



(12)发明专利申请

(10)申请公布号 CN 106328141 A

(43)申请公布日 2017. 01. 11

(21)申请号 201610803559.7

(22)申请日 2016.09.05

(71)申请人 南京大学

地址 210093 江苏省南京市鼓楼区汉口路
22号

(72)发明人 王晓亮 谭佳瑶 陆桑璐

(74)专利代理机构 南京苏高专利商标事务所
(普通合伙) 32204

代理人 许丹丹

(51) Int. Cl.

G10L 15/24(2013.01)

H04B 11/00(2006.01)

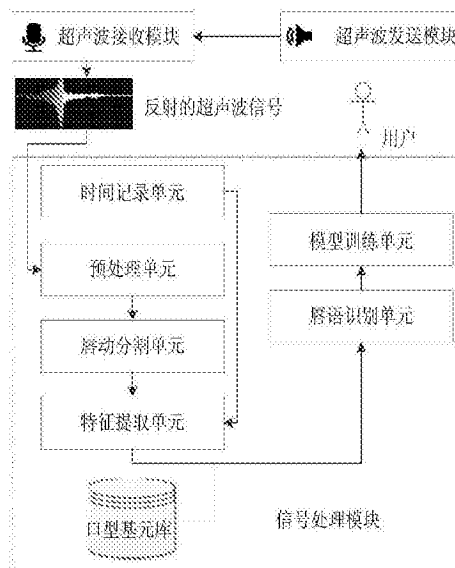
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种面向移动终端的超声波唇读识别装置及方法

(57)摘要

本发明公开了一种面向移动终端的超声波唇读识别装置及方法,装置包括超声波发送模块、超声波接收模块和信号处理模块;方法具体为超声波发送模块发送超声波信号,超声波信号在被嘴部反射后,由超声波接收模块接收,信号处理模块处理超声波接收模块获取的反射超声波信号以得到该信号的特征向量,根据特征向量识别口型,得到匹配结果。本发明的有益效果为:以移动终端作为超声波发送和接收模块,在无需额外硬件定制的基础上,利用超声波感知的能力识别唇语,拓展了超声波技术的应用场景,克服了传统唇读识别技术的不足,具有广泛的应用场景。



1. 一种面向移动终端的超声波唇读识别装置,其特征在于,包括:超声波发送模块、超声波接收模块和信号处理模块;超声波发送模块发送超声波信号,超声波信号在被嘴部反射后,由超声波接收模块接收,信号处理模块处理超声波接收模块获取的反射超声波信号以得到该信号的特征向量,根据特征向量识别口型,得到匹配结果。

2. 如权利要求1所述的面向移动终端的超声波唇读识别装置,其特征在于,超声波发送模块为移动终端的扬声器,超声波接收模块为移动终端的麦克风,均支持19KHz的超声波信号。

3. 如权利要求1所述的面向移动终端的超声波唇读识别装置,其特征在于,信号处理模块包括时间记录单元、预处理单元、唇动分割单元、特征提取单元、唇语识别单元、模型训练单元和口型基元库;时间记录单元记录每次发射超声波信号与接收超声波信号的时间差,预处理单元对超声波接收模块获取的反射超声波信号进行滤波得到唇动反射信号,唇动分割单元将唇动反射信号以单音节为单位分割成若干个单音节信号序列,特征提取单元针对每个单音节信号序列提取并处理对应口型的特征参数形成唇动特征向量,口型基元库应用特征提取单元提取的12种音节的特征参数,确立为12种基本口型,唇语识别单元调用口型基元库匹配所有单音节信号序列分量的结果,模型训练单元对分量匹配结果进行联合学习。

4. 一种面向移动终端的超声波唇读识别方法,包括如下步骤:

(1)超声波发送模块发送19KHz的超声波信号,同时信号处理模块的时间记录单元开始计时;

(2)超声波接收模块接收来自嘴部的超声波反射信号,并将反射信号传递到信号处理模块,同时信号处理模块的时间记录单元停止计时;

(3)信号处理模块的预处理单元对超声波接收模块的反射超声波信号进行滤波,从原始超声波反射信号中分离出唇动引起的反射信号即唇动反射信号,传递唇动反射信号给唇动分割单元;

(4)唇动分割单元对唇动反射信号进行分割,以单音节为单位进行分割,得到的音节信号序列传递给特征提取单元;

