



(12)发明专利申请

(10)申请公布号 CN 111488433 A

(43)申请公布日 2020.08.04

(21)申请号 202010594489.5

(22)申请日 2020.06.28

(71)申请人 南京银行股份有限公司

地址 210008 江苏省南京市玄武区中山路
288号

(72)发明人 丁晓平 姜欣荣 吴迪

(74)专利代理机构 南京理工大学专利中心
32203

代理人 马鲁晋

(51) Int. Cl.

G06F 16/33(2019.01)

G06F 16/332(2019.01)

G06N 5/02(2006.01)

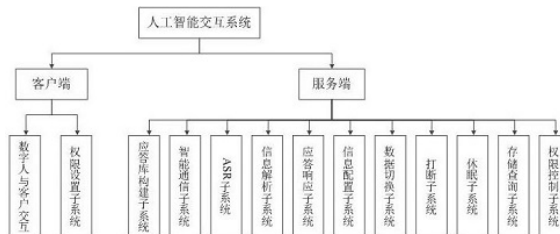
权利要求书5页 说明书12页 附图9页

(54)发明名称

一种适用于银行的提升现场体验感的人工智能交互系统

(57)摘要

本发明公开了一种适用于银行的提升现场体验感的人工智能交互系统,应用于银行业务办理终端,系统包括客户端,用于通过数字员工与用户进行面对面交互,接收用户语音输入信息并发送至服务端,响应并展示服务端反馈的应答内容,用户语音输入信息包括业务办理需求信息;服务端,用于响应用户语音输入信息进行应答,还用于配置客户端和服务端的信息。本发明能够通过自然语言理解、上下文理解和提槽技术,在不断的交互过程中充分理解用户的意图,与用户进行多轮交互,并由数字员工通过多媒体资源展示给客户查看,更加拟人智能化,提升客户体验感。此外,能够实现智能打断,及时响应用户的需求,使人机交互更加高效、快捷、自然,提高业务办理效率。



1. 一种适用于银行的提升现场体验感的人工智能交互系统,应用于银行业务办理终端,其特征在于,所述系统包括客户端和服务端;

所述客户端,用于通过数字员工与用户进行面对面交互,接收用户语音输入信息并发送至服务端,响应并展示服务端反馈的应答内容;所述用户语音输入信息包括业务办理需求信息;

所述服务端,用于响应所述用户语音输入信息进行应答,还用于配置所述客户端和服务端的信息;

所述服务端包括:

应答库构建子系统,用于构建客户端与用户交互的知识应答库,该知识应答库包括业务办理应答库;

智能通信子系统,用于提取并过滤所述用户语音输入信息中的噪声信息,并将过滤后的用户语音输入信息发送至ASR子系统;

ASR子系统,用于将所述用户语音输入信息解析为文字,并发送至信息解析子系统;

信息解析子系统,用于解析所述用户语音输入信息和文字,识别用户需求,包括业务办理需求,并将该需求信息发送至应答响应子系统;

应答响应子系统,用于响应所述需求信息,从所述应答库中提取应答内容并推送至客户端;所述应答内容包括数字员工应答语音、文字、驱动信息,以及业务交互窗口;所述数字员工驱动信息包括动作、表情和口型;

数据信息配置子系统,用于管理、配置客户端和服务端的数据信息。

2. 根据权利要求1所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述应答库构建子系统包括:

第一应答元素子库生成模块,用于生成每一项业务的办理流程,包括多级业务交互过程;

第二应答元素子库生成模块,用于将每一级业务交互过程转换为对应的交互语音和文字,生成多级交互语音和文字;

第三应答元素子库生成模块,用于生成每一级业务交互过程对应的业务交互窗口;

第四应答元素子库生成模块,用于生成每一级交互语音和文字对应的数字员工驱动信息;

映射关系建立模块,用于建立所述每一级业务交互过程与对应的交互语音和文字之间的映射关系,每一级业务交互过程与对应的业务交互窗口之间的映射关系,以及每一级交互语音和文字与其对应的数字员工驱动信息之间的映射关系。

3. 根据权利要求2所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述应答响应子系统包括:

第一响应模块,用于驱动所述第一应答元素子库响应所述需求信息,触发该需求信息对应的某一级业务交互过程,并生成第一触发指令;

第二响应模块,用于驱动所述第二应答元素子库响应所述第一触发指令,触发所述某一级业务交互过程对应的某一级交互语音和文字,并将所述某一级交互语音和文字发送至应答输出模块,同时生成第二触发指令;

第三响应模块,用于驱动所述第三应答元素子库响应所述第二触发指令,触发所述某

一级交互语音和文字对应的业务交互窗口,并将该业务交互窗口发送至应答输出模块;

第四响应模块,用于驱动所述第四应答元素子库响应所述第二触发指令,触发所述某一级交互语音和文字对应的数字员工驱动信息,并将该数字员工驱动信息发送至应答输出模块,同时生成第三触发指令;

应答输出模块,用于响应所述第三触发指令,将所述需求信息对应的所述某一级交互语音和文字、业务交互窗口,以及数字员工驱动信息同步推送至客户端。

4. 根据权利要求1所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述服务端还包括安全验证子系统,该子系统包括:

安全身份库构建模块,用于构建包括用户音频数据的安全身份库;所述安全身份库包括注册库和验证库,所述注册库包括用户注册的语音音频,所述验证库包括用户对某一指定文本录入的语音音频;

第一安全验证模块,用于在所述银行业务办理终端进行初次用户身份安全验证之后进行二次身份安全验证,验证通过后启动所述人工智能交互系统;

第二安全验证模块,用于根据用户语音输入信息实时判别用户是否为与银行账户关联的安全身份,若不是安全身份,触发锁定所述人工智能交互系统,并进一步触发锁定所述银行业务办理终端。

5. 根据权利要求4所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述根据用户语音输入信息实时判别用户是否为与银行账户关联的安全身份,具体过程包括:

从所述安全身份库中提取与当前银行账户关联的用户的某一语音音频,对该音频以及当前用户输入的语音音频分别进行特征提取,获得各自对应的音频向量;

利用余弦相似度公式求取两个音频向量的相似度,记为第一相似度;

判断所述第一相似度是否高于第一预设阈值,若高于,则判定当前用户为安全身份,否则执行下一步;

构建基于深度学习的声音相似度计算模型;具体过程包括:

从所述注册库中提取用户A的语音音频,对该音频进行特征提取,获得对应的音频向量,记为音频特征 a ;

对所述验证库中每一个用户的语音音频进行特征提取,获得对应的音频向量,分别记为音频特征 b_1, b_2, \dots, b_n ,构成音频特征集 B ; n 为验证库中语音音频的数目;

从所述音频特征集 B 中随机抽取若干音频特征,每个音频特征分别与所述音频特征 a 构成一个样本,形成训练样本集;

将所述训练样本集输入LSTM神经网络进行训练,生成声音相似度计算模型;

将所述当前用户输入的语音音频的音频向量,以及音频特征集 B 中该用户对应的音频特征输入至所述声音相似度计算模型,输出两者的相似度,记为第二相似度;

判断所述第二相似度是否高于第二预设阈值,若高于,则判定当前用户为安全身份,否则判定当前用户为不安全身份。

6. 根据权利要求5所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述特征提取的具体过程包括:

提取语音音频中的音调特征,具体过程包括:

利用YIN算法处理语音音频信号;

检测YIN算法处理后的信号的基频的周期即音调特征,具体包括:

提取YIN算法处理后的信号的极小值点;

判断是否存在极小值点小于预设判决阈值,若否,则将其中最小的极小值点与所述预设判决阈值的偏移量作为基频的周期;若是,则执行下述步骤;

获取所有小于预设判决阈值的极小值点,并将极小值的相对绝对差值小于预设误差的极小值点归入同一极小值点序列,获得若干极小值点序列,且对每个序列中的极小值点进行升序排列;

根据每个极小值点序列中极小值点的数目,求取该极小值点序列的权重;

获取权重值最大的极小值点序列中最小的极小值点,将该最小极小值点与所述预设判决阈值的偏移量作为基频的周期;

提取语音音频的梅尔频率倒谱系数;

提取语音音频的纹理特征;

将所述音调特征、梅尔频率倒谱系数及纹理特征拼接,形成音频向量。

7.根据权利要求5所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述特征提取的具体过程包括:

提取语音音频中的音调特征,具体过程包括:

构建函数:

$$HB_{\tau} = \frac{MI_{\tau} \cdot SU1_{\tau} \cdot SU2_{\tau}}{MA_{\tau} \cdot (L - \tau)^2}$$

$$YD_{\tau} = d_{\tau}^{HB_{\tau}}$$

其中, d_{τ} 为AMDF函数变换公式,公式为:

$$d_{\tau} = \frac{1}{L - \tau} \sum_{i=0}^{L - \tau - 1} |x_i - x_{i + \tau}|$$

式中, x 表示语音音频信号, L 为数据帧的长度, τ 为偏移量,其取值范围为 $0 \sim L - 1$; x_i 表示数据帧中 i 位置处的信号;

