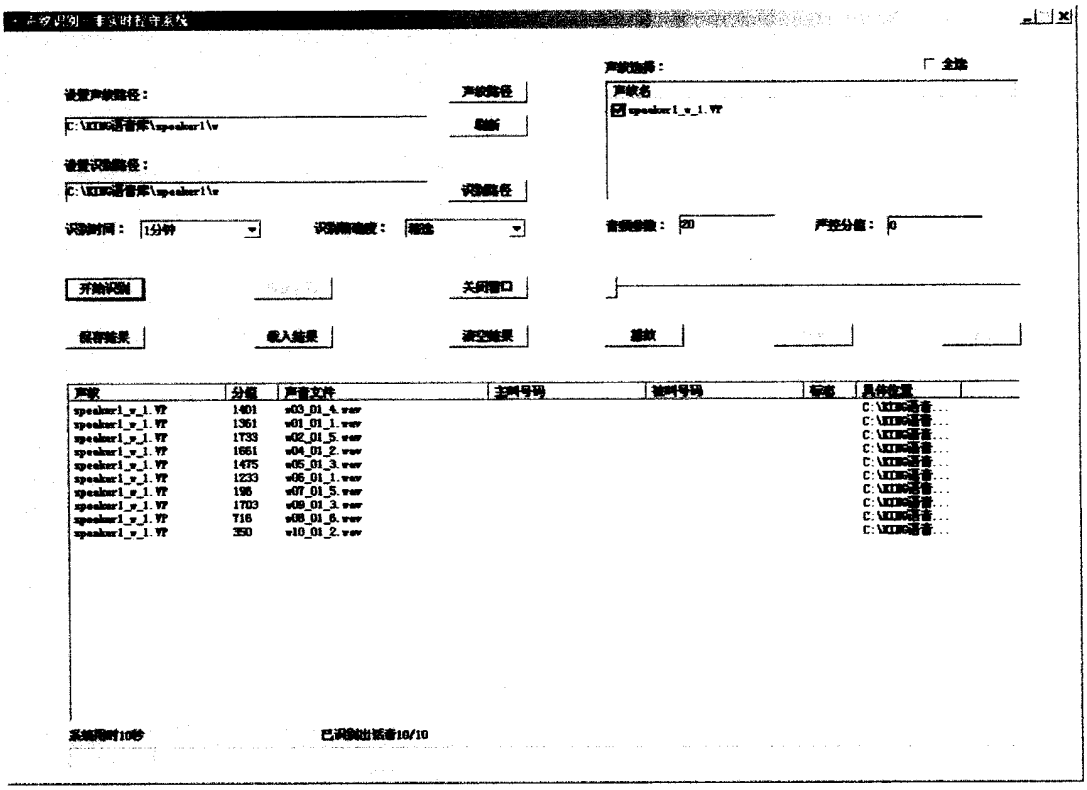


图 5

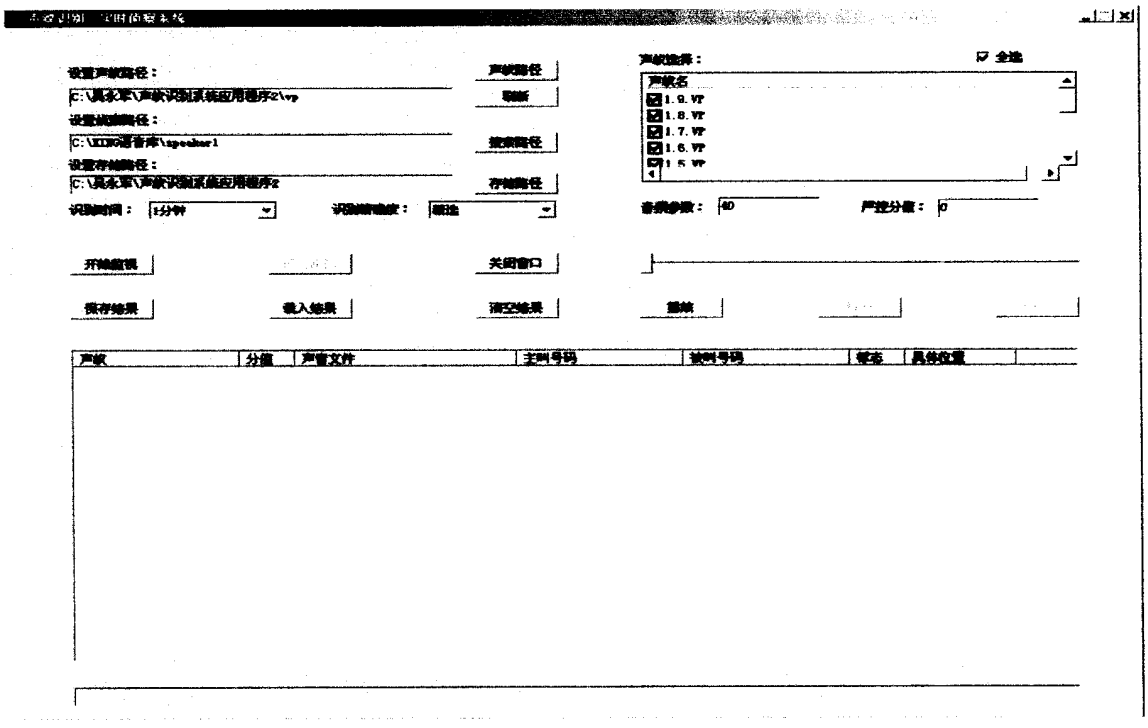


图 6

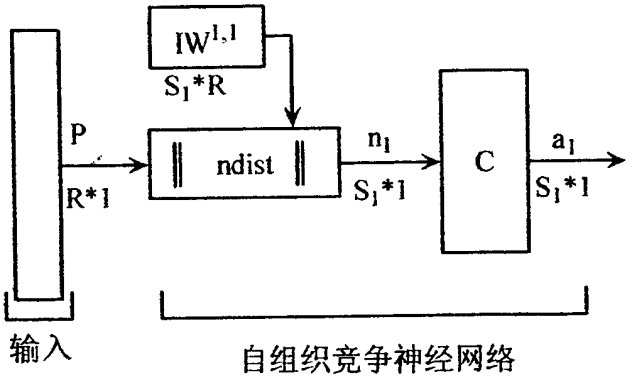

IW1,1
S1*R
P
R*1
ndist
n1
S1*1
C
a1
S1*1
输入