(5)特征提取单元检测每一个音节信号序列分量对应的特征参数,形成唇动特征向量;特征提取单元处理所有音节信号序列分量后,将每个分量的唇动特征向量传递给唇语识别单元;

(6)唇语识别单元对当前唇动特征向量进行识别,调用口型基元库,与口型基元库中的单口型唇动特征向量样本相匹配,根据口型基元库中单口型与音节的对应关系、单口型与特性向量对应关系,对当前唇动特征向量进行分类匹配,将所有分量匹配结果传递给模型训练单元进一步识别;

(7)模型训练单元基于马尔科夫假设,应用概率统计模型对分量匹配结果进行联合学习,结合语法规则集和基于上下文的纠错,统计具有最大概率的句子序列集,并输出为最终唇读结果。

5. 如权利要求4所述的面向移动终端的超声波唇读识别方法,其特征在于,步骤(4)中,唇动分割单元对唇动信号进行分割包括词内分割和词间分割;对唇动信号进行词间分割时,通过检测单词间存在的较为明显的停顿,设置一个长为 Tms 的滑动窗口,每次向前滑动

kms个单位,满足 $2k \leq T$,当检测到连续两次滑动窗口内不存在信号时,判断当前为停顿,为词间分割点;对唇动信号进行词内分割时,计算信号中19KHz的主频峰的个数n,根据个数对信号进行n均等词内分割。

6.如权利要求4所述的面向移动终端的超声波唇读识别方法,其特征在于,步骤(5)中的特征参数包括持续时间和频移;时间记录单元记录的发射超声波和接收超声波的时间差作为持续时间。

7.如权利要求4所述的面向移动终端的超声波唇读识别方法,其特征在于,步骤(5)中的频移特征提取步骤如下:

(1)采用快速傅里叶变换FFT计算主要频峰E和周围频带范围内的所有峰值点,主要频峰E为19KHz处的峰值,将小于19KHz的频段内的峰值存放入峰前数组F,大于19KHz频段内的峰值点存放在峰后数组A中;

(2)设置主要频峰和次要频峰的阈值比例k,扫描得到的频峰值数组A、F,若存在高于 $k \cdot E$ 的频峰值,则说明存在次要频峰,次要频峰即唇动引入的第二大频峰值;

(3)当确定了次要频峰位置后,对主要、次要频峰作差,得到唇动对应的频移 Δf 。

8.如权利要求4所述的面向移动终端的超声波唇读识别方法,其特征在于,步骤(7)中的模型训练单元应用概率统计模型统计最大概率的句子序列集,具体步骤如下:

(1)初始化概率统计模型的参数:口型状态O:定义为包括口型基元库中12种基本口型;音节状态S,即输出识别结果,定义为12种口型对应的所有元音辅音音节;转移概率 $P(O_i \rightarrow O_j)$:从口型状态 O_i 转移到口型状态 O_j 的概率;传输概率 $P(S_i | O_k, S_j)$:当前口型状态为 O_k 情况下,输出音节状态为 S_i 的概率;

(2)组合所有分量匹配结果时,第i个分量识别为音节状态 S_i 的概率与前一个口型状态 O_{i-1} 、当前口型状态 O_i 、后一个分量识别的音节状态 S_{i+1} 有关;具有最大概率的音节状态即作为当前分量的识别结果;即

$$P(S_i) = P(O_{i-1} \rightarrow O_i) \cdot P(S_i | O_i, S_{i+1})$$

(3)以此类推,计算到最后一个分量的识别结果,求解出对应的具有最大概率的序列 $S_1 S_2 \dots S_i \dots S_{n-1} S_n$ 。

一种面向移动终端的超声波唇读识别装置及方法

技术领域

[0001] 本发明涉及移动行为感知技术领域,尤其是一种面向移动终端的超声波唇读识别装置及方法。

背景技术

[0002] 现有技术中,移动终端上的行为感知技术被普遍应用,如呼吸检测、手势识别等,这种移动感知是一种更加智能化的人机交互方式。而唇读则是行为感知技术中更为细粒度的识别技术,是一项通过考察人们说话时嘴的运动模式,以此提高计算机的理解能力的技术。传统的唇读技术往往基于设备携带的传感器或者相机,借助移动设备携带的声学传感器进行语音识别,虽然能够精确的解释语音内容轻松实现唇读,但算法主要依赖于图像的质量,复杂度高。