MI_{τ} 为两帧信号中各自数据最小值的总和,公式为:

$$MI_{\tau} = \sum_{i=0}^{L - \tau - 1} [\text{sgn}(x_i, x_{i + \tau}) \cdot \text{Min}(|x_i|, |x_{i + \tau}|)]$$

式中, $\text{sgn}(x_i, x_{i + \tau}) = \begin{cases} 1, & x_i \cdot x_{i + \tau} \geq 0 \\ 0, & \text{其他} \end{cases}$;

MA_{τ} 为两帧信号中各自数据最大值的总和,公式为:

$$MA_{\tau} = \sum_{i=0}^{L-\tau-1} \{ \text{sgn}(x_i, x_{i+\tau}) \cdot \text{Max}(|x_i|, |x_{i+\tau}|) + [1 - \text{sgn}(x_i, x_{i+\tau})] \cdot (|x_i| + |x_{i+\tau}|) \}$$

$SU1_{\tau}$ 为两帧信号中对应位置数据值异号的数据点总数, 公式为:

$$SU1_{\tau} = \sum_{i=0}^{L-\tau-1} \text{sgn} |x_i, x_{i+\tau}|$$

$SU2_{\tau}$ 为两帧信号中相邻数据点呈正向变化的数据点总数与相邻数据点呈负向变化的数据点总数的最小值, 公式为:

$$SU2_{\tau} = \text{Min} \left\{ \left[\sum_{i=0}^{L-\tau-1} \text{sgn}(x_i - x_{i-1}, x_{i+\tau} - x_{i+\tau-1}) \right] \left[\sum_{i=0}^{L-\tau-1} \text{sgn}(x_{i+\tau} - x_{i+\tau-1}, x_i - x_{i-1}) \right] \right\}$$

利用所述函数对语音音频信号进行变换;

检测函数变换后的信号的基频的周期即音调特征, 具体为:

提取函数变换后的信号的极小值点;

判断是否存在极小值点小于预设判决阈值, 若否, 则将其中最小的极小值点与所述预设判决阈值的偏移量作为基频的周期; 若是, 则执行下述步骤;

获取所有小于预设判决阈值的极小值点, 并将极小值的相对绝对差值小于预设误差的极小值点归入同一极小值点序列, 获得若干极小值点序列, 且对每个序列中的极小值点进行升序排列;

根据每个极小值点序列中极小值点的数目, 求取该极小值点序列的权重;

获取权重值最大的极小值点序列中最小的极小值点, 将该最小极小值点与所述预设判决阈值的偏移量作为基频的周期;

提取语音音频的梅尔频率倒谱系数;

提取语音音频的纹理特征;

将所述音调特征、梅尔频率倒谱系数及纹理特征拼接, 形成音频向量。

8. 根据权利要求1所述的适用于银行的提升现场体验感的人工智能交互系统, 其特征在于, 所述服务端还包括数据切换子系统, 该子系统包括:

角色切换模块, 用于响应所述需求信息, 切换与该需求信息匹配的数字员工角色;

场景切换模块, 用于响应所述需求信息, 切换与该需求信息匹配的业务交互场景;

所述服务端还包括打断子系统, 用于在数字员工输出应答语音的过程中, 实时接收新的用户语音输入信息, 在接收到新的用户语音输入信息时暂停当前应答语音的输出, 同时生成并发送触发命令至ASR子系统, 对新的用户语音输入信息进行新一轮响应。

9. 根据权利要求1所述的适用于银行的提升现场体验感的人工智能交互系统, 其特征在于,

所述服务端还包括休眠子系统, 用于实时监测用户语音输入信息, 在未监测到用户语音输入信息的时常超过预设阈值时, 控制所述人工智能交互系统进入休眠状态; 在休眠过程中, 若监测到用户语音输入信息, 即时唤醒人工智能交互系统恢复至休眠前的状态;

所述服务端还包括存储查询子系统, 该子系统包括:

交互记录存储模块, 用于记录存储数字员工与用户的交互信息, 包括语音和文字;

交互记录查询模块,用于查询所述交互记录存储模块存储的信息;
所述服务端还包括权限控制子系统,用于控制用户访问服务端各子系统的权限。

10.根据权利要求2至9任意一项所述的适用于银行的提升现场体验感的人工智能交互系统,其特征在于,所述数据信息配置子系统包括:

第一配置模块,用于配置所述数字员工的信息;

第二配置模块,用于配置所述应答库的信息;

第三配置模块,用于配置所述安全验证子系统进行身份安全验证的验证信息;

第四配置模块,用于配置所述数据切换子系统的切换数据信息;

第五配置模块,用于配置所述休眠子系统进入或暂停休眠状态的信息;

第六配置模块,用于配置所述存储查询子系统实现记录存储的信息;

第七配置模块,用于配置所述权限控制子系统的权限信息。

一种适用于银行的提升现场体验感的人工智能交互系统

技术领域

[0001] 本发明属于金融AI领域,特别涉及一种适用于银行的提升现场体验感的人工智能交互系统。

背景技术

[0002] 随着人工智能技术的蓬勃发展,人工智能的落地场景变得更加丰富。各种智能终端越来越多,使得人工智能和银行人工服务的结合上升到了一个新的起点。传统银行业务的数字化、移动化快速发展,将为传统银行等金融行业在人工服务方面带来更广阔的发展空间,多样化的业务办理方式可以让用户获得更多的服务体验以及提高业务的办理效率、节省客户办理时间。

[0003] 目前,传统银行业等金融领域依然大量采用文本提示、人工讲解等为客户服务,其中文本提示包括短信、邮件等,人工讲解包括人工语音,现场讲解等方式,常见的有在线客服和大堂经理,在工作时间内为客户讲解金融业务知识、介绍金融产品、回答客户疑问、解决客户难题等,但由于银行服务人员数量限制,当客户较多时,会造成客户排队等候、客服电话无法接入、非工作时间问题无法及时解决等问题,增加了客户对银行服务的不认同感,一定程度上可能造成客户的流失,进而增加银行重新获客成本。当下已经有银行推广了一些交互系统想要克服上述问题,但都还只是停留在智能客服系统层面,能够实现基本的问答,其问答系统主要分为面向任务、面向知识、面向聊天三个方向,技术原理是一种基于知识图谱与信息检索相结合的方法,检索出匹配度最高的一个答案返回给客户,从而实现智能问答的模式。

[0004] 现有的智能客服系统存在以下问题:1、无法精准的识别客户的意图:现有的智能客服系统只能通过跳转至原生页面实现基础业务的办理,如转账,查余额、查积分等,智能客服系统绝大部分是通过检索客户的语音文字,当检索到匹配的关键词如“转账”,则直接跳转至转账原生页面,只是减少了客户自行点击转账按钮跳转至页面的操作;2、业务场景割裂、无法形成全流程的闭环管理:智能语音客服系统虽然可以通过交互实现闲聊、简单业务办理的功能,但是业务之间是相互独立的,跳转至原生页面办理业务后不具备带参数返回智能语音客服系统的功能,因此,对于客户而言,需要重新发起一轮新的交互进行新的业务办理;3、无法实现角色的选择、场景感弱:现有智能客服系统均没有真正的形象作为支撑,而是冠以“机器人”的称号为客户提供服务,没有角色代入感和现场体验感,一定程度上降低了客户的信任度;4、启动时安全性较低无法验证:

智能客服系统在执行语音交互操作时不具备身份识别的能力,安全性有待提高。

发明内容

[0005] 本发明的目的在于针对上述现有技术存在的问题,提供一种适用于银行的提升现场体验感的人工智能交互系统。

[0006] 实现本发明目的的技术解决方案为:一种适用于银行的提升现场体验感的人工智

能交互系统,应用于银行业务办理终端,所述系统包括客户端和服务端;

所述客户端,用于通过数字员工与用户进行面对面交互,接收用户语音输入信息并发送至服务端,响应并展示服务端反馈的应答内容;所述用户语音输入信息包括业务办理需求信息;

所述服务端,用于响应所述用户语音输入信息进行应答,还用于配置所述客户端和服务端的信息。

[0007] 进一步地,所述服务端包括:

应答库构建子系统,用于构建客户端与用户交互的知识应答库,该知识应答库包括业务办理应答库;

智能通信子系统,用于提取并过滤所述用户语音输入信息中的噪声信息,并将过滤后的用户语音输入信息发送至ASR子系统;

ASR子系统,用于将所述用户语音输入信息解析为文字,并发送至信息解析子系统;

信息解析子系统,用于解析所述文字,识别用户需求,包括业务办理需求,并将该需求信息发送至应答响应子系统;

应答响应子系统,用于响应所述需求信息,从所述应答库中提取应答内容并推送至客户端;所述应答内容包括数字员工应答语音、文字、驱动信息,以及业务交互窗口;所述数字员工驱动信息包括动作、表情和口型;

数据信息配置子系统,用于管理、配置客户端和服务端的数据信息。

[0008] 进一步地,所述应答库构建子系统包括:

第一应答元素子库生成模块,用于生成每一项业务的办理流程,包括多级业务交互过程;

第二应答元素子库生成模块,用于将每一级业务交互过程转换为对应的交互语音和文字,生成多级交互语音和文字;

第三应答元素子库生成模块,用于生成每一级业务交互过程对应的业务交互窗口;

