

基于语言认知的智能电视 VUI 适老化设计研究

郭会娟¹, 汪海波², 殷堃²

(1.东南大学, 南京 211189; 2.安徽工业大学, 安徽 马鞍山 243002)

摘要: **目的** 基于语言认知老化机制, 研究智能电视语音交互界面的适老化设计策略。**方法** 对语言产生和语言理解过程中语言认知老化的现象和基础机制进行研究, 总结出语音交互过程中可能影响老年人交互效率和交互体验的主要因素; 通过 KANO 模型对影响老年用户使用智能电视满意度的设计因素分类, 并识别用户需求的重要性, 以此为依据提出智能电视语音交互界面的适老化设计策略。**结论** 从用户语言认知的角度出发, 适度采取简洁自然的语言交互环境、灵活包容的信息识别策略、精准有效的语音输出迭代等补偿性策略, 可有效减少老年用户与智能电视语音交互的障碍, 提升绩效并优化使用体验。

关键词: 语言认知; 智能电视; 语音交互界面 (VUI); 适老化设计

中图分类号: TB472; J511 **文献标识码:** A **文章编号:** 1001-3563(2022)08-0050-05

DOI: 10.19554/j.cnki.1001-3563.2022.08.007

Adaptive Aging Design of Smart TV VUI Based on Language Cognition

GUO Hui-juan¹, WANG Hai-bo², YIN Shuang²

(1.Southeast University, Nanjing 211189, China; 2.Anhui University of Technology, Anhui Maanshan 243002, China)

ABSTRACT: This paper aims to study the adaptive aging design strategy of smart TV voice interaction interface based on language cognitive aging mechanism. The phenomenon and underlying mechanism of language cognitive aging in the process of language generation and language understanding are studied, and the main factors that may affect the interaction efficiency and interaction experience of the elderly in the process of human-computer voice interaction are summarized. The KANO model is used to classify the design factors affecting the satisfaction of elderly users in the process of using a smart TV, and the importance of user needs is recognized. Based on this, an aging-friendly design strategy is proposed for the voice interaction interface of smart TV. From the perspective of user language cognition, moderate compensatory design strategies such as a simple and natural language interaction environment, flexible and inclusive information recognition strategy and accurate and effective voice output iteration can effectively reduce the obstacles to voice interaction between elderly users and smart TV, improve the interaction performance and optimize the user experience.

KEY WORDS: language cognitive; smart TV; voice user interface; adaptive aging design

人工智能时代, 技术革新不断推动机器智能的进步, 极大地拓展了人机交互的空间。近年来, 在多模态交互中, 语音交互技术的突破为人机之间自然交互的达成贡献了巨大力量, 逐渐成为市场主流和研究热点。现代家庭场景中, 智能电视因兼具智慧数据连接和优质观赏体验, 逐渐成为家庭娱乐的核心, 特别深受老年用户的青睐。相对于图形交互界面 (GUI, Graphical User Interface) 密集的选项、复杂的操作,

语音交互界面 (VUI, Voice User Interface) 的信息线性、直觉识别性、操作无边界性和反馈的自然性更能迎合智能产品使用经验不足的老年用户。

语音交互是以语言为媒介、基于某种目的、围绕特定任务展开的人机对话, 是人机之间意义和认知协调的过程, 见图 1。智能产品对用户语音输入的处理起始于语音识别, 其使用声源定位、降噪、回音消除和端点检查等技术清洗出目标信息, 并用声学 and 语言

收稿日期: 2021-12-24

基金项目: 非遗数字文创的文化认知与用户体验研究 (AHSKY2020D99)。

作者简介: 郭会娟 (1980—), 女, 博士生, 副教授, 主要研究方向为交互界面设计。

学模型将语音数据转译为文本;通过文本分析将用户的指令转化为机器的结构化数据,完成对用户语言的认知;接着从数据库中提取出响应数据,将数据文本合成语音输出。用户在感知相关的机器语言信息后,也会通过对感知材料的分析建构交互的内容和意义,完成对机器语言的认知;并依据认知的结果确定自身要表达的内容和概念,经词汇择取、结构组织等过程将交互意图表达出来。