发明内容

[0003] 本发明所要解决的技术问题在于,提供一种面向移动终端的超声波唇读识别装置及方法,对来自嘴部的反射信号进行特征提取匹配,实现唇读识别。

[0004] 为解决上述技术问题,本发明提供一种面向移动终端的超声波唇读识别装置,包括超声波发送模块、超声波接收模块和信号处理模块;超声波发送模块发送超声波信号,超声波信号在被嘴部反射后,由超声波接收模块接收,信号处理模块处理超声波接收模块获取的反射超声波信号以得到该信号的特征向量,根据特征向量识别口型,得到匹配结果。

[0005] 优选的,超声波发送模块为移动终端的扬声器,超声波接收模块为移动终端的麦克风,均支持19KHz的超声波信号。

[0006] 优选的,信号处理模块包括时间记录单元、预处理单元、唇动分割单元、特征提取单元、唇语识别单元、模型训练单元和口型基元库;时间记录单元记录每次发射超声波信号与接收超声波信号的时间差,预处理单元对超声波接收模块获取的反射超声波信号进行滤波得到唇动反射信号,唇动分割单元将唇动反射信号以单音节为单位分割成若干个单音节信号序列,特征提取单元针对每个单音节信号序列提取并处理对应口型的特征参数形成唇动特征向量,口型基元库应用特征提取单元提取的12种音节的特征参数,确立为12种基本口型,唇语识别单元调用口型基元库匹配所有单音节信号序列分量的结果,模型训练单元对分量匹配结果进行联合学习。

[0007] 一种面向移动终端的超声波唇读识别方法,包括如下步骤:

[0008] (1)超声波发送模块发送19KHz的超声波信号,同时信号处理模块的时间记录单元开始计时;

[0009] (2)超声波接收模块接收来自嘴部的超声波反射信号,并将反射信号传递到信号处理模块,同时信号处理模块的时间记录单元停止计时;

[0010] (3)信号处理模块的预处理单元对超声波接收模块的反射超声波信号进行滤波,从原始超声波反射信号中分离出唇动引起的反射信号即唇动反射信号,传递唇动反射信号

给唇动分割单元；

[0011] (4)唇动分割单元对唇动反射信号进行分割,以单音节为单位进行分割,得到的音节信号序列传递给特征提取单元；

[0012] (5)特征提取单元检测每一个音节信号序列分量对应的特征参数,形成唇动特征向量;特征提取单元处理所有音节信号序列分量后,将每个分量的唇动特征向量传递给唇语识别单元；

[0013] (6)唇语识别单元对当前唇动特征向量进行识别,调用口型基元库,与口型基元库中的单口型唇动特征向量样本相匹配,根据口型基元库中单口型与音节的对应关系、单口型与特性向量对应关系,对当前唇动特征向量进行分类匹配,将所有分量匹配结果传递给模型训练单元进一步识别；

[0014] (7)模型训练单元基于马尔科夫假设,应用概率统计模型对分量匹配结果进行联合学习,结合语法规则集和基于上下文的纠错,统计具有最大概率的句子序列集,并输出为最终唇读结果。

[0015] 优选的,步骤(4)中,唇动分割单元对唇动信号进行分割包括词内分割和词间分割;对唇动信号进行词间分割时,通过检测单词间存在的较为明显的停顿,设置一个长为 Tms 的滑动窗口,每次向前滑动 kms 个单位,满足 $2k \leq T$,当检测到连续两次滑动窗口内不存在信号时,判断当前为停顿,为词间分割点;对唇动信号进行词内分割时,计算信号中 $19KHz$ 的主频峰的个数 n ,根据个数对信号进行 n 均等词内分割。

[0016] 优选的,步骤(5)中的特征参数包括持续时间和频移;时间记录单元记录的发射超声波和接收超声波的时间差作为持续时间。

[0017] 优选的,步骤(5)中的频移特征提取步骤如下:

[0018] (1)采用快速傅里叶变换FFT计算主要频峰 E 和周围频带范围内的所有峰值点,主要频峰 E 为 $19KHz$ 处的峰值,将小于 $19KHz$ 的频段内的峰值存入峰前数组 F ,大于 $19KHz$ 频段内的峰值点存放在峰后数组 A 中;