第四应答元素子库生成模块,用于生成每一级交互语音和文字对应的数字员工驱动信息;

映射关系建立模块,用于建立所述每一级业务交互过程与对应的交互语音和文字之间的映射关系,每一级业务交互过程与对应的业务交互窗口之间的映射关系,以及每一级交互语音和文字与其对应的数字员工驱动信息之间的映射关系。

[0009] 进一步地,所述应答响应子系统包括:

第一响应模块,用于驱动所述第一应答元素子库响应所述需求信息,触发该需求信息对应的某一级业务交互过程,并生成第一触发指令;

第二响应模块,用于驱动所述第二应答元素子库响应所述第一触发指令,触发所述某一级业务交互过程对应的某一级交互语音和文字,并将所述某一级交互语音和文字发送至应答输出模块,同时生成第二触发指令;

第三响应模块,用于驱动所述第三应答元素子库响应所述第二触发指令,触发所述某一级交互语音和文字对应的业务交互窗口,并将该业务交互窗口发送至应答输出模块;

第四响应模块,用于驱动所述第四应答元素子库响应所述第二触发指令,触发所述某一级交互语音和文字对应的数字员工驱动信息,并将该数字员工驱动信息发送至应答输出

模块,同时生成第三触发指令;

应答输出模块,用于响应所述第三触发指令,将所述需求信息对应的所述某一级交互语音和文字、业务交互窗口,以及数字员工驱动信息同步推送至客户端。

[0010] 进一步地,所述服务端还包括安全验证子系统,该子系统包括:

安全身份库构建模块,用于构建包括用户音频数据的安全身份库;所述安全身份库包括注册库和验证库,所述注册库包括用户注册的语音音频,所述验证库包括用户对某一指定文本录入的语音音频;

第一安全验证模块,用于在所述银行业务办理终端进行初次用户身份安全验证之后进行二次身份安全验证,验证通过后启动所述人工智能交互系统;

第二安全验证模块,用于根据用户语音输入信息实时判别用户是否为与银行账户关联的安全身份,若不是安全身份,触发锁定所述人工智能交互系统,并进一步触发锁定所述银行业务办理终端。

[0011] 进一步地,所述客户端包括权限设置子系统,用于设置多个与银行账户关联的安全身份,每一个安全身份信息包括其语音信息。

[0012] 进一步地,所述服务端还包括数据切换子系统,该子系统包括:

角色切换模块,用于响应所述需求信息,切换与该需求信息匹配的数字员工角色;

场景切换模块,用于响应所述需求信息,切换与该需求信息匹配的业务交互场景。

[0013] 进一步地,所述服务端还包括打断子系统,用于在数字员工输出应答语音的过程中,实时接收新的用户语音输入信息,在接收到新的用户语音输入信息时暂停当前应答语音的输出,同时生成并发送触发命令至ASR子系统,对新的用户语音输入信息进行新一轮响应。

[0014] 进一步地,所述服务端还包括休眠子系统,用于实时监测用户语音输入信息,在未监测到用户语音输入信息的时常超过预设阈值时,控制所述人工智能交互系统进入休眠状态;在休眠过程中,若监测到用户语音输入信息,即时唤醒人工智能交互系统恢复至休眠前的状态。

[0015] 进一步地,所述服务端还包括存储查询子系统,该子系统包括:

交互记录存储模块,用于记录存储数字员工与用户的交互信息,包括语音和文字;

交互记录查询模块,用于查询所述交互记录存储模块存储的信息。

[0016] 进一步地,所述服务端还包括权限控制子系统,用于控制用户访问服务端各子系统的权限。

[0017] 进一步地,所述数据信息配置子系统包括:

第一配置模块,用于配置所述数字员工的信息;

第二配置模块,用于配置所述应答库的信息;

第三配置模块,用于配置所述安全验证子系统进行身份安全验证的验证信息;

第四配置模块,用于配置所述数据切换子系统的切换数据信息;

第五配置模块,用于配置所述休眠子系统进入或暂停休眠状态的信息;

第六配置模块,用于配置所述存储查询子系统实现记录存储的信息;

第七配置模块,用于配置所述权限控制子系统的权限信息。

[0018] 本发明与现有技术相比,其显著优点为:1)通过形象建模技术构建数字员工,相比

于其他智能客服系统,给客户提供了更有温度的、更人性化的服务;2)能够通过自然语言理解、上下文理解和提槽技术,在不断的交互过程中充分理解用户的意图,与用户进行多轮交互(相比于现有智能客服系统只能直接跳转至原生页面进行业务办理,本发明能够直接实现关键业务要素的提取,完成业务办理的相关信息填写,只需用户在原生页面确认即可,提高了客户办理业务的便捷度,且跳转至原生页面办理业务后代参返回,可以持续不断的进行新的业务办理,不需要从最初启动交互开始发起完全新一轮的交互),并由数字员工通过多媒体资源展示给客户查看,更贴近人工业务办理的模式,更加拟人智能化,且相比于人工服务,业务办理的效率更高;3)可以实现多场景切换,包含厅堂场景、转账场景、理财场景等,带给客户身临其境的体验,一定程度上增强用户的信任度;4)能够实现随时智能打断,及时响应用户的需求,使人机交互更加高效、快捷、自然,有助于增强客户体验;5)具有更为强大的安全保密功能,通过多级安全验证,实时全方位保证客户的信息安全和账户安全;6)提出的特征提取方法综合考虑了音频的音调、梅尔频率倒谱系数以及纹理特征,能在一定程度上提高后续相似度的计算精度,进而提高安全验证精度;7)在通过语音判定安全身份时,采用了两级不同方式的安全判断,第一级安全判断过程相对简单,若第一级直接判别结束,可以降低算法的复杂度;第二级安全判断基于深度学习方法构建相似度模型,计算精度更高。

[0019] 下面结合附图对本发明作进一步详细描述。

附图说明

[0020] 图1为一个实施例中适用于银行的提升现场体验感的人工智能交互系统的原理图。

[0021] 图2为一个实施例中适用于银行的提升现场体验感的人工智能交互系统的结构示意图。

[0022] 图3为一个实施例中应答库构建子系统结构图。

[0023] 图4为一个实施例中应答响应子系统结构图。

[0024] 图5为一个实施例中数字营业厅转账流程图。

[0025] 图6为一个实施例中数字营业厅信用卡账单查询流程图。

[0026] 图7为一个实施例中转账业务中某一级业务交互过程对应的业务交互窗口示意图。

[0027] 图8为一个实施例中判别用户是否为与银行账户关联的安全身份的流程图。

[0028] 图9为一个实施例中构建基于深度学习的声音相似度计算模型的流程图。

[0029] 图10为一个实施例中特征提取的流程图。

[0030] 图11为一个实施例中特征提取的流程图。

具体实施方式

[0031] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0032] 在一个实施例中,结合图1和图2,提供了一种适用于银行的提升现场体验感的人

工智能交互系统,应用于银行业务办理终端,所述系统包括客户端和服务端;

这里,人工智能交互系统也可以称为数字营业厅,银行业务办理终端包括手机银行APP等,所述数字营业厅以SDK的形式嵌入至手机银行APP内。客户通过点击手机银行APP内的数字营业厅入口及一级页面的悬浮框可以跳转进入数字营业厅,进入后会建立WEBRTC会话通道。

[0033] 所述客户端,用于通过数字员工与用户进行面对面语音交互,接收用户语音输入信息并发送至服务端,响应并展示服务端反馈的应答内容;所述用户语音输入信息包括业务办理需求信息;

这里,数字员工是由一个人的一张或者几张不同角度的照片,若干句语音,制作一个有表情、有动作、能说话的‘活人’,以此来实现一个人的‘数字克隆体’。

[0034] 这里,所述语音交互的方式支持通话模式,一直保持会话状态,实现实时会话;支持对讲模式,即可通过长按语音按钮进行对话,防止环境噪音被实时抓取分析,提高应答的效率和准确率。

[0035] 所述服务端,用于响应所述用户语音输入信息进行应答,还用于配置所述客户端和服务端的信息。

[0036] 进一步地,在其中一个实施例中,所述服务端包括:

应答库构建子系统,用于构建客户端与用户交互的知识应答库,该知识应答库包括业务办理应答库;

这里,知识应答库还包括与业务不相关的日常交流应答库。

[0037] 智能通信子系统,用于提取并过滤所述用户语音输入信息中的噪声信息,并将过滤后的用户语音输入信息发送至ASR子系统;

ASR子系统,用于将所述用户语音输入信息解析为文字,并发送至信息解析子系统;

这里,语音识别技术(ASR, Automatic Speech Recognition),是让机器通过识别和理解过程使之听懂人类语言的技术。语音识别技术是信息技术中人机交互的关键技术,目前已经在呼叫中心、电信增值业务、企业信息化系统中有着广泛的应用。随着语音识别在语音搜索、语音控制等全新应用领域的深入应用,语音识别技术被业界权威人士誉为有可能引发人机界面领域革命的关键技术。近年来,语音识别技术取得了足够的进步,本发明优选采用了当前最为先进的DNN(深层神经网络)技术,为构建高价值自助语音服务提供强大、稳定、易用的核心动力引擎。