自组织竞争神经网络

图 7

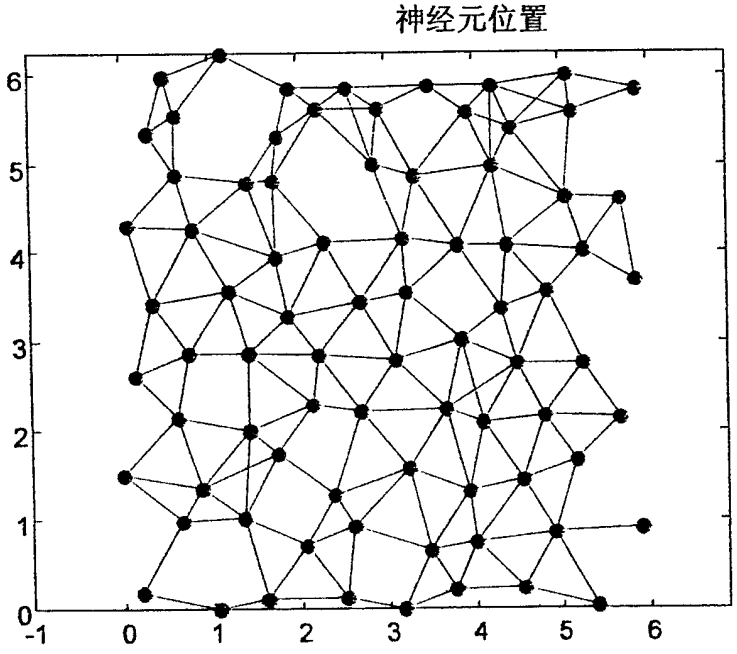

神经元位置
6
5
4
3
2
1
0
-1
0
1
2
3
4
5
6

图 8

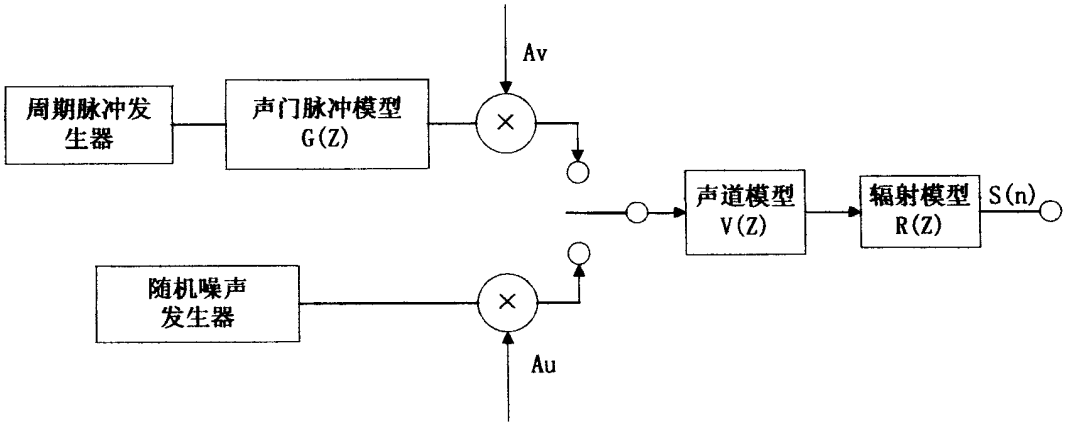

Av
周期脉冲发生器
声门脉冲模型 G(Z)
声道模型 V(Z)
辐射模型 R(Z)