[0019] (2)设置主要频峰和次要频峰的阈值比例 k ,扫描得到的频峰值数组 A 、 F ,若存在高于 $k \cdot E$ 的频峰值,则说明存在次要频峰,次要频峰即唇动引入的第二大频峰值;

[0020] (3)当确定了次要频峰位置后,进一步对主要、次要频峰作差,得到唇动对应的频移 Δf 。

[0021] 优选的,步骤(7)中的模型训练单元应用概率统计模型统计最大概率的句子序列集,具体步骤如下:

[0022] (1)初始化概率统计模型的参数:口型状态 O ,定义为包括口型基元库中12种基本口型;音节状态 S ,即输出识别结果,定义为12种口型对应的所有元音辅音音节;转移概率 $P(O_i \rightarrow O_j)$,从口型状态 O_i 转移到口型状态 O_j 的概率;传输概率 $P(S_i | O_k, S_j)$,当后一个音节状态为 S_j ,当前口型状态为 O_k 情况下,输出音节状态为 S_i 的概率;

[0023] (2)组合所有分量匹配结果时,第 i 个分量识别为音节状态 S_i 的概率与前一个口型状态 O_{i-1} 、当前口型状态 O_i 、后一个分量识别的音节状态 S_{i+1} 有关;具有最大概率的音节状态即作为当前分量的识别结果;即

[0024]
$$P(S_i) = P(O_{i-1} \rightarrow O_i) \cdot P(S_i | O_i, S_{i+1})$$

[0025] (3)以此类推,计算到最后一个分量的识别结果,求解出对应的具有最大概率的序

列 $S_1S_2 \dots S_i \dots S_{n-1}S_n$ 。

[0026] 本发明的有益效果为:以移动终端作为超声波发送和接收模块,在无需额外硬件定制的基础上,利用超声波感知的能力识别唇语,拓展了超声波技术的应用场景,克服了传统唇读识别技术的不足,具有广泛的应用场景。

附图说明

[0027] 图1是本发明的整体装置结构示意图。

[0028] 图2是本发明的方法流程图。

[0029] 图3是本发明的移动终端配置示意图。

[0030] 图4是本发明的特征提取流程图。

[0031] 图5是本发明的应用概率统计模型识别唇读示意图。

[0032] 图6是本发明的口型状态数字标号与口型、音节对应关系图。

[0033] 图7是本发明的口型状态转移关系图。

[0034] 图8是本发明的口型状态到音节状态转移关系图。

具体实施方式

[0035] 如图1所示,一种面向移动终端的超声波唇读识别装置,包括超声波发送模块、超声波接收模块和信号处理模块;超声波发送模块发送超声波信号,超声波信号在被嘴部反射后,由超声波接收模块接收,信号处理模块处理超声波接收模块获取的反射超声波信号以得到该信号的特征向量,根据特征向量识别口型,得到匹配结果。

[0036] 超声波发送模块为移动终端的扬声器,超声波接收模块为移动终端的麦克风,均支持19KHz的超声波信号。

[0037] 信号处理模块包括时间记录单元、预处理单元、唇动分割单元、特征提取单元、唇语识别单元、模型训练单元和口型基元库;时间记录单元记录每次发射超声波信号与接收超声波信号的时间差,预处理单元对超声波接收模块获取的反射超声波信号进行滤波得到唇动反射信号,唇动分割单元将唇动反射信号以单音节为单位分割成若干个单音节信号序列,特征提取单元针对每个单音节信号序列提取并处理对应口型的特征参数形成唇动特征向量,口型基元库结合汉语的发音特征,应用特征提取单元提取的12种音节的特征参数,确立为12种基本口型,唇语识别单元调用口型基元库匹配所有单音节信号序列分量的结果,模型训练单元对分量匹配结果进行联合学习。

[0038] 如图1和2所示,一种面向移动终端的超声波唇读识别方法,包括如下步骤:(1)超声波发送模块发送19KHz的超声波信号,同时信号处理模块的时间记录单元开始计时;

[0039] (2)超声波接收模块接收来自嘴部的超声波反射信号,并将反射信号传递到信号处理模块,同时信号处理模块的时间记录单元停止计时;

