

面向老年人的智能家居多模态交互系统研究

谷学静^{1,2} 王志良¹ 贺 杰¹ 郑思仪¹ 王 巍¹

(北京科技大学信息工程学院 北京 100083)¹ (河北联合大学电气工程学院 唐山 063009)²

摘 要 提出了应用于智能家居老年人帮助的新型多模态人机交互模式。构建了基于 avatar 的智能家居人机交互原型系统,集成了语音处理和视线追踪功能,实现了视觉和听觉双通道交互;同时采用基于规则的任务推理方法感知用户任务信息。测试结果表明,该交互模式提高了老年人的交互体验。

关键词 多模态,视线追踪,人机交互,智能家居,老年人

中图法分类号 TP183 **文献标识码** A

Study on Multimodal Interaction in Smart Home to Support Senior User

GU Xue-jing^{1,2} WANG Zhi-liang¹ HE Jie¹ ZHENG Si-yi¹ WANG Wei¹

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(School of Electrical Engineering, Hebei United University, Tangshan 063009, China)²

Abstract A new multimodal Human Computer Interaction(HCI) model was proposed to support senior user in smart home. The prototype system which integrates the functions of chinese language speech information processing and eye gaze tracking in the network environment of smart home was developed. By using the avatar technologies, achieved capability of visual auditory dual-channel interaction. User's intention was perceived by rule-based task reasoning. The experimental results show that user experience is improved.

Keywords Multimodal, Eye gaze tracking, HCI, Smart home, Senior people

1 引言

20 世纪 70 年代后,逐渐兴起智能家居的研究,智能家居可以提供给居住者舒适、安全、便捷的生活环境,拥有巨大的市场潜力。智能家居中包括大量的、可操纵的、协同工作的传感器和执行器,目前绝大多数研究都聚焦于智能家居的使能技术,如基础构架、软件、硬件等,关于人与智能家居和谐交互的研究还处于初级阶段。研究表明,提高智能家居的价值,关键不仅在于技术水平的提高,而且在于从实用、易用和人性化的角度做出改进^[1,2],具有和谐交互功能的智能家居服务有很大发展空间^[3]。

随着社会的发展,越来越多的国家开始步入老年社会,空巢老人的比例也越来越大,预计到 2030 年我国 65 岁以上的老年人口将达到 15.98%,届时中国将进入老龄社会,而 2040 年将成为超高龄社会^[4]。因此,如何提供一个可用性良好的界面来满足老年人用户的需求,显得相当关键。智能家居是老年人生活辅助的重要发展方向。许多经济发达国家已有大量的科研和工业界人士先后投入面向老年人的智能家居研究,具有代表性的工作包括乔治亚理工的 Aware Home^[5]、英特尔西雅图研究院的 Ageing-in-Place^[6]、英国布鲁内尔大学

的 Millennium Home^[7] 以及美国佛罗里达大学的 Houseof-Matilda^[8] 等。上述智能家庭主要面向老年人用户,通过嵌入在智能家庭中的智能感知、通信、交互设备来监测、引导老年人的日常生活,从而达到延长他们独立生活时间的目标,并为其提供较好的安全保障。认知学研究表明,人的视线具有直接性、自然性和双向性等其它信息所无法具备的特点^[9]。而且,如果缺少非语言信号的交流,不但会减弱人机交互中的社会表现力,导致情感交流的缺失,还会影响到人机交互的参与对等性。社会心理学的研究表明,注视行为在建立良好的关系和和谐的交互中发挥着非常重要的作用^[10]。因此,在人-机交互方式朝着更加自然、和谐的方向发展的今天,比起传统的手动输入方式及新发展起来的语音或手势输入方式来说,“视线方向”无疑是一种更好的能使人机对话变得简便、自然的输入通道^[11]。