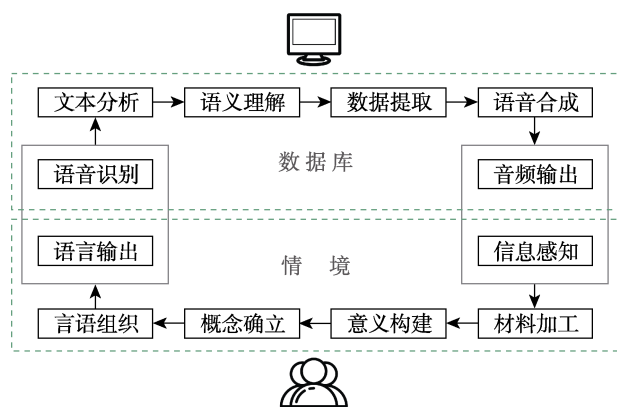


图 1 语音交互中的语言认知过程

Fig.1 Language cognitive process in voice interaction

从语言认知的角度来看,目前语音交互过程中语言信号的处理方式和响应机制基本可按照信息加工的模式来认识。交互中,人机双方需要在对话中注意和捕捉对方的语言信号及信息需求,进而推断其期望和意图。与普通语言交流不同,语音交互过程中并非时时遵守“听者设计”的原则,产品无论是作为“说者”还是“听者”,都需要以用户为中心建立交流基础,以促进用户对语言的认知,增强人机的合作性和共享性,降低语音交互的歧义和错误。与图形交互使用结构化数据驱动不同,语音交互的人机对话过程中使用的是自然语言,VUI 要给出正确的响应除受语音识别、语音合成等技术的影响之外,用户语言交流的能力也是其重要制约因素。特别是老年用户,因生理机能的老化必然产生相应语言认知能力的下降,从而在语音交互中产生操作和反馈障碍。因此,基于语言认知展开对智能电视 VUI 适老化设计研究十分必要。

1 老年用户的语言认知老化研究

伴随生理机能的老化,老年人的工作记忆不断衰退、认知资源逐渐减少,在语言认知方面的能力也呈现逐渐下降的态势^[1]。语言认知主要涉及 2 个部分:语言理解和语言产生。在传统的认知心理学研究中,多数认为语言理解和语言产生具有不同的神经基础,二者是可以相互分离的。但近些年的研究表明,尽管神经机制不同,但两类加工过程中都包括了概念、语义、音韵、字形等表征,语言理解和语言产生存在着相互影响,可以整合看待。^[2]在老年人与智能电视的

语音交互中,语言理解和语言产生也是不断交互循环的。因此,对语言认知的老化研究,要在认知机制的基础上同时关注语言理解和语言产生的全过程。

1.1 语言理解过程中的认知老化

语言理解是基于感官通道获取的语言刺激,通过大脑的注意和信息加工机制,积极主动建构意义的过程。首先是对语言刺激的感知,人机语音交互中,用户感官获取语言刺激的来源有 3 种:来自机器的言语感知、来自图形界面的阅读感知和情境感知。听觉语境下,言语感知中的认知老化主要由 2 个因素构成:一是语音知觉的恒常性老化^[3],老人对相似音的区分存在明显困难,尤其当面临多个对象,且其个体间存在发音差异时,老人的词汇识别能力受损呈现更加明显;二是口语环境下语音边界区分能力的老化^[4],在一定范围内,当词汇间呈现间隔的时间越短,老人的词汇识别效率越低。视觉语境下,相比年轻人,老人在阅读过程中对单个词汇的关注更多,且注视点呈多次反复趋势^[5],特别是由于抑制干扰信息能力衰退,老人在拼写规则复杂的阅读区域中,需要付出更多的认知资源。与以上 2 种不同,可能是基础经验的丰厚,老人通过符号、图片、语气、动作和其他知觉等情境因素预测词汇的能力较年轻人强,他们对语义的依赖程度更高。其次是对语言材料的分析,从信息加工的角度来说,这个过程包括语言结构的建构、干扰信息的抑制和语境语用知识的整合。由于工作记忆老化,老人在语言分析中对关联信息的激活水平和命题编码效率明显下降,导致复杂句式的加工时间相较于年轻人显著增加^[6],他们更习惯采用“点”加工的模式去分析句子中的关键词汇,以抑制干扰信息,消解歧义,眼动监测中的跳读现象印证了此观点,但明显加工策略更加冒险,常存在对低频词和低预测词的忽略,从而导致语言分析和意义建构的失败。