[0040] (3)信号处理模块的预处理单元对超声波接收模块的反射超声波信号进行滤波,从原始超声波反射信号中分离出唇动引起的反射信号即唇动反射信号,传递唇动反射信号给唇动分割单元;

[0041] (4)唇动分割单元对唇动反射信号进行分割,以单音节为单位进行分割,得到的音节信号序列传递给特征提取单元;

[0042] (5)特征提取单元检测每一个音节信号序列分量对应的特征参数,形成唇动特征向量;特征提取单元处理所有音节信号序列分量后,将每个分量的唇动特征向量传递给唇语识别单元;

[0043] (6)唇语识别单元对当前唇动特征向量进行识别,调用口型基元库,与口型基元库中的单口型唇动特征向量样本相匹配,根据口型基元库中单口型与音节的对应关系、单口型与特性向量对应关系,对当前唇动特征向量进行分类匹配,将所有分量匹配结果传递给模型训练单元进一步识别;

[0044] (7)模型训练单元基于马尔科夫假设,应用概率统计模型对分量匹配结果进行联合学习,结合语法规则集和基于上下文的纠错,统计具有最大概率的句子序列集,并输出为最终唇读结果。

[0045] 如图3所示,为根据本发明实施方式设计的支持不同唇动模式的移动终端配置的示意图,具体实施过程如下:

[0046] (1)超声波发送模块,即移动终端的扬声器发送出19KHz的超声波信号。同时启动信号处理模块的时间记录单元开始计时。

[0047] (2)由于多普勒效应,超声波信号碰到嘴部后,不同的唇动模式将导致超声波信号的时间频率分布发生变化。根据图3所示,若移动终端包括一个超声波发送模块和一个超声波接收模块,以及用户的嘴部做不同的运动模式,则估计的反射超声波信号频率如下所示:

$$[0048] \quad f' = \left(\frac{C \pm V}{C} \right) f_0$$

[0049] 其中, f' 为在用户的手处接收超声波信号的频率, f_0 为超声波信号的发送频率19KHz, C 为声波在空气中传播的速度, V 为嘴部相对于介质的速度。式中+、-分别表示嘴部靠近、远离移动终端。

[0050] (3)超声波接收模块,即移动终端的麦克风接收来自嘴部的反射信号。信号处理模块的时间记录单元停止计时。

[0051] 如图4所示,为特征提取单元提取唇动特性向量的流程图,包括以下步骤:

[0052] (1)采用快速傅里叶变化FFT运算对唇动信号进行时频变换。

[0053] (2)计算频谱上主要频峰值E,即19KHz对应的峰值点以及周围所有频段内其他所有峰值点。

[0054] (3)对计算的所有频峰点进行判断,频率低于19KHz频段内的峰值点存入峰前数组F,频率高于19KHz的频段内的峰值点存入峰后数组A。

[0055] (4)扫描得到峰前、峰后数组,搜索是否存在次要频峰。首先,设置主要频峰和次要的阈值比例k,若在数组中存在大于 $k \cdot E$ 的峰值点,则判断为次要频峰e。

[0056] (5)对主要频峰E、次要频峰e作差,与时间记录单元记录的时间差,分别作为频移、持续时间特征参数。

[0057] 如图5所示,为模型训练单元应用隐式马尔可夫模型统计最大概率的句子序列集的过程。

[0058] (1)首先,初始化概率统计模型四个主要参数:

[0059] 口型状态0:定义为口型基元库对应的12种基本口型。为了更加形象描述,将12种口型用1-12的数字标号代替,对应关系如图6所示。

[0060] $O = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

[0061] 音节状态S:定义为如图6所示的12种基本口型对应的所有元音辅音音节。

[0062] 转移概率 $P(O_i \rightarrow O_j)$:从口型状态 O_i 转移到口型状态 O_j 的概率。如图7所示,根据语法规则对各个口型状态之间的转移进行了约束。例如状态1只可转移到状态4或者状态5,这是由于语法规则中状态1对应的音节a只与状态4对应的音节i可形成复合音节/ai/,或者与状态5对应的u复合发音形成/au/,因此, $P(1 \rightarrow 4) + P(1 \rightarrow 5) = 1$ 。