本课题组已经在人机和谐交互方面做了一部分研究工作^[12,13]。本文的主要研究工作针对智能家居环境,面向老年人用户,从人机交互设计的角度,力求使交互的双方以及环境之间具有和谐一致的协作关系;在图形用户界面 WIMP(Window, Icon, Menu, Pointer)的基础上,基于 Microsoft Speech SDK 的 Application-Level Interfaces 提供的 API、识别语音及语音合成,基于 ICTCLAS(Institute of Computing

到稿日期:2010-12-12 返修日期:2011-03-25 本文受国家自然科学基金(60903067),863 国家高技术研究计划(2007AA04Z218),北京市自然科学基金(KZ200810028016)资助。

谷学静(1972-),女,博士生,副教授,主要研究方向为人工心理、情感交互、虚拟人技术,E-mail:guxuejing@sina.com;王志良(1956-),男,教授,博士生导师,主要研究方向为人工心理、机器人技术、物联网;贺杰(1985-),女,硕士生,主要研究方向为情感识别、智能家居;郑思仪(1981-),女,博士生,主要研究方向为图像识别、虚拟人技术;王巍(1983-),男,博士生,主要研究方向为机器人技术、情感建模。

Technology, Chinese Lexical Analysis System)和句模匹配技术,理解用户交互语言;基于 PCCR(瞳孔-角膜反射)技术,提取用户视线在屏幕上的落点坐标,从而构建了基于 avatar 的集语音和视线交互为一体的多模态交互模式,实现了适合老年人的智能家居系统自然、和谐人机交互终端。

2 基于 avatar 的语音视线交互系统设计

为了避免奇异古怪现象的影响和充分体现注视交互的效果,特别制作了大眼睛的卡通动画形象,作为用于智能家居老年人帮助系统的 avatar 开发模板。avatar 模块由 5 个模块实现:avatar 形象模块、视线追踪模块、Socket 模块、语音识别与合成模块和任务推理规则模块。其中视线追踪模块提取用户眼睛注视区域;Socket 模块用于 avatar 形象模块与视线追踪模块间的相互通信;语音识别与合成模块完成语音交互。由于互相独立地利用多个通道并不是真正意义上的多通道界面,不能有效地提高人机交互的效率,因此本设计利用任务推理规则模块协调从语音、视线两个并行、协作和互补的通道的非精确输入获得的任务信息。

2.1 avatar 语音交互功能的设计

语音识别功能利用微软的 Speech SDK 5.1 提供的 Application-Level Interfaces API 函数进行二次开发。在程序结构中加入 Microsoft Speech SDK 5.1 提供的识别引擎,使 avatar 可识别中文语音。微软的 Speech SDK 5.1 采用 COM 组件的形式,具体实现步骤如下:1)程序初始化 COM 环境;2)初始化语音识别引擎;3)加载语法文件;4)设置识别引擎返回消息。当语音识别事件发生之后,识别引擎自动向程序窗口发送识别成功或识别失败消息,开发者根据自身程序的情况进行相应的处理。关于语法文件的加载,采用基于数据库的动态语法规则加载方式。这种方法主要是将语法结构与语句词组互相分离,语法结构可以看作是识别语句的骨架,仍然保存在 XML 文件中,各部分的内容分别保存在数据库的不同表中。这种方法的优点在于利用数据库可以存储大量的单词或语句,数据管理简单,在不改变程序的情况下可以添加或删除识别语法的语句。

同样采用 Microsoft Speech SDK 5.1 实现语音合成,实现步骤为:1)调用 API 函数 CoInitialize 初始化 COM 组件;2)使用 SpFindBestToken 函数,传入参数,设置语言类型;3)调用 CoCreateInstance API 函数创建 COM 语音合成接口实例 IspVoice;4)调用 IspVoice 接口方法 SetVoice,加载前面所设置的声音类型;5)初始化工作全部成功后,则可以在程序需要的地方调用 IspVoice 接口中的 Speak 方法,将合成的语句以宽字符的形式作为参数即可。

2.2 avatar 视线交互功能的设计

视线追踪系统是一套利用人眼来进行人机交互的装置。设计中集成了本课题组自主开发的基于瞳孔-角膜反射(PCCR)技术视线追踪算法^[14],集成该功能的程序流程如图 1 所示。