1.2 语言产生过程中的认知老化

在语言交流中,语言产生是对语言理解的使用,其整个过程包含概念化、组织化和语言输出 3 个阶段。首先是概念化阶段,依据理解的结果确立对话的意图和期望表达的概念,产生前词汇信息。相对于年轻人,老人在这方面的认知老化主要表现在词汇的提取方面,即著名的“舌尖效应”。由于脑生理机能的老化,老人的工作记忆容量衰减、信息加工能力受损,以致出现明显感觉存储存在相关的语义信息,却无法即时映射到词汇的现象^[7]。其次是言语组织阶段,即将词汇信息转换成具体的语言结构形式。由观察可见,老人在口语交流中更倾向简单的句式,而且在结构性方面的出错率也明显高于年轻人,这表明老人在复杂句式的加工方面已经表现出明显的能力不足。最后是语言输出阶段,即以书面、口语或手语的形式将交流的意图表达出来。这一过程相对复杂,程序性较

强,更多依赖于记忆、注意和执行控制,其老化突出表现在整体的连贯性、聚焦性和完整性方面,局部老化的效应不明显^[8]。这意味着老年人对长文本的把握能力明显下降,可能在较长的对话中容易出现偏离主题、前后表达不一致、语义混乱等现象。

从用户语言认知角度的分析可见,语音交互过程中可能影响老年人交互效率和交互体验的因素主要包括:单通道的语音感知、交互的语速、文本结构的复杂程度、对话的长度、语义的提示、容错性的处理、人机交互的情境等。下面将以此为调研依据,以老年用户使用智能电视过程中对 VUI 的设计需求展开研究。

2 智能电视 VUI 适老化设计需求分析

鉴于智能电视功能的综合性和复杂性,此次需求分析仅针对人机交互中涉及语言认知的部分进行。结合用户访谈和问卷调查,依据老年用户与智能电视语音交互的环节,设置听觉感知、视觉感知、语言分析与产生、工作记忆 4 个层面、10 个体验要素和 16 个具体需求点(见表 1),用 KANO 模型以设置正反问题的方式来获取用户对每项需求的态度,以确认体验要素的优先级和识别设计需求的重要性。

受访的 20 位用户样本取自某市老年大学,年龄在 60~70 岁,男性、女性各 10 位,教育背景良好,学历均在高中及以上,大专及以上学历 13 位;其中使用智能电视语音交互功能不满 3 个月的见习用户 5 位,使用语音交互功能 3 个月至 1 年的中级用户 10 位,使用语音交互功能经历 1 年以上的专家用户 5 位。为确保问卷结果的有效性,使用 SPSS24.0

表 1 老年用户智能电视语音交互需求汇总表
Tab.1 Summary of voice interaction requirements between elder users and smart TV

语言认知	体验要素	设计需求点	序号
听觉感知	声音情绪	活泼、亲切	A1
	交互语速	更慢	A2
视觉感知	视觉情境	形象拟人化	A3
		倾听表现	A4
		思考表现	A5
	辅助界面	字号加大	A6
		色彩清晰	A7
语言分析与产生	对话界面	对话信息标记	A8
	结果反馈界面	有指令对话记录	A9
		选项少量显示	A10
	语义示例提示	按历史推荐显示	A11
		唤醒时凸显示例	A12
工作记忆容量	语音错误处理	分析错误原因	A13
	输入等待时间	输入等待时间更长	A14
	多轮对话	多轮对话标识	A15
		结果收至提示库	A16

对问卷数据进行了信度分析,结果内部一致性良好,数据具备可信度。

2.1 体验要素的优先级确认

收集问卷中体验要素的重要程度相关数据,以赋值加权平均的方式进行优先级计算。按不重要、不太重要、一般、比较重要、非常重要分别赋值 1、2、3、4、5 分,具体计算结果见表 2。

表 2 老年用户智能电视语音交互体验要素优先级排序
Tab.2 Prioritization of experience elements of voice interaction between elder users and smart TV

体验要素/ 重要性	占比/%					重要 性/%	优先 级
	1 分/	2 分/	3 分/	4 分/	5 分/		
声音情绪	15	25	45	10	5	6.5	低
语速	20	30	40	10	0	6.0	低
视觉情境	5	10	50	20	15	9.0	中
辅助界面	10	20	35	20	15	10.3	中
对话界面	5	10	35	35	20	12.0	高
结果反馈界面	15	15	40	15	15	9.0	中
语音示例提示	5	10	30	35	20	13.3	高
语音错误处理	10	5	15	40	30	16.5	高
输入等待时间	15	20	35	15	15	9.5	中
多轮对话	10	15	25	20	30	13.5	高