[0063] 传输概率 $P(S_i | O_k, S_j)$:当后一个音节状态为 S_j 时,当前口型状态为 O_k 情况下,输出音节状态为 S_i 的概率;如图8所示,定义了所有口型状态到音节状态之间转移的关系,进一步基于现有的语料库统计概率。

$$[0064] \quad P(S_i | O_k, S_j) = \frac{N_{O_k, S_i, S_j}}{\sum_{m=1}^n N_{O_k, S_m, S_j}}$$

[0065] 其中, N_{O_k, S_i, S_j} 表示在语料库中当前口型为 O_k ,后一个为音节 S_j 时,输出音节状态为 S_i 的数目。 $\sum_{m=1}^n N_{O_k, S_m, S_j}$ 表示在语料库中,当前口型为 O_k ,后一个音节为 S_j 时,输出为 O_k 状态下所有音节状态的数目。例如,当识别“book(/buk/)”,基于现有的英文语料库,利用上式可得

$$P(b|6, u) = \frac{N_{6, b, u}}{N_{6, b, u} + N_{6, p, u} + N_{6, m, u}} = 31.1\%, \text{ 同时 } P(p|6, u) = 28.9\%, \text{ 由此识别第一个音节为b}$$

而不是p。

[0066] (2)依次识别每个分量,第i个分量识别为音节状态 S_i 的概率与前一个口型状态 O_{i-1} 、当前口型状态 O_i 、后一个分量识别的音节状态 S_{i+1} 有关;具有最大概率的音节状态即作为当前分量的识别结果;即

$$[0067] \quad P(S_i) = P(O_{i-1} \rightarrow O_i) \cdot P(S_i | O_i, S_{i+1})$$

[0068] (3)以此类推,计算到最后一个分量的识别结果,求解出对应的具有最大概率的序列 $S_1 S_2 \dots S_i \dots S_{n-1} S_n$ 。