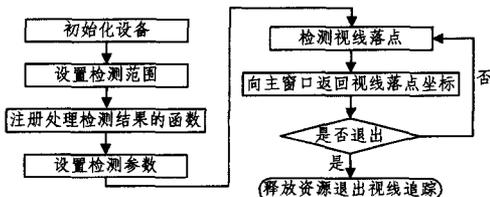


图 1 集成视线追踪功能程序设计流程

(1)初始化图像采集卡和摄像头等硬件设备。调用视线追踪模块的 gzInit 接口进行初始化。

(2)设置检测范围。默认检测范围为全屏幕(也可以指定检测范围为屏幕上某一区域)。调用视线追踪模块的 gzSetScreenRect 接口进行设置。

(3)注册检测结果处理函数。本模块的结果处理函数是 SendXY(),用来向主窗口发送视线落点坐标。通过为两个函数指针 OnRegion 和 OnPos 传递值来实现注册。

(4)设置检测参数。设置用户检测参数有利于提高检测精确度。用户参数存储在指定的文本文件中,通过调用视线追踪模块的 gzSetParameters 接口读取用户参数进行设置。

(5)检测视线落点。通过调用视线追踪模块的 gzStart 接口开始进行视线追踪。

(6)向主窗口发送视线落点坐标。调用(3)中注册的结果处理函数向主窗口发送视线落点坐标,坐标全局变量为(Mousex, Mousey)。

(7)判断是否退出视线追踪检测。如果不退出,则回到(5);如果退出,则转到(8)。

(8)释放资源退出视线追踪。释放占用的内存空间和和设备资源并退出。

屏幕坐标分区如图 2 所示,avatar 利用视线追踪程序得到眼睛定位坐标(Mousex, Mousey)。

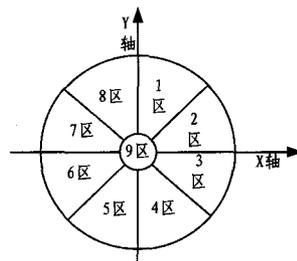


图 2 屏幕坐标分区示意图

按照表 1 的对应关系,avatar 响应用户视线方向,利用 socket 模块向虚拟人发送位置消息,从而实现虚拟管家眼动视线跟随,让交互更加生动。

表 1 用户视线落点与 avatar 注视方向对应关系

分区	avatar 注视方向角度值
1 区	67.5°
2 区	22.5°
3 区	-22.5°
4 区	-67.5°
5 区	-112.5°
6 区	-157.5°
7 区	157.5°
8 区	112.5°
9 区	正视图用户
X 轴正半轴(9 区外部分)	0°
X 轴负半轴(9 区外部分)	180°
Y 轴正半轴(9 区外部分)	90°
Y 轴负半轴(9 区外部分)	-90°



图 3 利用视线追踪技术代替鼠标单击按钮操作截图

在实现 avatar 视线跟随功能的同时,系统计算视线在具有该坐标值的功能按钮上停留的时间来响应按钮消息,即代替鼠标作为一种指点装置。图 3 所示为用户注视天气预报按钮的程序响应。

2.3 任务推理规则

利用语音和视线两个通道的信息获取来识别用户的任务信息,比利用单个通道的信息会更有效地提高系统对用户命令识别的准确性。采用实验的方法,记录用户语音视线操作习惯,选择 10 名 65~70 岁之间的老年人作为测试对象,其中男性 6 名、女性 4 名。每人完成指定的 10 个操作任务,观测并记录其操作习惯,结果表明,在 10 个人 100 次任务中,89 次任务操作时用户的语音指令和视线注视的方向是一致的,仅出现了 11 次二者不统一的情况。考虑到语音和视线的识别问题以及可能出现的两个通道的任务信息不一致的情况,采用了基于规则的任务推理方式来提高系统的可靠性,如表 2 所列。

