

基于生理电的移动视频卡顿占比体验的研究

谭浩¹, 梁晨曦², 孙家豪², 马梦云²

(1.湖南大学 先进车身设计制造国家重点实验室, 长沙 410082;

2.湖南大学 设计艺术学院, 长沙 410082)

摘要: **目的** 卡顿因素对移动视频体验的影响进行测量研究。**方法** 以人的生理电信号与主观量表相结合的方式, 通过人类生理信号对卡顿的有效反应进行分级研究。**结果** 得出人类对移动视频卡顿的最大忍耐限度, 以及不同设备对卡顿因素的影响。**结论** 不同设备对卡顿占视频时间比的体验影响不大; 卡顿占比极限为卡顿时间占视频时间比 20%~25%。

关键词: 用户研究; 生理电信号; 主观量表; 移动视频体验; 卡顿

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2017)02-0120-07

Mobile Video Lag-Time Proportion Experience Based on Physiological Signal

TAN Hao¹, LIANG Chen-xi², SUN Jia-hao², MA Meng-yun²

(1.State Key Laboratory of Advanced Design and Manufacture for Vehicle Body, Hunan University, Changsha 410082, China; 2.School of Design, Hunan University, Changsha 410082, China)

ABSTRACT: It aims to conduct research on the influence of the mobile video lag-time proportion on experience. Based on the combination of human physiological signal and subjective scale, lag effective response is classified by human physiological signals. We conclude the endurance limit of mobile video lag-time and the influence of the lag-time on different devices. There is little influence on the experience of the different devices. The limit of the lag-time proportion is between 20%~25%.

KEY WORDS: user research; physiological; subjective scale; mobile video experience; lag

移动互联网业务已经成为人类消费生活的重要组成部分。目前,在这之中属于内容类的视频点播服务是运营商所关注的,对移动互联网视频日趋重要的项目。本文研究的对象即为视频的卡顿占视频时间比,目的在于探索卡顿对于用户观看移动视频体验的影响,进而判断受单一因素即卡顿占比影响的移动视频质量。

1 卡顿因素的研究现状

视频卡顿时长是指因客户端或流媒体的处理能力,或因网络传输质量等原因导致用户视频任务出现卡顿的持续时间的长度。通常一个视频播放过程中会出现一次或以上的卡顿。视频卡顿时长占比指一个视

频播放过程中,除了首次缓冲所需时间外,其余播放时出现的暂停时间总和占视频总时长的比例^[1]。

目前,关于在线视频质量体验的研究主要集中在通信学科,但研究卡顿占视频时长比这一概念的较少,多数情况下对于此课题的客观研究方法主要是利用通信测试技术,即视频压缩及网络传输特征参数来评估视频质量的方法,但该方法并不能直接反应用户的心理情感感受。而主观测试方法多为主观评价,如 Fidel Liberal 于 2009 年网页质量评估的研究中利用五点量表进行评价^[2],通过问卷或者访谈的形式,在用户与产品交互过程结束后收集用户的主观感受。这种方式会受到主试效应和社会赞许效应的影响,最重要的是忽略了用户感受会在使用过程中产生多重变化