[0069] 尽管本发明就优选实施方式进行了示意和描述,但本领域的技术人员应当理解,只要不超出本发明的权利要求所限定的范围,可以对本发明进行各种变化和修改。

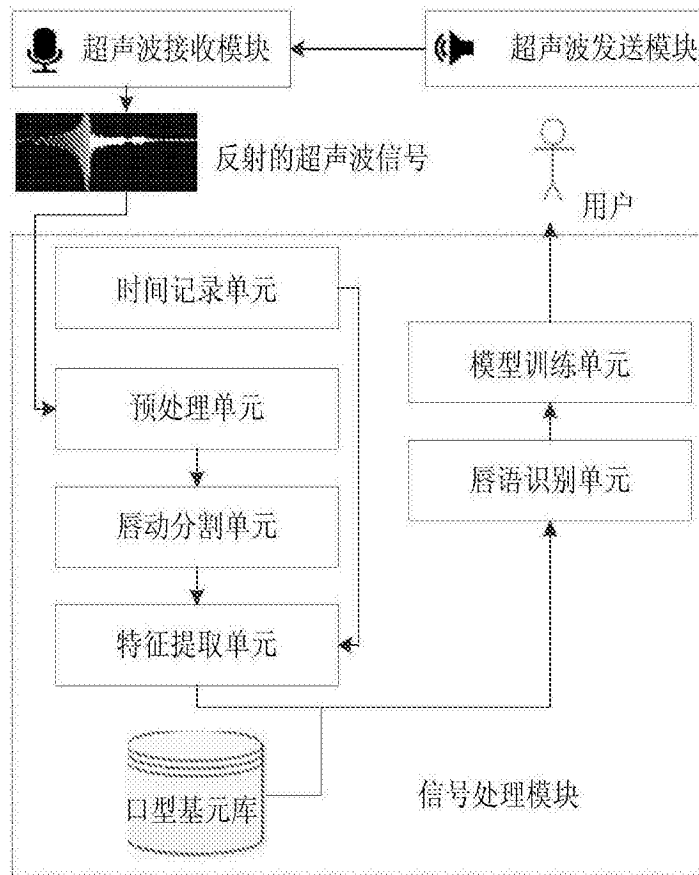


图1

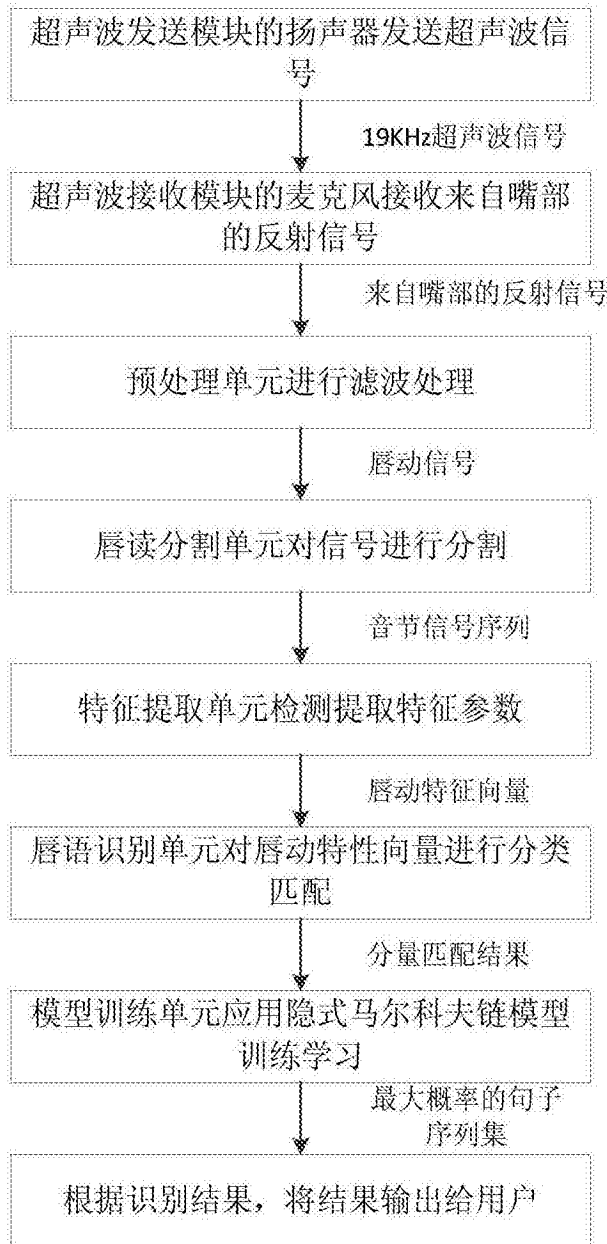


图2

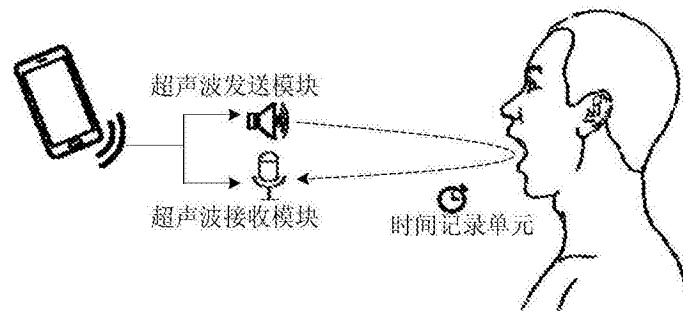


图3

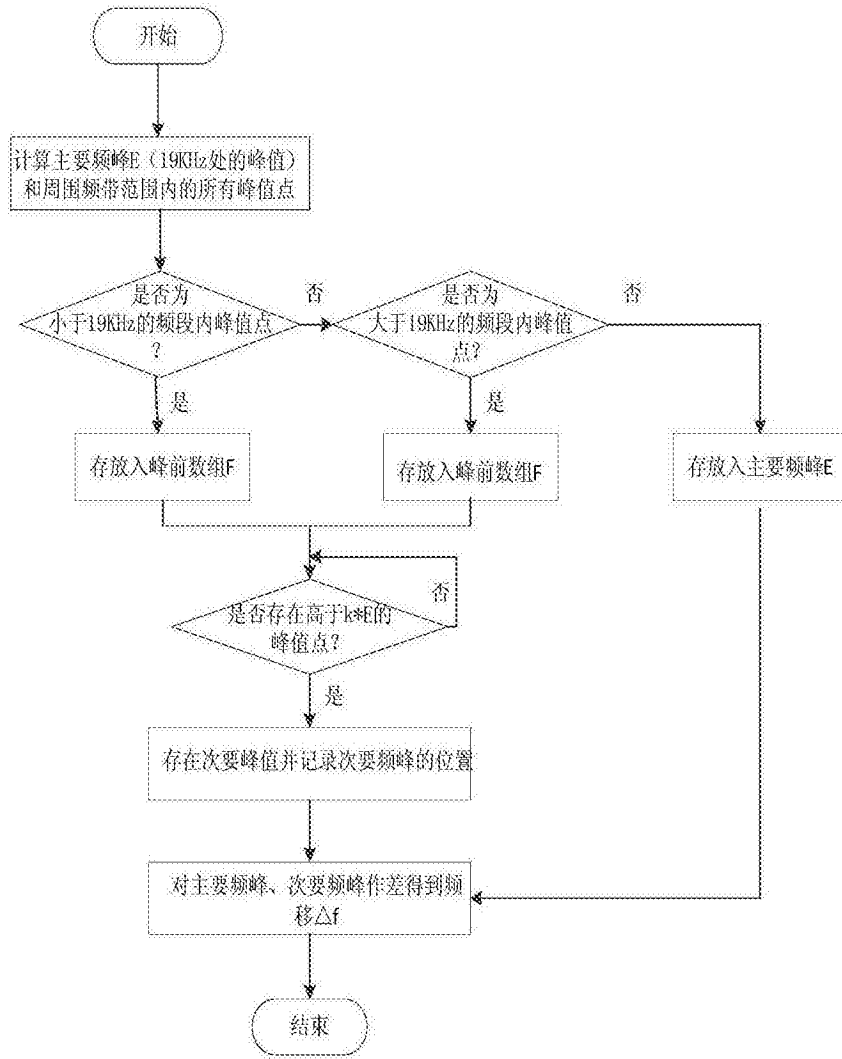


图4