由计算结果可见:对话界面、语音提示示例、语音错误处理、多轮对话的需求优先级别较高,目前老年用户在词汇识别、词汇提取、文本长度和语言输出部分的交互障碍比较明显;对视觉情境、辅助界面、结果反馈界面、输入等待时间的需求级别中等,可见老年用户在视觉与听觉的融合感知、工作记忆容量方面也存在部分障碍;对声音情绪和语速的需求级别较低,说明目前智能电视的 VUI 界面在交互语速和声音情绪方面交互障碍不明显。

2.2 设计需求的重要性识别

统计并分析设计需求部分数据,对照 KANO 模型将其按兴奋型需求(A)、期望型需求(O)、必备型需求(M)、无差异需求(I)和反向需求(R)进行归类分析,得出老年用户智能电视语音交互需求属性分类,见表 3。

表 3 智能电视 VUI 交互设计需求 KANO 属性分类表
Tab.3 KANO attribute classification of smart TV VUI design requirements

序号	A	O	M	I	R	序号	A	O	M	I	R
A1	4	8	5	3	0	A9	4	8	7	1	0
A2	2	7	5	6	0	A10	5	8	6	1	0
A3	4	10	5	1	0	A11	7	5	4	4	0
A4	5	7	6	1	1	A12	5	8	5	2	0
A5	3	9	7	1	0	A13	8	5	6	1	0
A6	2	7	3	6	2	A14	9	4	4	3	0
A7	3	5	7	4	1	A15	5	7	4	3	1
A8	4	5	7	3	1	A16	8	5	4	3	0

根据用户满意度指数 (Customer Satisfaction Index, CSI) 表示智能电视的某种功能对用户满意度影响可分为用户满意度增加指数 (Satisfaction Increment Index, SII) 和用户不满意度降低指数 (DissatisfactionDecrement Index, DDI) 其中, SII (S) 的计算公式为: $S=(A+O)/(A+O+M+I)$, DII(D) 的计算公式为: $D=-(M+O)/(A+O+M+I)$ 。

对老年用户智能电视语音交互需求属性分类结果进行计算, 结果见表 4。

为对设计需求的重要性进行排序, 并将其对老年用户体验的影响程度可视化, 根据表 4 的计算结果, 构建以满意度增加指数绝对值|S|和不满意度降低指数绝对值|D|为纵、横坐标的四象限图进行数据分析。其中象限临界点为满意度指数均值, 纵轴临界线为 0.591, 横轴临界线为 0.615, 见图 2。

其中第一象限为期望型需求, 包含 A1、A3、A4、A5、A9、A10、A12 等 7 个需求点, 说明声音情绪的活泼可亲、视觉情境的营造、指令对话的记录、内容选项的少量显示和语义提示示例的凸显, 对老年用户比较重要, 这些需求的满足程度越高, 老年用户的交互满意度就越好, 是产品提升用户体验的主要亮点。位于第二象限的是兴奋型需求, 包括 A11、A13、A14、A15、A16 等 5 个需求点, 说明语义示例的历史保存和推荐、对语音错误处理的分析和反馈、智能的输入等待、多轮对话的标识等功能是老年用户对智能电视

VUI 界面的潜在需求, 虽然不会因缺乏而表现出明显的不满, 但一旦被关注, 即使需求未得到完全满足, 用户的满意度也会大幅上升。第三象限内的是无差异型需求, 包括 A2 和 A6, 即目前智能电视语音交互的语速和辅助界面的字体字号比较适中, 老年用户对其语速的减慢和字号的加大无明显需求, 设计过程中可以延续。第四象限是必备型需求, 包括 A7 和 A8, 说明辅助界面的清晰程度、对话界面中重要的信息标记是老年用户对智能电视的基本要求, 这明显与用户生理视觉机能的老化和语言阅读相关, 在设计中需要特别重视, 以防影响用户的基本使用, 导致用户对产品的满意度降低。

依据 KANO 模型的重要性排序规则: 必备型需求>期望型需求>兴奋型需求, 得出结论如下。

在智能电视的适老性需求中, 重要的是色彩清晰、对话信息标记等需求, 涉及交互信息识别和语言产生的促进, 是智能电视必须要提供的, 否则会大幅降低用户的满意度; 声音的情绪、形象的拟人、对话记录、凸显示例等涉及感知情境营造、语义认知提示方面的需求重要程度次之, 是竞品之间比较的重点; 历史记忆、致错分析、提示库、多轮对话标识等涉及老人工作记忆容量的需求再次之, 对用户满意度和忠诚度的提升非常有效。