表 2 任务推理规则

V	E	V==E?	T
V=T1	E=T2	Yes	T=T1
V=T1	E=A	No	T=T1
V=0	E=T1	No	T=T1
V=T1	E=0	No	T=T1
V=T1	E=T2	No	T=VR
V=D	E任意	No	T=D

其中 V 表示语音, E 表示视线, T 表示任务, VR 表示开始语音识别,即在 V、E 均指向某个任务而二者又不一致的情况下,系统通过语音询问用户任务信息,以用户的语音回答作为判别结果。D 表示对话,此情况下开始和 avatar 的聊天交互。在这一部分使用了基于文本的句模匹配的方法,主要实现步骤为:自动分词、词性标注、问题类型归类、关键词提取、基于规则的句模匹配。通过以上基于关键词定位、句模规则匹配的处理过程,已经确定问句的内容与类型,通过数据库查找技术,得到适合的答案,使 avatar 具有基于智能家居管理内容的受限范围的自然语言理解能力。

3 实验结果及分析

本研究将具有多模态交互能力的 avatar 作为智能家居人机交互界面。为了验证这种人机交互方法对老年用户的有效性及其满意度,选择问卷的方式进行情感测量。虽然已有心理学研究表明^[15],在使用问卷调查的方法时,有时会出现报告与行为不一致的情况,但是由于我们所使用的量表为语义差异量表,每对词所表示的情感在其所属维度上的量值相反,因此避免了不一致情况的发生。基于本系统的目标用户为中国人,所以采用适合中国人情感评价的中文简化版情感量表,并结合 PAD(愉悦度 Pleasure-displeasure、激活度 Arousal-non-arousal、优势度 Dominance-submissiveness)三维情感空间,从 3 个维度上分析本系统对老年用户情感体验的影响。仍然选择 10 名 65~70 岁之间的老年人被试,其中男性 6 名、女性 4 名,分别使用第一组传统的触摸屏界面(去掉本研究中的 avatar 和语音视线交互功能的界面)和第二组本研究开发的基于语音视线交互功能的 avatar 界面,完成相同的 10 项家居管理任务,并在完成每一组界面的所有 10 项操作任务后,记录其在 PAD 3 个维度上的体验数据。

$$x_i = (P_i \quad A_i \quad D_i), 1 \leq i \leq N, N=10 \quad (1)$$

实验结果表明,针对第二组界面的 3 个维度的平均值分别为

$$\bar{x}_2 = (\bar{P}_2 \quad \bar{A}_2 \quad \bar{D}_2) = (21.39 \quad 3.92 \quad 11.8) \quad (2)$$

第一组界面的 3 个维度的平均值分别为

$$\bar{x}_1 = (\bar{P}_1 \quad \bar{A}_1 \quad \bar{D}_1) = (10.24 \quad 6.92 \quad -1.02) \quad (3)$$

由实验结果可知,在 P 维度上第二组平均值大于第一组的平均值,说明在操作的愉快性方面第二组有显著优势。在 A 维度上第二组低于第一组的原因是由于操作过程是在相对放松的交互过程中进行的,因此唤醒度相对较低。在 D 维度上第二组平均值又显著地高于第一组,是由于在任务执行的过程中基于 avatar 的交互是以用户为中心的,用户不会感觉被程序控制,因此会有较高的优势度。由此,验证了系统的对于帮助老年人完成智能家居的操作的有效性,使老年人用户可以在较放松的状态下完成操作,并在操作的过程中获得较高的愉悦感和成就感。

结束语 数字化时代智能系统的重要特性之一就是要有良好的交互性(Interactivity)。对于旨在提高人类家庭生活质量的智能家居服务系统,良好的交互显得更加重要。本文完成了应用于智能家居的基于 avatar HCI(Human Computer Interaction)系统的开发,提出并实现了语音和视线双通道交互,实现了无需手动参与的交互方式,弥补了智能家居领域目前对以用户为中心的自然人机交互研究的不足。利用中文简化版情感量表,对系统进行实验测试,得到了基于 PAD 情感空间的情感体验描述,实验结果表明具有视线和语音交互功能的 avatar 智能家居服务系统可以提升用户在交互中的正向情绪。