[0038] 信息解析子系统,用于解析所述用户语音输入信息和文字,识别用户需求,包括业务办理需求,并将该需求信息发送至应答响应子系统;

这里,解析文字具体采用自然语言处理NLP技术实现,通过语义理解、上下文理解等识别用户需求。

[0039] 采用该种方式,针对用户对同一需求的不同表达话语,均能实现充分理解用户的意图,提高了解析的准确率。例如对于转账业务需求,用户无需准确的描述出“转账”“汇款”等专业词汇,即使描述为“我要转钱”“我要汇钱”等语句,信息解析子系统也能充分理解用户的意图即为“转账”“汇款”。

[0040] 采用该种方式,信息解析子系统可以直接解析、应答用户的最终需求,该最终需求可能是某一业务办理过程中的某一环节,而不是仅仅解析出要办理的业务,然后从业务办

理的初始环节开始依次询问应答,直至到达所述最终需求,更具针对性,提高了信息解析的效率,进而提高后续应答的效率。例如用户的需求为“我要转账100元给张三”,信息解析子系统会直接解析出转账和转账对象,然后直接应答出转账对象交互窗口如图7所示。

[0041] 应答响应子系统,用于响应所述需求信息,从所述应答库中提取应答内容并推送至客户端;所述应答内容包括数字员工应答语音、文字、驱动信息,以及业务交互窗口;所述数字员工驱动信息包括动作、表情和口型;

数据信息配置子系统,用于管理、配置客户端和服务端的数据信息。

[0042] 进一步地,在其中一个实施例中,结合图3,所述应答库构建子系统包括:

第一应答元素子库生成模块,用于生成每一项业务的办理流程,包括多级业务交互过程;

这里,每一级业务交互过程对应业务办理中的一次“问答”,问、答的主体可以为用户,也可以为数字员工。所有的“问答”交替连续,实现完整的业务办理流程。如图5和图6分别为转账、信用卡账单查询的流程,每一次判别都对应一次“问答”。

[0043] 第二应答元素子库生成模块,用于将每一级业务交互过程转换为对应的交互语音和文字,生成多级交互语音和文字;

这里,将每一级业务交互过程转换为对应的交互语音是通过语音合成TTS技术(又称为文语转化)实现,该技术是将文本信息转化为语音数据的技术,涉及声学、语言学、数字信号处理、多媒体等多种前沿的高新科技,是中文信息处理领域的一项前沿技术。本发明优选采用人工智能语音合成系统 STTS实现。语音合成系统的合成引擎在完成文本到语音数据的转化过程中可以简单分解为两个步骤的处理:文本先经过前端的语法分析,通过词典和规则的处理,获得格式规范,携带语法层次的信息,传送到后端;后端在前端分析结果的基础上,经过对韵律的分析处理,获得语音的时长、音高等韵律信息,再根据这些信息生成语音单元(临时合成模式)或在音库中挑选最合适的语音单元(提前录好拼接模式),通过调整和拼接语音单元,即能得到最终的语音数据。

[0044] 第三应答元素子库生成模块,用于生成每一级业务交互过程对应的业务交互窗口,如图5所示为转账业务中业务交互过程“用户:我要转账100元给张三;数字员工:找到张三的以下账户信息,请问转给哪个账户”对应的业务交互窗口。

[0045] 第四应答元素子库生成模块,用于生成每一级交互语音和文字对应的数字员工驱动信息;

映射关系建立模块,用于建立所述每一级业务交互过程与对应的交互语音和文字之间的映射关系,每一级业务交互过程与对应的业务交互窗口之间的映射关系,以及每一级交互语音和文字与其对应的数字员工驱动信息之间的映射关系。

[0046] 进一步地,在其中一个实施例中,结合图4,所述应答响应子系统包括:

第一响应模块,用于驱动所述第一应答元素子库响应所述需求信息,触发该需求信息对应的某一级业务交互过程,并生成第一触发指令;

这里,通过上述所述自然语言处理NLP技术对所述需求信息进行上下文语义理解,获得该需求信息对应的某一级业务交互过程,该级业务交互可以是数字员工对用户问题的回答,也可以为向用户的主动提问。

[0047] 第二响应模块,用于驱动所述第二应答元素子库响应所述第一触发指令,触发所

述某一级业务交互过程对应的某一级交互语音和文字,并将所述某一级交互语音和文字发送至应答输出模块,同时生成第二触发指令;

第三响应模块,用于驱动所述第三应答元素子库响应所述第二触发指令,触发所述某一级交互语音和文字对应的业务交互窗口,并将该业务交互窗口发送至应答输出模块;

第四响应模块,用于驱动所述第四应答元素子库响应所述第二触发指令,触发所述某一级交互语音和文字对应的数字员工驱动信息,并将该数字员工驱动信息发送至应答输出模块,同时生成第三触发指令;

这里,上述第三响应模块、第四响应模块可以不限于上述顺序执行,可以同步响应,也可以由第四响应模块先响应。

[0048] 这里,第三触发指令可以由第三相应模块生成。

[0049] 应答输出模块,用于响应所述第三触发指令,将所述需求信息对应的所述某一级交互语音和文字、业务交互窗口,以及数字员工驱动信息同步推送至客户端。

[0050] 这里示例性地,结合图5和图7,以转账为例,假设用户的初始需求是“我要转账100元给张三”,则第一响应模块通过话术分析、语义理解,触发的某一级业务交互过程为:数字员工将向用户提问“找到张三的以下账户信息,请问转给哪个账户”,之后第二响应模块至第四响应模块以及应答输出模块会相继触发,获取“找到张三的以下账户信息,请问转给哪个账户”对应的应答语音、文字和交互窗口并同步推送至客户端,实现数字员工与用户对该级业务交互过程的交互。之后用户从主观角度会针对数字员工的提问输入新的语音信息,由此会触发智能通信子系统、ASR子系统、信息解析子系统以及应答响应子系统进行新一轮的响应,以此循环实现多轮交互,直至用户退出本系统。

[0051] 采用本实施例的方案,在不断的交互过程中充分理解用户的意图,与用户进行多轮交互,并由数字员工通过多媒体资源展示给客户查看,更贴近人工业务办理的模式,更加拟人智能化,且相比于人工服务,业务办理的效率更高。

[0052] 进一步地,在其中一个实施例中,所述服务端还包括安全验证子系统,该子系统包括:

安全身份库构建模块,用于构建包括用户音频数据的安全身份库;所述安全身份库包括注册库和验证库,所述注册库包括用户注册的语音音频,所述验证库包括用户对某一指定文本录入的语音音频;

第一安全验证模块,用于在所述银行业务办理终端进行初次用户身份安全验证之后进行二次身份安全验证,验证通过后启动所述人工智能交互系统;

这里,二次身份安全验证包括声纹验证或人脸识别验证等等方式。

[0053] 第二安全验证模块,用于根据用户语音输入信息实时判别用户是否为与银行账户关联的安全身份,若不是安全身份,触发锁定所述人工智能交互系统,并进一步触发锁定所述银行业务办理终端。

[0054] 采用本实施例的方案,能够实现实时、多重保障用户个人信息和账户信息的安全。

[0055] 进一步地,在其中一个实施例中,结合图8,所述根据用户语音输入信息实时判别用户是否为与银行账户关联的安全身份,具体过程包括:

从所述安全身份库中提取与当前银行账户关联的用户的某一语音音频,对该音频以及当前用户输入的语音音频分别进行特征提取,获得各自对应的音频向量;

利用余弦相似度公式求取两个音频向量的相似度,记为第一相似度;

判断所述第一相似度是否高于第一预设阈值,若高于,则判定当前用户为安全身份,否则执行下一步;

构建基于深度学习的声音相似度计算模型;结合图9,具体过程包括:

从所述注册库中提取用户A的语音音频,对该音频进行特征提取,获得对应的音频向量,记为音频特征 a ;

对所述验证库中每一个用户的语音音频进行特征提取,获得对应的音频向量,分别记为音频特征 b_1, b_2, \dots, b_n ,构成音频特征集 B ; n 为验证库中语音音频的数目;

从所述音频特征集 B 中随机抽取若干音频特征,每个音频特征分别与所述音频特征 a 构成一个样本,形成训练样本集;

将所述训练样本集输入LSTM神经网络进行训练,生成声音相似度计算模型;

将所述当前用户输入的语音音频的音频向量,以及音频特征集 B 中该用户对应的音频特征输入至所述声音相似度计算模型,输出两者的相似度,记为第二相似度;

判断所述第二相似度是否高于第二预设阈值,若高于,则判定当前用户为安全身份,否则判定当前用户为不安全身份。

[0056] 这里优选地,第一预设阈值、第二预设阈值分别设为80%、50%。

[0057] 采用本实施例的方案,采用了两级不同方式的安全判断,第一级安全判断过程相对简单,若第一级直接判别结束,可以降低算法的复杂度;第二级安全判断基于深度学习方法构建相似度模型,计算精度更高。

[0058] 进一步地,在其中一个实施例中,结合图10,所述特征提取的具体过程包括:

提取语音音频中的音调特征,具体过程包括:

利用YIN算法处理语音音频信号;

检测YIN算法处理后的信号的基频的周期即音调特征,具体包括:

提取YIN算法处理后的信号的极小值点;

判断是否存在极小值点小于预设判决阈值,若否,则将其中最小的极小值点与所述预设判决阈值的偏移量作为基频的周期;若是,则执行下述步骤;

获取所有小于预设判决阈值的极小值点,并将极小值的相对绝对差值小于预设误差的极小值点归入同一极小值点序列,获得若干极小值点序列,且对每个序列中的极小值点进行升序排列;

根据每个极小值点序列中极小值点的数目,求取该极小值点序列的权重;

获取权重值最大的极小值点序列中最小的极小值点,将该最小极小值点与所述预设判决阈值的偏移量作为基频的周期;

提取语音音频的梅尔频率倒谱系数;

提取语音音频的纹理特征;纹理特征可以包括:均匀局部二值模式、局部二值模式的直方图傅里叶、局部相位量化等等;

将所述音调特征、梅尔频率倒谱系数及纹理特征拼接,形成音频向量。

[0059] 进一步地,在其中一个实施例中,结合图11,所述特征提取的具体过程包括:

提取语音音频中的音调特征,具体过程包括:

构建函数：

$$HB_{\tau} = \frac{MI_{\tau} \cdot SU1_{\tau} \cdot SU2_{\tau}}{MA_{\tau} \cdot (L - \tau)^2}$$

$$YD_{\tau} = d_{\tau}^{HB_{\tau}}$$

其中， d_{τ} 为AMDF函数变换公式，公式为：

$$d_{\tau} = \frac{1}{L - \tau} \sum_{i=0}^{L - \tau - 1} |x_i - x_{i + \tau}|$$

式中， x 表示语音音频信号， L 为数据帧的长度， τ 为偏移量，其取值范围为 $0 \sim L - 1$ ； x_i 表示数据帧中 i 位置处的信号；

MI_{τ} 为两帧信号中各自数据最小值的总和，公式为：

$$MI_{\tau} = \sum_{i=0}^{L - \tau - 1} [\text{sgn}(x_i, x_{i + \tau}) \cdot \text{Min}(|x_i|, |x_{i + \tau}|)]$$

$$\text{sgn}(x_i, x_{i + \tau}) = \begin{cases} 1, & x_i \cdot x_{i + \tau} \geq 0 \\ 0, & \text{其他} \end{cases};$$

MA_{τ} 为两帧信号中各自数据最大值的总和，公式为：

$$MA_{\tau} = \sum_{i=0}^{L - \tau - 1} \{ \text{sgn}(x_i, x_{i + \tau}) \cdot \text{Max}(|x_i|, |x_{i + \tau}|) + [1 - \text{sgn}(x_i, x_{i + \tau})] \cdot (|x_i| + |x_{i + \tau}|) \}$$

$SU1_{\tau}$ 为两帧信号中对应位置数据值异号的数据点总数，公式为：

$$SU1_{\tau} = \sum_{i=0}^{L - \tau - 1} \text{sgn} |x_i, x_{i + \tau}|$$

$SU2_{\tau}$ 为两帧信号中相邻数据点呈正向变化的数据点总数与相邻数据点呈负向变化的数据点总数的最小值，公式为：

$$SU2_{\tau} = \text{Min} \left\{ \left[\sum_{i=0}^{L - \tau - 1} \text{sgn}(x_i - x_{i - 1}, x_{i + \tau} - x_{i + \tau - 1}) \right] \left[\sum_{i=0}^{L - \tau - 1} \text{sgn}(x_{i + \tau} - x_{i + \tau - 1}, x_i - x_{i - 1}) \right] \right\}$$

利用所述函数对语音音频信号进行变换；

检测函数变换后的信号的基频的周期即音调特征，具体为：

提取函数变换后的信号的极小值点；

判断是否存在极小值点小于预设判决阈值，若否，则将其中最小的极小值点与所述预设判决阈值的偏移量作为基频的周期；若是，则执行下述步骤；

获取所有小于预设判决阈值的极小值点，并将极小值的相对绝对差值小于预设误差的

极小值点归入同一极小值点序列,获得若干极小值点序列,且对每个序列中的极小值点进行升序排列;

根据每个极小值点序列中极小值点的数目,求取该极小值点序列的权重;

获取权重值最大的极小值点序列中最小的极小值点,将该最小极小值点与所述预设判决阈值的偏移量作为基频的周期;

提取语音音频的梅尔频率倒谱系数;

提取语音音频的纹理特征;

将所述音调特征、梅尔频率倒谱系数及纹理特征拼接,形成音频向量。

[0060] 采用上述实施例的方案,综合考虑了音频的音调、梅尔频率倒谱系数以及纹理特征,能在一定程度上提高后续相似度的计算精度,进而提高安全验证精度。

[0061] 进一步地,在其中一个实施例中,所述客户端包括权限设置子系统,用于设置多个与银行账户关联的安全身份,每一个安全身份信息优选包括其语音信息,这是由于本发明为倾向于语音交互的系统。

[0062] 这里,安全身份信息也可以是声纹、面容等。

[0063] 进一步地,在其中一个实施例中,所述服务端还包括数据切换子系统,该子系统包括:

角色切换模块,用于响应所述需求信息,切换与该需求信息匹配的数字员工角色;

这里,若需求为业务咨询或其他日常咨询,匹配数字员工角色可以为大堂经理;若需求为业务办理,匹配数字员工角色可以为柜员或业务办理员。

[0064] 场景切换模块,用于响应所述需求信息,切换与该需求信息匹配的业务交互场景。

[0065] 这里,业务交互场景根据银行网点实际场景构建,根据功能划分涵盖多种场景,包括迎宾、业务办理、理财等营销产品推荐等等,场景可以为动态视频或者宣传图等等。所述迎宾场景在人工智能交互系统唤醒时自动切换。

[0066] 采用本实施例的方案,能够提高用户的现场体验感(包括愉悦感、信任感等等)。

[0067] 进一步地,在其中一个实施例中,所述服务端还包括打断子系统,用于在数字员工输出应答语音的过程中,实时接收新的用户语音输入信息,在接收到新的用户语音输入信息时暂停当前应答语音的输出,同时生成并发送触发命令至ASR子系统,对新的用户语音输入信息进行新一轮响应。

[0068] 采用本实施例的方案,相比于现有的打断方式如:小度要借助称呼触发语句“小度”进行打断,siri必须要通过触发对话按钮才能进行打断,本发明的实时打断方式,能够实现以无接触、更直接的方式进行话语打断,及时响应用户的需求,使人机交互更加高效、快捷、自然,有助于增强客户体验。此外,即使用户不清楚系统操作手册和触发语句,也可以无障碍使用,更加拟人智能化,如同正常的人与人交流,降低了系统的使用门槛。

[0069] 进一步地,在其中一个实施例中,所述服务端还包括休眠子系统,用于实时监测用户语音输入信息,在未监测到用户语音输入信息的时常超过预设阈值时,控制所述人工智能交互系统进入休眠状态;在休眠过程中,若监测到用户语音输入信息,即时唤醒人工智能交互系统恢复至休眠前的状态。

[0070] 这里,所述休眠状态可以为人工智能交互系统锁定,也可以为推送活动或产品宣传至客户端,进行滚动播放。

[0071] 采用本实施例的方案,使得业务办理状态不会因用户突发事宜自动中断而导致多次重启业务办理,能够实现保持当前业务办理状态不间断且被及时唤醒(除非用户主动提出退出当前业务办理的流程),提高了用户的体验感。

[0072] 采用本实施例的方案,通过锁定可以进一步保障用户账户信息的安全。

[0073] 采用本实施例的方案,通过宣传推送,能够充分利用资源实时进行银行业务、产品的推销,一定程度上提高了人工推销的成本,以及推销的成功率。

[0074] 进一步地,在其中一个实施例中,所述服务端还包括存储查询子系统,该子系统包括:

交互记录存储模块,用于记录存储数字员工与用户的交互信息,包括语音和文字;

交互记录查询模块,用于查询所述交互记录存储模块存储的信息。

[0075] 采用本实施例方案,有利于后续对客户的进一步分析,如对客户意向进行分类,获得客户最频繁的需求,以便针对性的回访跟进客户,提升客户的好感度。

[0076] 进一步地,在其中一个实施例中,所述服务端还包括权限控制子系统,用于控制用户访问服务端各子系统的权限。

[0077] 采用本实施例的方案,能够充分保障整个人工交互系统的安全性以及用户的隐私。

[0078] 进一步地,在其中一个实施例中,所述数据信息配置子系统包括:

第一配置模块,用于配置所述数字员工的信息;

这里,数字员工的信息包括数字员工的角色身份、形象等。

[0079] 第二配置模块,用于配置所述应答库的信息;

这里,应答库的信息包括应答库本身的规格、功能区域划分、数据记载方式等,应答语音单元的数量、音色等,应答文字的字体、排列方式等,业务交互窗口的显示方式(包括动画、色彩、显示位置等)以及显示内容,数字员工的表情、动作、口型等文件。

[0080] 第三配置模块,用于配置所述安全验证子系统进行身份安全验证的验证信息;