3 智能电视 VUI 适老化设计策略提出

通过对老年用户语言认知老化机制和智能电视适老化设计需求重要性的映射和比照, 可以从影响信息感知和加工的交互情境、影响文本分析和语义加工的信息识别、补偿工作记忆容量的输出迭代等方面对智能电视的 VUI 设计提出相应的设计策略, 以增强语音交互中人机交流目标的一致性和交流过程的合作性, 降低老年用户的操作和反馈障碍, 提升语音交互的质量和满意度。

3.1 营造简洁自然的语言交互情境

在智能电视的使用过程中, 交互情境对信息的感知、意图的预测、语义的加工以及交互过程中的情绪和交互的绩效都有重要影响。首先, 对于认知能力衰退的老年用户而言, 交互情境中的信噪比是影响其信息感知的重要指标, 因此, 要尽可能保持 VUI 中信息呈现的简洁性。如使用简短、温和、舒适的提示音, 视觉辅助界面保持清晰、简洁的风格等, 都可避免因信息过载产生过多的注意占用, 提高信息感知的有效性。其次, 由于老年用户口语交流中有更高语义依赖的倾向, 智能电视 VUI 要尽可能营造自然的交互方式和交互情境。语音自然交互的基础之一, 是拟人化的声音合成, 在音色、节奏、语调、语速等方面遵循人类的语言学特征, 加入情感化设计的要素, 如使用纯真的音色、舒缓的节奏、尾音上扬的语调、积极鼓励

表 4 智能电视 VUI 设计需求满意度指数计算汇总
Tab.4 Design requirements CSI summary of smart TV VUI

序号	分类	S	D	序号	分类	S	D
A1	M	0.600	-0.650	A9	O	0.600	-0.750
A2	O	0.450	-0.600	A10	O	0.650	-0.700
A3	M	0.700	-0.750	A11	O	0.600	-0.450
A4	O	0.632	-0.684	A12	A	0.650	-0.650
A5	O	0.600	-0.800	A13	A	0.650	-0.550
A6	O	0.500	-0.556	A14	O	0.650	-0.400
A7	O	0.421	-0.632	A15	A	0.632	-0.579
A8	M	0.474	-0.632	A16	A	0.650	-0.450

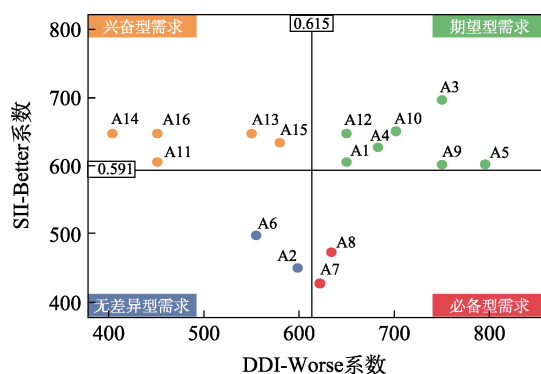


图 2 智能电视 VUI 设计需求四象限
Fig.2 Design requirements quadrant chart of smart TV VUI

的语气都更能迎合老年用户的心理预期,减轻其认知过程中的负荷。同时自然交互还包括必要的场景模拟,如生动可爱的视觉形象与适度配合的表情、动作都可增强老年用户对语义的预测,以抑制干扰信息,消解歧义。

3.2 采用灵活包容的信息识别策略

相对于年轻人,老年用户在语言输出环节的认知能力老化比较明显,因而在智能电视的语音识别、文本分析和语义理解环节,要给予老年用户更多的包容性。随着年龄的增长,老年用户信息处理需要的时间逐步加长,智能电视要根据语音输入的环测,给予用户灵活、适度的语音采集区间,以防因固定的时长设置结束导致用户指令录入不全。因地域口音、语音边界不清、用户指令的多元化等原因导致的表述问题也相对较多,因此智能电视交互系统要适度提升容错性,对输入语音采取适度的自动校正策略,扩大相似或相近语音区域的激活,并以上下文、结构化列表等形式对用户指令做出合理推测。记忆库的建立也是提升信息识别包容性的策略之一,除相似问题的循环应对记忆外,语音交互系统可适度增加示例语料库的容量,在老年用户再次开启相似对话时优先提示,以减少交互循环的冗余;还可考虑在智能电视常用的功能区域,建立用户语言习惯与系统数据之间的智能映射,记忆用户的口语发音和识别结果,不断增加人机之间的交互匹配程度,提升系统语义理解的准确率。另外,在用户需求相对模糊的情况下,灵活主动地开启预测式对话,并及时提供可选择的反馈方案,帮助老年用户有效触发可用对话,也可大幅提升用户满意度^[9]。