参 考 文 献

- [1] Yeoh C-M, Tan He-Y, Kok C-K, et al. e2Home: A Lightweight Smart Home Management System[C]// International Conference on Convergence Information Technology. 2008, 250: 80-86
- [2] Teng W-C, Pao Y-C, Chung S-L. Design of My Server: A Residential Server in Smart Home Systems[C]// IEEE Asia-Pacific Conference on Services Computing. 2008, 191: 82-87
- [3] Noh M J, Kim J S. Factors influencing the user acceptance of digital home services[J]. Telecommunications Policy, 2010, 34: 672-682
- [4] 周兴社, 於志文. 面向老年人生活的智能辅助[J]. 中国计算机学会通讯, 2010, 6(6): 57-67
- [5] Kidd C D, Orr R, Abowd G D, et al. The Aware Home: A Living Laboratory for Ubiquitous Computing Research[C]// Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture. 1999, 1670: 191-198
- [6] Adam D, Roberts. Exploring Design Dialogues for Ageing in Place[J]. Anthropology in Action, 2009, 16(1): 72-88
- [7] Dowdall A, Perry M. The Millennium Home: Domestic Technology to Support Independent-living Older People[C]// The 1st Equator IRC Workshop on Ubiquitous Computing in Domestic Environments, Nottingham, 2001: 1-15
- [8] Helal S, Winkler B, Lee C H. Enabling Location-aware Pervasive Computing Applications for the Elderly[C]// Proc. of 1st IEEE

- [9] Jacob R J K, Karn K S. Eye tracking in human computer interaction and usability research; Ready to deliver the promises [J]. Computer Vision and Image Understanding, 2003, 24: 682-701
- [10] Bavelas J B, Coates L, Johnson T. Listener responses as a collaborative process; The role of gaze[J]. Journal of Communication, 2002, 52(3); 566-580
- [11] 冯成志, 沈模卫. 视线跟踪技术及其在人机交互中的应用[J]. 浙江大学学报, 2002, 29(2); 25-232
- [12] 祝长生, 王志良. 基于注意度评价的人工意识模型[J]. 计算机科

- [13] 孟秀艳, 王志良, 李娜, 等. 情感机器人的情感模型研究[J]. 计算机科学, 2008, 35(6); 158-162
- [14] Zhang Peng-yi, Wang Zhi-liang, Zheng Si-yi, et al. A Design and Research of Eye Gaze Tracking System Based on Stereovision [C]//5th International Conference on Intelligent Computing, 2009, 5754; 278-286
- [15] Mohammad Y, Xu Y, Matsumura K, et al. The H3R explanation corpus; human-human and base human-robot interaction dataset [C]//The Fourth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP2008). 2008; 1-47

(上接第 199 页)

是时间比例一样。实验计算机条件为 CPU DELL 3GHz, 主板 915, 内存 512MB。通过实验可知, 基于偏移率的方法与现行的模糊推理方法相比, 每次计算量少一些, 但总的学习执行时间缩短到约一半。基于偏移率的方法比现行方法在降水量预测的场合快 5.2 倍, 安全态势预测的场合快 1.37 倍。因为安全态势预测的场合, 模糊变量取 3 个, 时间差比例比较小。随着模糊规则的增加, 两个方法的时间差越大。降水量预测的学习误差得到改善(7.35%), 而且安全态势预测误差稍有下降(0.937%)。提出的方法的准确度比文献[18]的 PSO_FNN 的准确度 83.34% 有所改善。安全态势预测准确度下降的原因是模糊变量的个数只是 3 个, 但是随着模糊变量的增加, 准确度也随之提高。学习执行时间缩短的原因是, Sugeno 方法在模糊推理过程中要计算较复杂的指数函数形式的隶属函数, 但本文提出的方法只计算用隶属函数的宽度与输入信息表示的距离, 且通过阈值的计算去掉不必要的模糊推理过程, 所以计算复杂度相对比较小。通过测试, 最终证明, 本文提出的基于偏移率的模糊推理方法与学习算法比现行的模糊推理方法与学习算法具有更高的效率。