收稿日期: 2016-11-01

基金项目: 华为—湖南大学 体验极限研究; 华为—湖南大学 基于人体感知的移动视频和网页浏览体验的 MOS 研究

作者简介: 谭浩(1977—),男,四川人,博士,湖南大学副教授,主要研究方向为工业设计。

的问题^[3]。为解决这一问题,引入电生理技术无疑是一种新的尝试^[4]。

2 生理信号的研究成果与应用

在心理学研究中,基本情绪主要表现为和人的心理需要相联系的一种内心体验。根据费尔和罗塞尔的大规模调查可以看出,几乎所有受试者都把情绪定义为他们的感受或体验^[5],而我国心理学家孟绍兰将情绪定义为一种内在体验和外在表情^[6-7]。根据 James-Lange 的情绪理论,情绪是对刺激引起的外周生理变化感知的结果,每种情绪体验都伴随着生理指标的变化。负性情绪是个体面对负性的刺激或心理期待与现实产生差距时,所产生的悲伤、不安、愤怒、痛苦等情绪,代表个体对某种负性的或厌恶的情绪体验的程度^[8]。负性情绪对个体产生多方面的影响,如 1995 年孟绍兰与徐景波发现在负性情绪下被试心率显著增加,指尖脉搏容积显著下降;2006 年贾静影片片段诱发情绪发现悲伤情绪引发皮肤电电位增高。同时这种生理变化也存在个体差异,如 1995 年孟绍兰与徐景波等人发现负性情绪下女性心率反应比男性高,皮电基础水平和反应水平也比男性低^[9-10](本文实验基于个体数据做比较)。其中,葛燕等人在《电生理测量在用户体验的应用》中详细归纳了心电和皮电近年来在心理学研究领域的进展^[11]。主观评价与生理信号和认知情感测量的关系见图 1。

2.1 心电

心电中的心率变异性表示心脏跳动快慢的变化情况,由两个相邻的 R-R 间期的时间长短确定,是人机交互过程中监控心理负荷和情绪状态的有效指标^[4, 12-13]。Ward 和 Marsden 曾在 2003 年发现被试在使用良好设计的网站和糟糕设计的网站时心电指标有所不同^[14]。而 Lin, Imamiya 和 Mao, 在 2008 年采用生理测量来评估用户玩游戏时的感受和行为,发现心率变异性对游戏难度变化敏感,与任务完成时的心理负荷成正比^[15]。

2.2 皮电

皮电指的是皮肤表面汗腺由于应激或其他刺激引起的激活而导致的电传导能力的变化。皮肤电导是皮肤电活动的通用指标,在皮电与情绪体验方向的研究发现,皮肤电导与唤醒水平呈线性相关^[16],可反映用户的情绪唤醒度。2006 年, Mandryk, Inkpen 和 Calvert 探索了使用生理指标评测用户玩电脑游戏时的情感体验的有效性。他们要求用户对自我的无聊,兴奋等情绪主观评分,结果发现皮肤电导反应与兴奋度相关^[17],因此在前人的研究基础上,本文的客观评价方法采用生理仪采集用户的心率变异性与皮电数

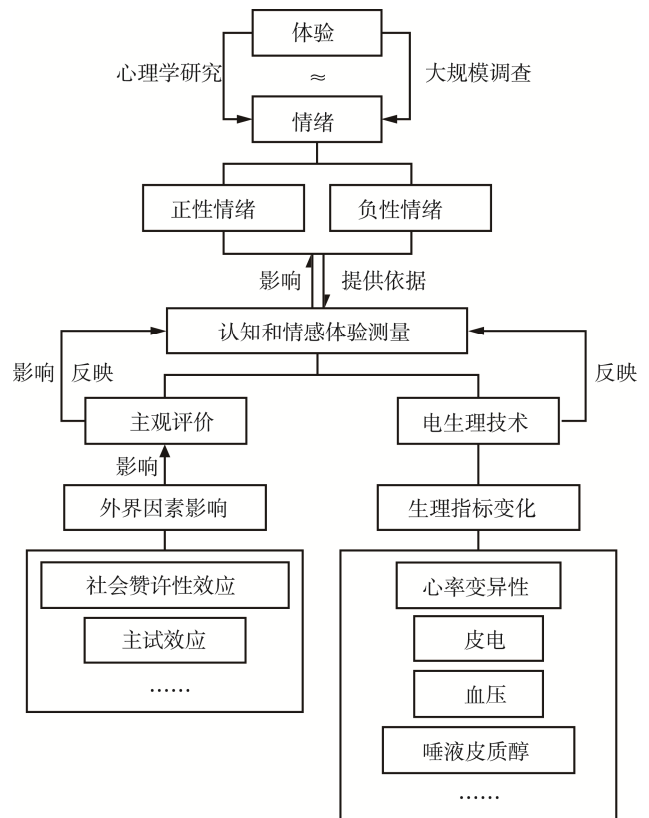


图 1 主观评价与生理信号和认知情感测量的关系

Fig.1 The relationship between the cognitive emotion measurement and the subjective evaluation or the physiological signal