这里,不同验证方式对应不同的验证信息,验证信息可以包括用户名和密码,面容、指纹、声音等的一种或多种。

[0081] 第四配置模块,用于配置所述数据切换子系统的切换数据信息;

这里,切换数据信息包括所有的待切换数据(角色、场景等),以及切换的指令、切换的方式等。

[0082] 第五配置模块,用于配置所述休眠子系统进入或暂停休眠状态的信息;

这里,信息包括所述预设阈值的具体数值,进入或暂停休眠状态的触发指令,以及休眠状态的具体方式等。

[0083] 第六配置模块,用于配置所述存储查询子系统实现记录存储的信息;

这里,信息包括需要存储的数据内容(记录日期、时间点以及业务办理事项等)、数据存储的格式或方式等(例如某一段时间的数据会形成一个数据包,便于后期数据查找和分析)。

[0084] 第七配置模块,用于配置所述权限控制子系统的权限信息。

[0085] 这里,权限信息包括每个服务端用户对应的访问子系统的权限(例如是否有权限访问,存在权限的有效时间段等等),权限的类型(只读、可修改等等)。

[0086] 上述实施例阐明的系统、模块,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。为了描述方便,描述以上系统时以功能分为各种模块分别进行描述。当然,在实施本说明书一个或多个时可以把各模块的功能在同一个或多个软件和/或硬件中实现,也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。以上所描述的系统实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个模块或单元组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性机械或其它的形式。

[0087] 以上显示和描述了本发明的基本原理、主要特征及优点。本行业的技术人员应该了解,本发明不受上述实施例的限制,上述实施例和说明书中描述的只是说明本发明的原理,在不脱离本发明精神和范围的前提下,本发明还会有各种变化和改进,这些变化和进步都落入要求保护的本发明范围内。本发明要求保护范围由所附的权利要求书及其等效物界定。