3.3 构造精准有效的语音输出迭代

语音交互的时间决定其输出信息的数量。听觉信息传输是线性的,对短时记忆的要求较高,为减少用户认知负荷,现有智能电视使用的主流策略是利用视觉表现弥补听觉感知的不足。视听界面的融合可以加强用户信息的感知强度和感知数量,丰富信息加工材料,提供更多线索以帮助其在信息加工中剔除干扰、消解歧义,支持用户对交互内容的认知、意图的推测和意义的构建。在视听融合策略的支持下,针对老年用户,需要对智能电视的反馈做进一步的输出迭代,以保证其信息获取和加工的有效性及准确率。首先是降低对话中句子的复杂程度,避免句式加工超出老年用户的语言认知能力范围。同时,在多轮对话中可建立层级明确的推进模式,以免对话中断或无效重复,增强交互进行的可持续性和交流完成的完整性。其次,要做适当的目标信息标记和语义提示,保持交互目标的一致性和聚焦性。在系统的语音合成中,通过语气、语调加强对目标信息的标记,以集中老年用户的注意力,聚焦对象指令;在视觉界面中,可适当加强对重点词汇的标记,并保留历史对话中的关键信息,扩充语音交互的空间阈,方便老年用户语言认知中的反复、

回跳,帮助其在上下文的提示下保持交互目标的一致性。在语义提示中设置示例也是变相的输出迭代,其一方面可以提供交互样本、有效聚焦问题,提升人机交互目标的一致性;另一方面又可帮助老年用户及时提取有效词汇,缓解“舌尖效应”,提升和优化用户体验。

4 结语

智慧养老的背景下,语音交互因其简单自然的操作方式迅速成为解决智能产品适老化的关注重点,但其在传输特征、信息容量、结构化驱动方面仍有不可忽视的劣势。从老年用户语言认知能力衰退的角度出发,在分析其使用需求的基础上,适度考虑 VUI 设计中的补偿性策略,可有效减少老年用户与智能电视人机交互的障碍,提升语音交互的绩效,优化老年用户的智能产品使用体验,从而增进他们对智慧生活的向往,为他们享受信息时代带来的生活便利提供必要支持。

参考文献:

- [1] 韩笑,梁丹丹.正常老化脑的语言加工及其自适应机制[J].当代语言学,2019,21(4):586-601.
HAN Xiao, LIANG Dan-dan. Language Processing and Its Adaptive Mechanism of Normal Aging Brain[J]. Contemporary Linguistics, 2019, 21(4): 586-601.
- [2] PICKERING M J, GARROD S. An Integrated Theory of Language Production and Comprehension[J]. The Behavioral and Brain Sciences, 2013, 36(4): 329-347.
- [3] YONAN C A, SOMMERS M S. The Effects of Talker Familiarity on Spoken Word Identification in Younger and Older Listeners[J]. Psychology and Aging, 2000, 15(1): 88-99.
- [4] HARRIS W H. Cognition, language and aging[M]. Amsterdam: John Benjamins Publishing Company, 2016.
- [5] 王丽红,白学军,闫国利,等.词频和语境预测性在老年人阅读中的作用:眼动研究[J].中国老年学杂志,2012,32(16):3503-3507.
WANG Li-hong, BAI Xue-jun, YAN Guo-li, et al. The Role of Word Frequency and Word Predictability in Reading of Elderly People[J]. Chinese Journal of Gerontology, 2012, 32(16): 3503-3507.
- [6] DEDE G. Sentence Comprehension in Older Adults: Evidence for Risky Processing Strategies[J]. Experimental Aging Research, 2014, 40(4): 436-454.
- [7] VALENTE A, LAGANARO M. Ageing Effects on Word Production Processes: An ERP Topographic Analysis[J]. Language, Cognition and Neuroscience, 2015, 30(10): 1259-1272.
- [8] WRIGHT H H, CAPILOUTO G J, KOUTSOFTAS A. Evaluating Measures of Global Coherence Ability in Stories in Adults[J]. International Journal of Language & Communication Disorders, 2013, 48(3): 249-256.
- [9] SANTOS R, ABREU J, BEÇA P, et al. Voice Interaction on TV: Analysis of Natural Language Interaction Models and Recommendations for Voice User Interfaces[J]. Multimedia Tools and Applications, 2020, 79(47-48): 35689-35716.