S(n)
随机噪声发生器
Au

图 9

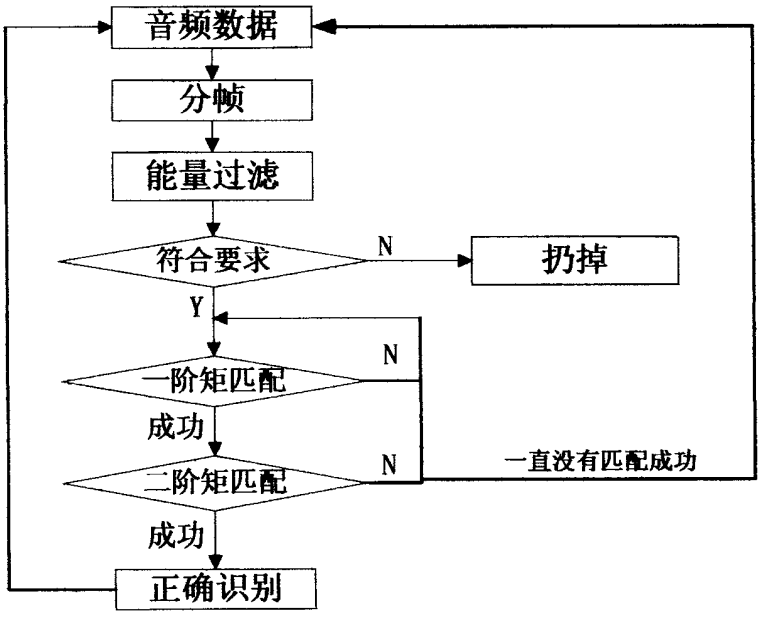

音频数据
分帧
能量过滤
符合要求
N
扔掉
Y
一阶矩匹配
N
成功
二阶矩匹配
N
一直没有匹配成功
成功
正确识别

图 10

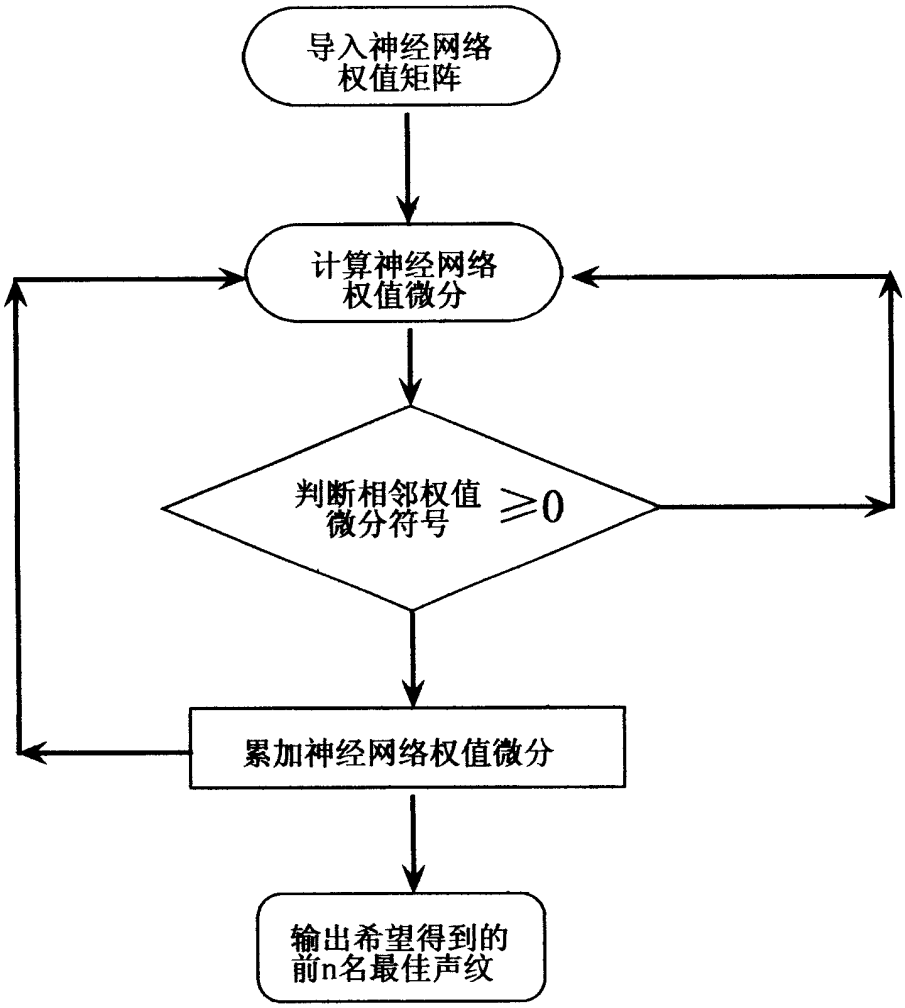