图1

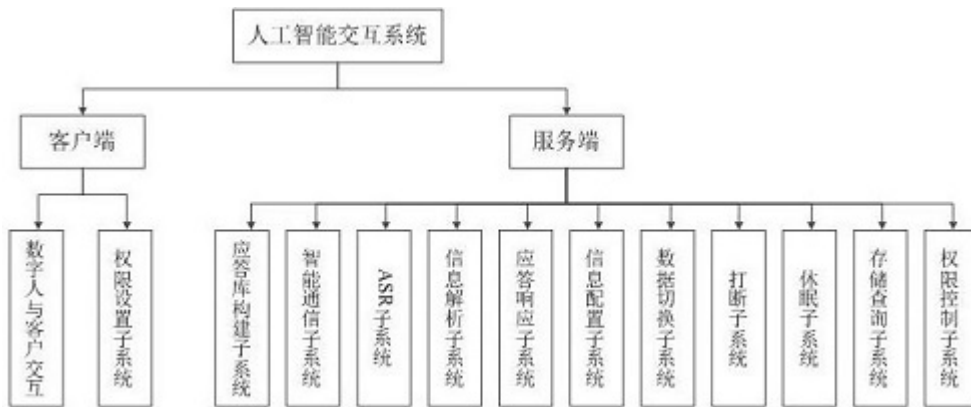


图2

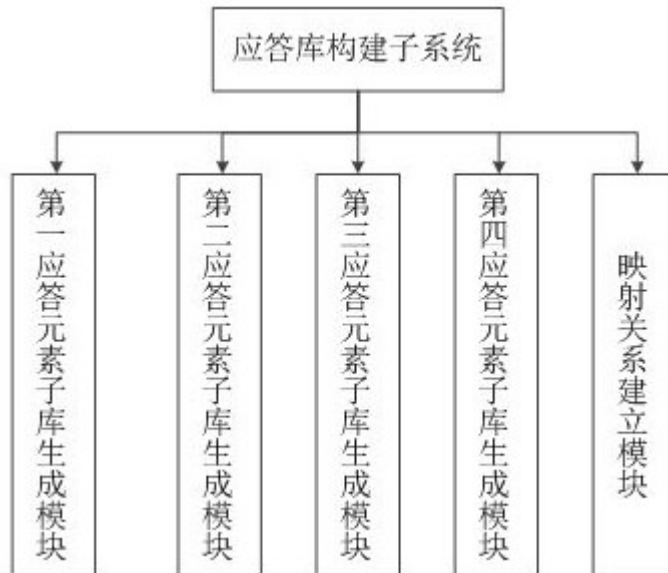


图3

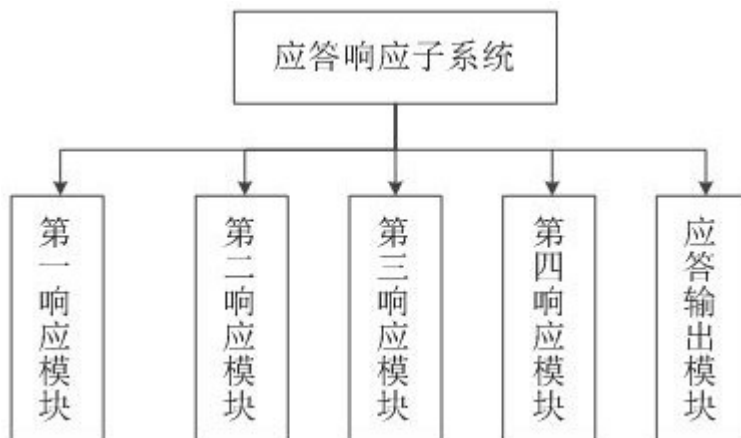


图4

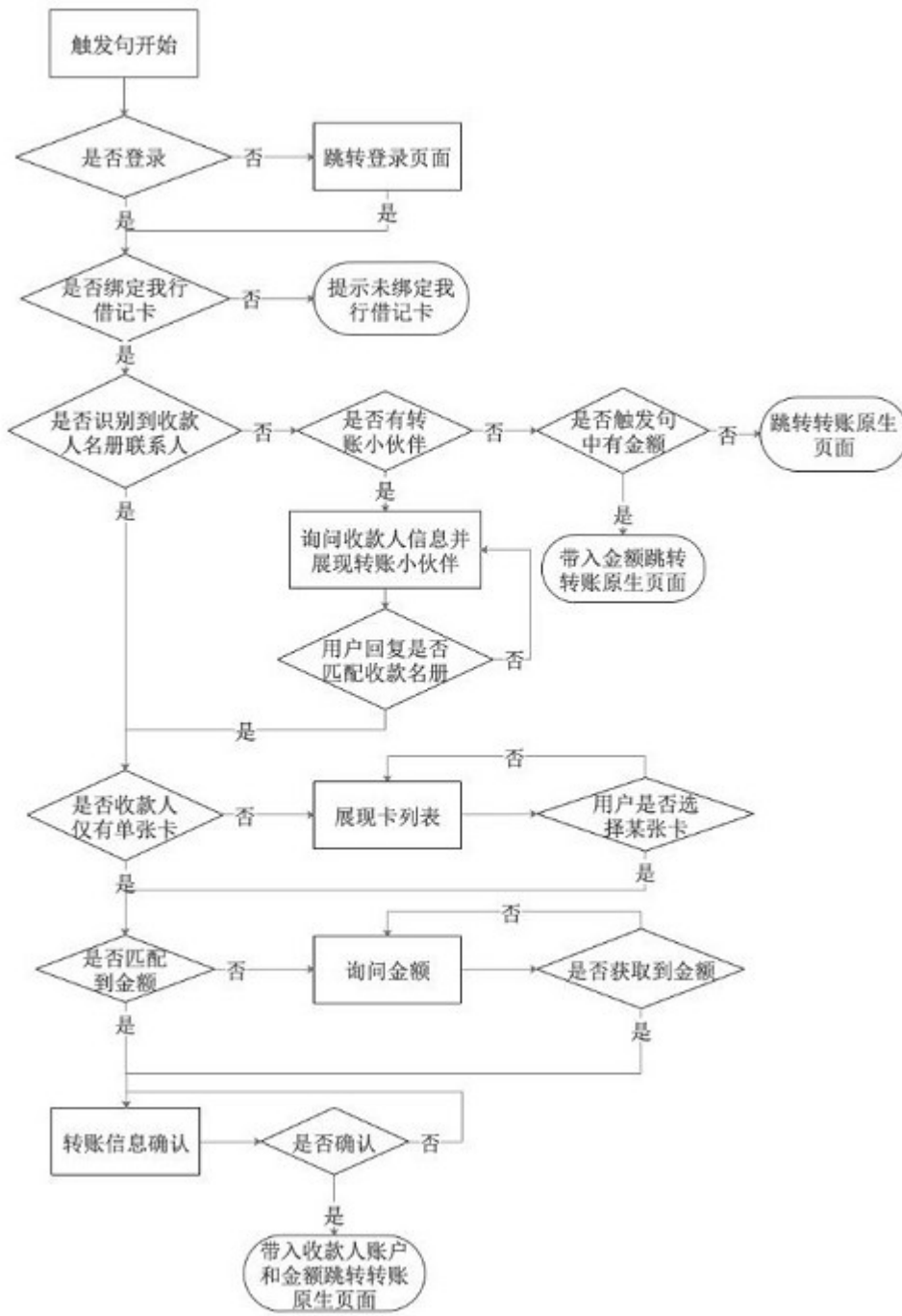


图5

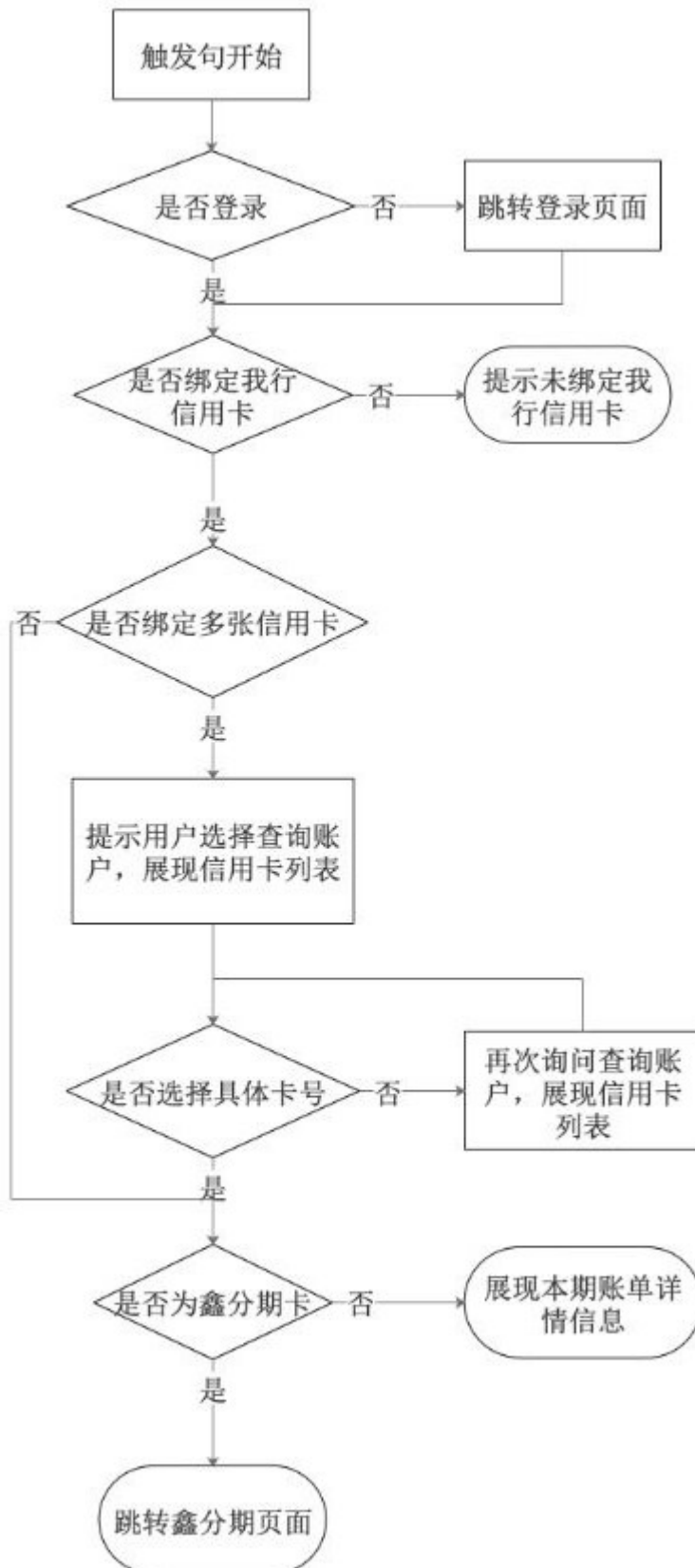