据,同时使用五点量表(1~5分,可保留两位小数)打分来进行主观测试,通过运用主观、客观相结合的方法,来反映用户对视频业务的主观感受。

3 实验方法

3.1 实验被试与设备

本次实验环境为室内实验室环境,噪声恒定为 50 Hz 左右,光照 100 LUX 左右。实验设备为 BioLab 多导生理仪,用于测量记录被试生理数据;华为 mate7 手机一部, iPad Air 一部,用于播放实验材料并研究不同设备下实验结果的变化;摄像机一台,用于记录实验过程;视频分值主观测评表,用于给用户为每段实验素材评分。

实验被试来自于湖南大学师生,18 名男性,16 名女性,通过初步筛选合格被试 30 人,男性女性各 15 人。被试均经常上网,平均年龄 6 年,拥有一部以上的移动可联网设备,其中 80%的被试经常使用移动设备观看视频。

3.2 实验素材

本次实验的素材在分辨率上选取 4K 的高清视频,目的是为了排除视频不清晰对被试情绪造成的负面影响。在内容上,选取多个内容的视频为素材,目的是

让用户在实验选取一个最感兴趣的素材，以此方式排除内容对用户情绪的影响。内容选取方式是按照 Alexa 网站近 3 个月（2015 年 7 月至 10 月）的流量排名选取国内前 9 大视频网站，分析得出用户常用的，

具备差异性的，内容影响较小的并且用户覆盖面广的 6 大类别：新闻、综艺与娱乐、原创、动画、体育和纪录片。每段视频时长为 5 min。素材列表见表 1。

表 1 素材列表
Tab.1 The table of the materials

素材类别		卡顿占视频时长比						总时长/min
新闻	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5
综艺&娱乐	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5
原创	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5
动画	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5
体育	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5
纪录片	1%/3 s	5%/15 s	10%/30 s	15%/45 s	20%/60 s	25%/75 s	30%/90 s	5

3.3 实验方法

在实验前，被试对本次实验的目的，要求和方法进行了解，选取视频素材。实验人员为被试黏贴生理信号收集贴片，同时播放优雅舒缓的音乐，让用户平复心情，准备实验。实验前，先记录被试 5 min 的平静状态下的生理数据做基准值。

卡顿实验模仿视频卡顿情况，实验前用户选择自己最感兴趣的视频类别，以此减少由内容产生的负面情绪。实验流程示意图 2：屏幕显示视频播放页面，同时生理仪开始记录生理数据，用户点击“播放按钮”，页面开始正常播放 30 s，之后开始多次卡顿，

卡顿次数随机，范围为 5~7 次。每次卡顿时间随机，随机范围为 1~5 s，卡顿间隔时间随机，随机范围 1~30 s 不等，每一次卡顿加卡顿间隔为一卡顿周期，规律循环。当卡顿时间占比达到预期时，正常播放余下视频直至结束。之后实验人员提示请为该段视频体验打分，用户打分之后进入下一个材料。卡顿占视频时间比（预期卡顿占比）共 7 类，分别为：1%、5%、10%、15%、20%、25%和 30%。顺序测试，生理仪在这段时间内收集数据（心率变异性 and 皮电）。当被试为某段素材评分为 1（最差）时，则默认之后的材料主观量表分数为 1 并结束该实验。

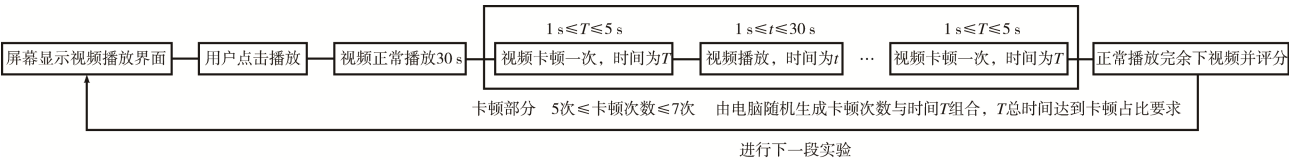


图 2 实验流程示意
Fig.2 The schematic diagram of experimental process

4 数据处理

4.1 处理生理数据得到变化点

用户在观看素材时会产生生理信号变化，当心率变异性和皮电都发生变化时，则该变化为有效变化（以下简称变化点）。

心率变异性用频域分析方法，变化与否以 T 检验方法判定，分析软件 SPSS11.0。从第一段素材所得数据开始，与基准值（来自第一段实验前平静时的素材数据）做 T 检验，P 值小于 0.05 时为显著变化。依次对比，直至找到第一个 P 值小于 0.05 的素材，此素材为心率变异性变化的第一个素材（以下简称第一个心电变化素材），之后的素材数据依次与第一个心电变化素材数据做 T 检验，直至找到第二个心率变异