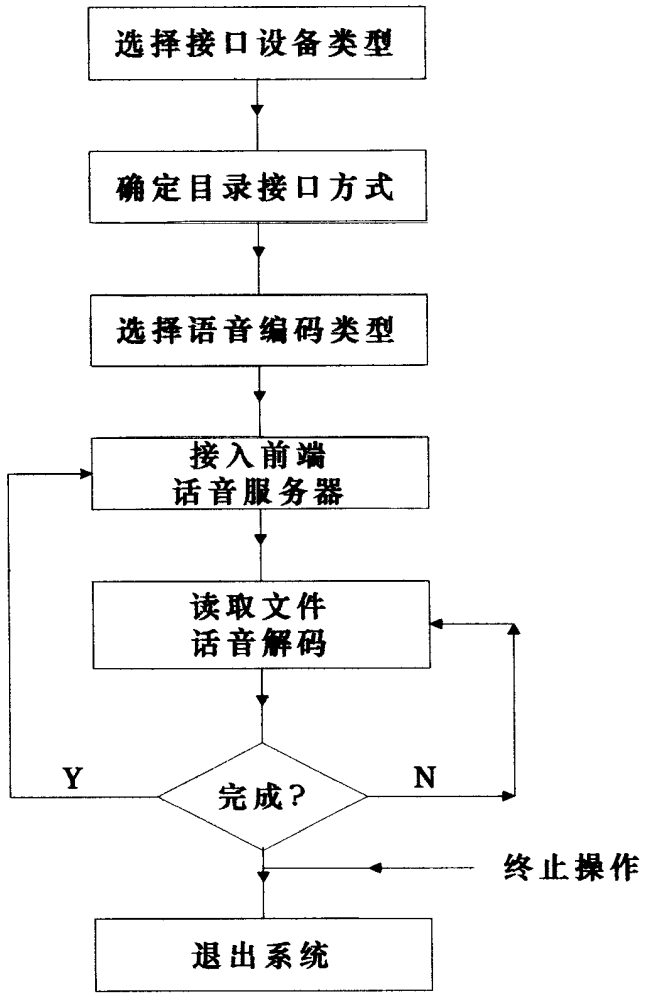

导入神经网络
权值矩阵
计算神经网络
权值微分
判断相邻权值
微分符号
≥0
累加神经网络权值微分
输出希望得到的
前n名最佳声纹

图 11


选择接口设备类型
确定目录接口方式
选择语音编码类型
接入前端
话音服务器
读取文件
话音解码
Y
完成?
N
终止操作
退出系统

话音接口流程图

图 12