图6



图7

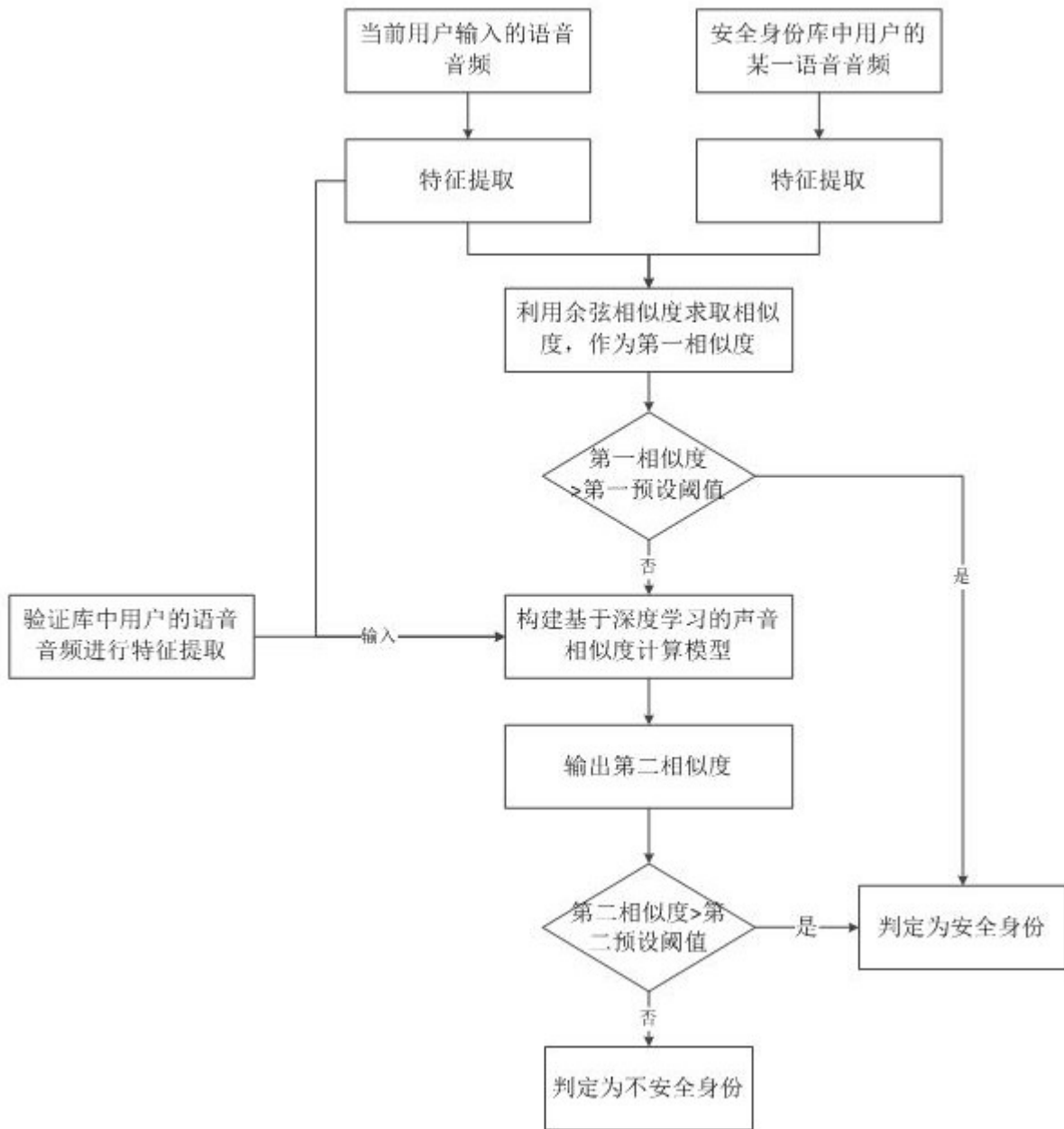


图8

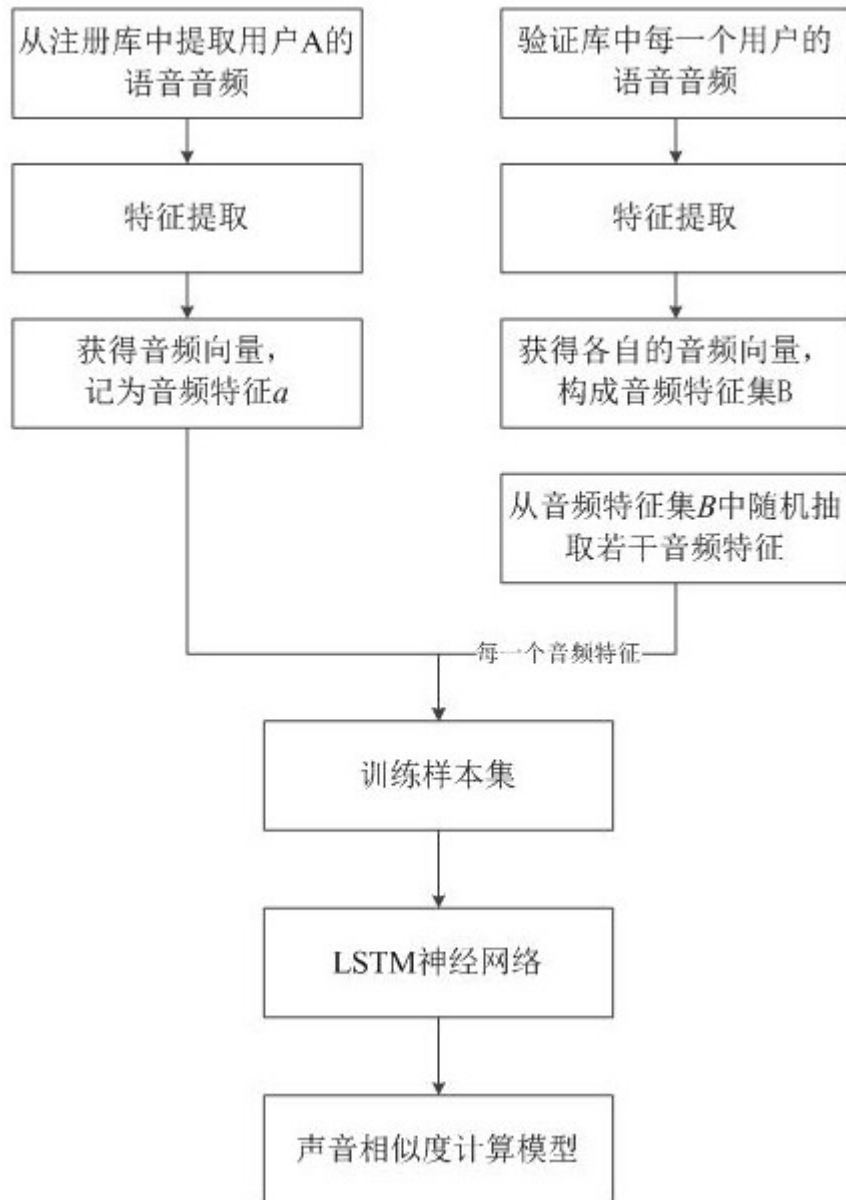


图9

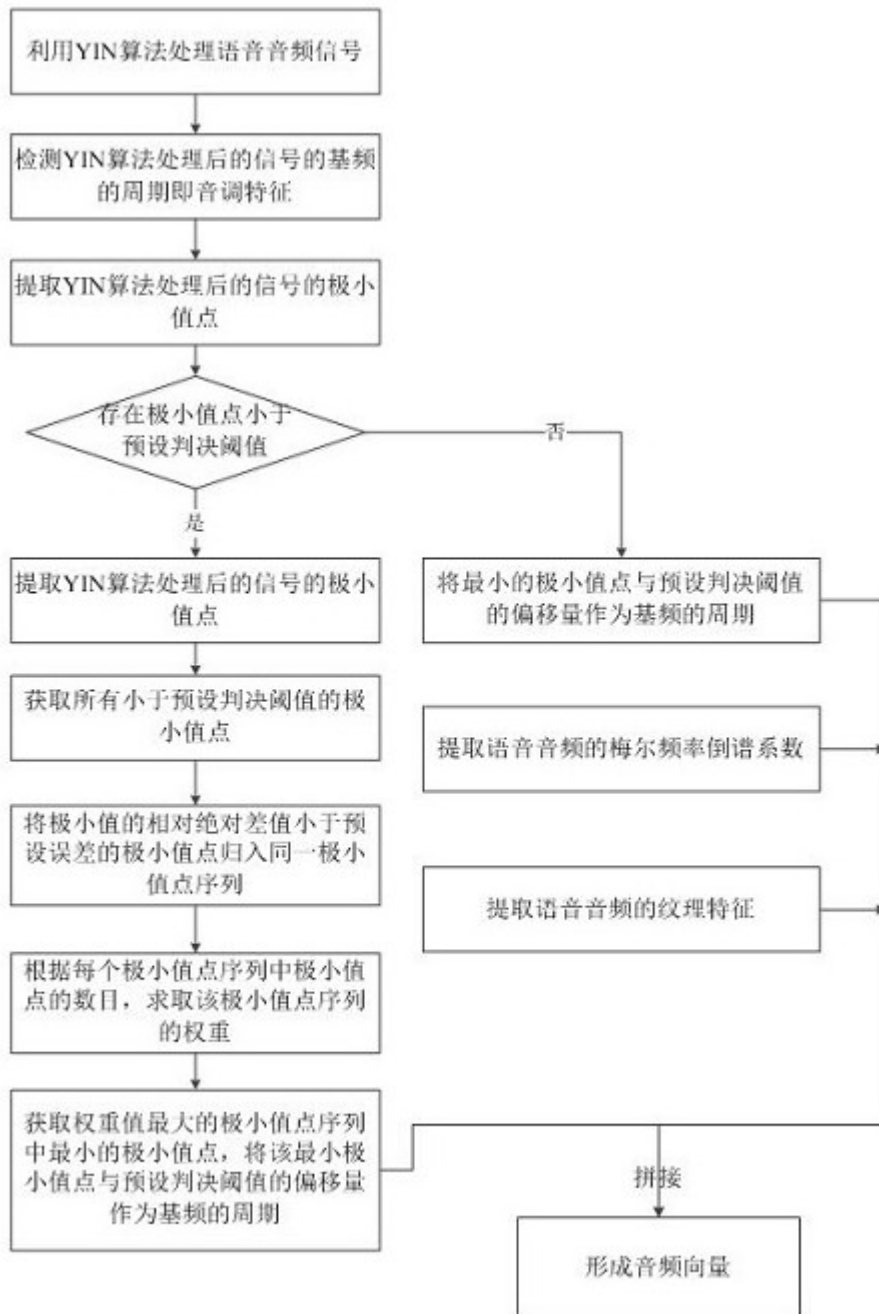


图10

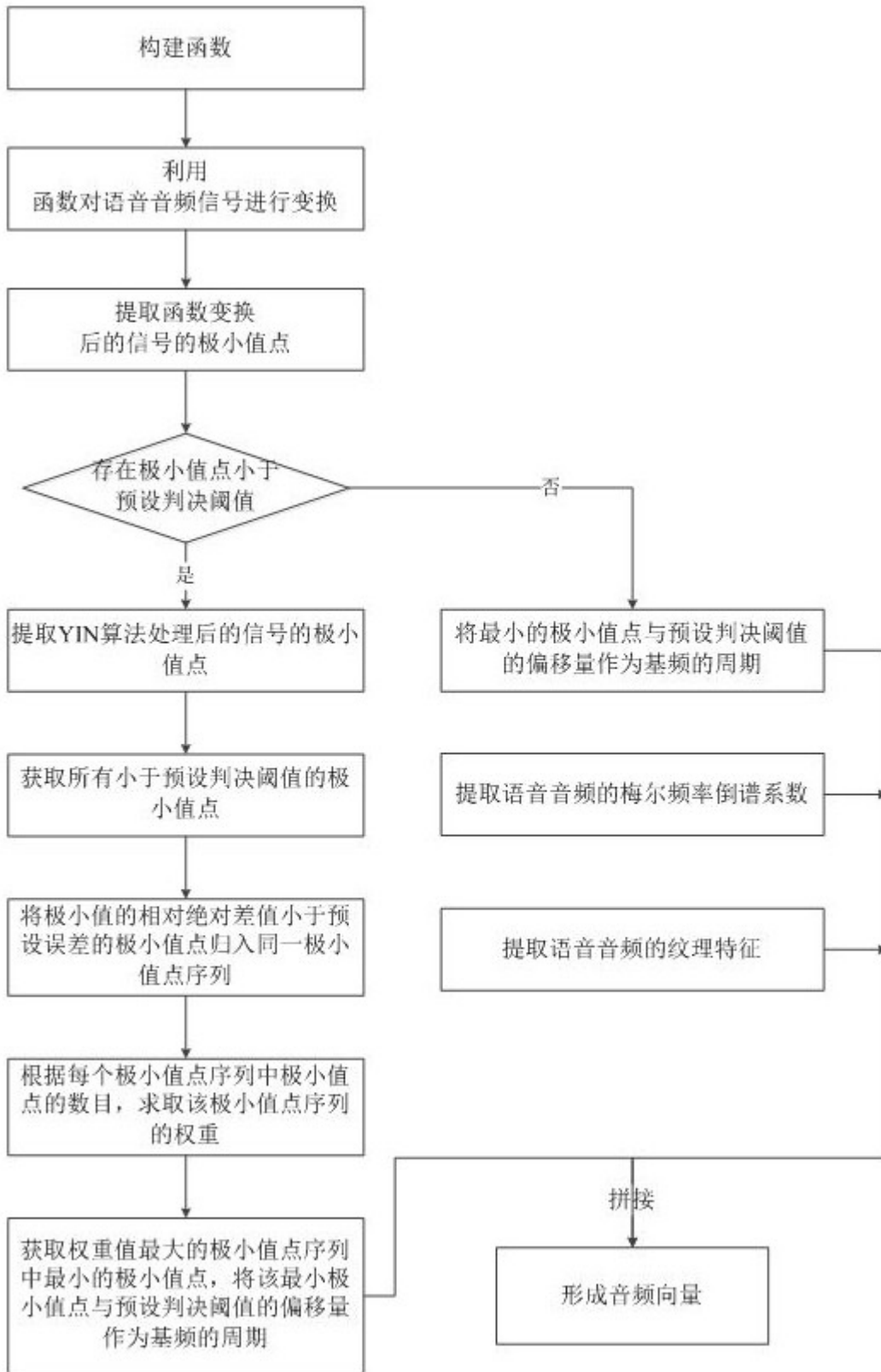


图11