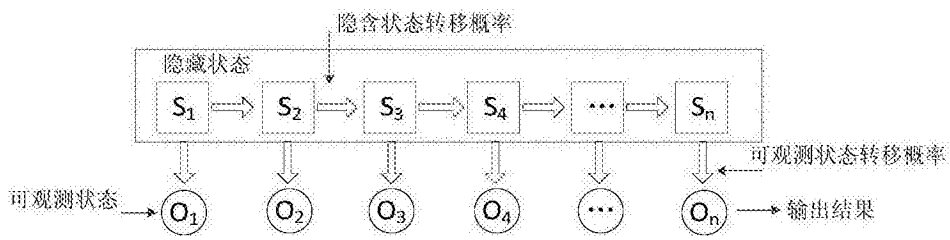


图5

元音	辅音
1: $\Lambda/a:$	6:b p m
2: $\text{ɔ}/\text{ɔ}:$	7:f
3: $\text{ə}/\text{ə}:$	8:d t n l
4: $i/i:$	9:g k h
5: $u/u:$	10:j
	11:dʒ ʃ tʃ r
	12:z tz s ts θ

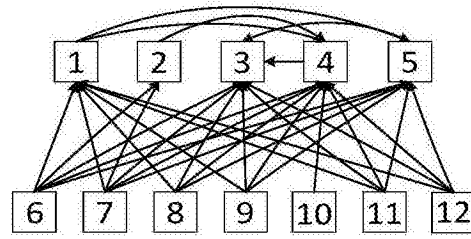


图7

图6

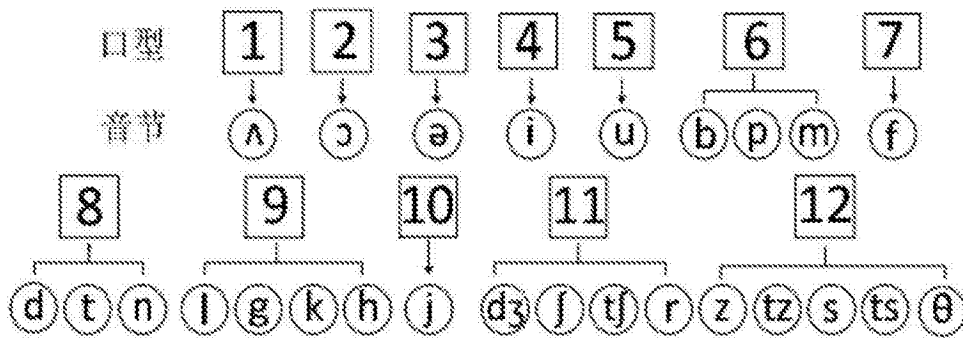


图8