结束语 本文针对模糊神经网络学习算法计算量过大、误差比较大的问题, 提出了一种基于偏移率的模糊推理方法与模糊神经网络学习算法, 并与现行的模糊推理方法的理论比较, 得出新方法的优点。通过降水量与安全态势预测模型设计实验, 证明了新的推理方法与学习算法的有效性。仿真结果表明, 预测模型设计中用基于偏移率的方法比现行的方法学习时间短, 而且学习误差比较小, 本文提出的方法为模糊推理方法与模糊神经网络学习算法的研究提供了一种新的途径。

参 考 文 献

- [1] Buckley J, Hayashi J Y. Fuzzy neural networks; a survey [J]. Fuzzy Sets and Systems, 1994, 66(1); 1-13
- [2] 王士同. 神经模糊系统及其应用 [M]. 北京: 北京航空航天大学出版社, 1998; 126-209
- [3] Jang J R. Self learning fuzzy logic controllers based on temporal back-propagation [J]. IEEE Trans. Neural Network, 1992, 3(5); 714-723
- [4] Wang Li, Mendel J M. Generating fuzzy rules by learning from example [C]//Intelligent Control, Proceedings of the 1991 IEEE

International Symposium on. 1991; 263-268

- [5] 金龙, 史旭明. 模糊输入对模糊输入网络预报模型的影响研究 [J]. 计算机工程与应用, 2007, 43(31); 192-195
- [6] 萧海东. 网络安全态势评估与趋势感知的分析研究 [D]. 上海: 上海交通大学, 2007
- [7] Lee K C, Cho H R, Kim J S. A self-organizing feature map-driven approach to fuzzy approximate reasoning [J]. Expert Systems with Applications, 2007, 33(2); 509-521
- [8] Fukuda S, Hiramatsu K. Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (*Oryzias latipes*) [J]. Ecological Modelling, 2008, 215(4); 301-313
- [9] Lin Da, Wang Xing-yuan, Nian Fu-zhong, et al. Dynamic fuzzy neural networks modeling and adaptive backstepping tracking control of uncertain chaotic systems [J]. Neurocomputing, 2010, 73(16-18); 2873-2881
- [10] 白玉清, 孙云山. 模糊神经网络的发展 [J]. 科技情报开发与经济, 2007, 17(21); 157-158
- [11] 张健. 模糊神经网络模型算法研究与应用 [D]. 大庆: 大庆石油学院, 2002
- [12] 许少华, 王颖, 王皓. 一种基于数值积分的过程神经网络训练算法 [J]. 计算机科学, 2010, 37(11); 203-205
- [13] 李少远, 王景成. 智能控制 (第二版) [M]. 北京: 机械工业出版社, 2009; 25-51
- [14] Yu Xia, Wang Shuo-yu, Zhao Xian-chao. A health-check system for health-care robot Robotics and Biomimetics (ROBIO) [C]//2005 IEEE International Conference on. 2005; 709-712
- [15] Jiang Yin-lai, Wang Shuo-yu. Adapting directional intention identification in running control of a walker to individual difference with fuzzy learning [C]//Mechatronics and Automation (ICMA). 2010 International Conference on. 2010; 693-698
- [16] Ho C, Li Jing-jiao, Gwak S. Research of a New Fuzzy Reasoning Method By Moving of Fuzzy Membership Functions [C]//IPTC. IEEE computer society, 2010; 297-300
- [17] Lohani A K, Goel N K, Bhatia K K S, Takagi-Sugeno fuzzy inference system for modeling stage-discharge relationship [J]. Journal of Hydrology, 2006, 331(1/2); 146-160
- [18] 周岩, 王盛, 高传善, 等. 基于改进粒子群算法的模糊神经网络及其在短时天气预报中的应用 [J]. 计算机应用与软件, 2010, 27(5); 234-237