性变化的素材。之后的心电变化素材寻找方法同上，直至所有素材都做过 T 检验。

皮电变化判定与心率变异性变化判断相同，利用 T 检验方法，数据为每段素材的皮电数据。T 检验对象匹配见图 3。某被试实验皮电原始图像见图 4。

由于被试的变化点通常情况下大于一个，故按照每位被试的变化点出现顺序进行编号：变化点 1，变化点 2，变化点 3 等，以此类推。变化点所在素材称为变化素材。通过初步统计可得，30 位被试中 20 位被试（约 67%）变化点个数为 3 个，6 位被试（约 20%）变化点多于 3 个，4 为被试（约 13%）变化点少于 3 个。

4.2 生理变化点验证关键主观分值的准确性

本次实验中的主观量表打分为 1~5 分制，可保留两位小数，所有整数分（5 分，4 分，3 分，2 分，

1 分) 为关键分。把握好关键分的准确性就能极大程度上提高整体主观分值的可信度。

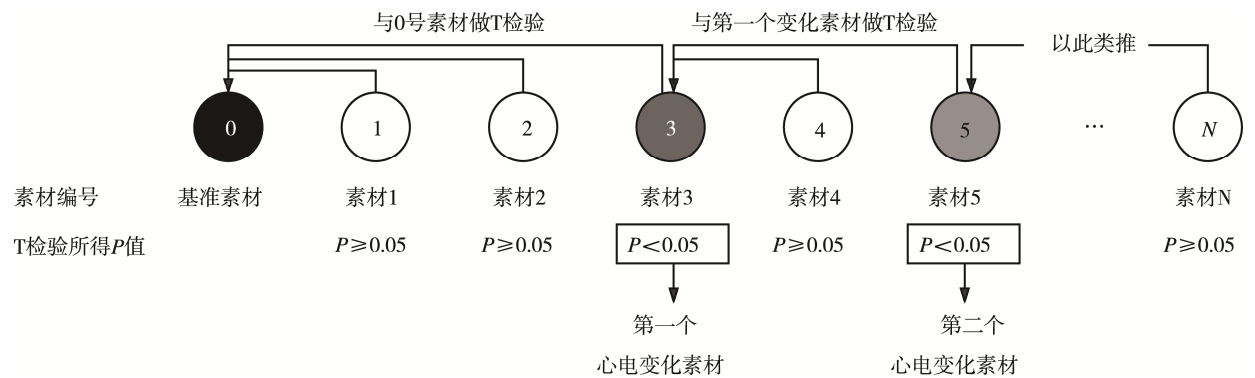


图 3 T 检验对象匹配
Fig.3 The matching pattern of T test

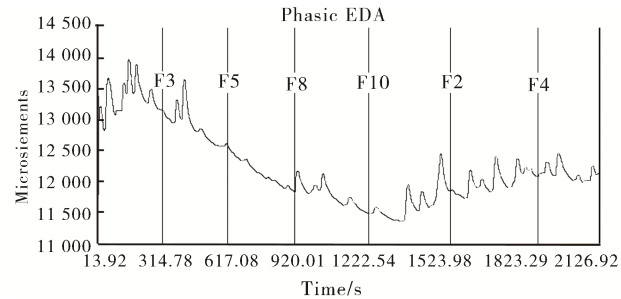


图 4 某被试实验皮电原始图像
Fig.4 The EDA original image of some testee

在该主观分系统下, 5 分对应视频的卡顿没有给被试带来任何负面体验的状态, 也就是人类的中常状态, 即无感知, 是情绪变化的起点。1 分对应视频的卡顿给被试带来极大负面体验的状态。余下的 4 分, 3 分, 2 分 3 个分值对应从无感知到极差体验的过度状态, 并且把主观分数序列划分成 4 个部分, 即 5 ~ 4 分, 4 ~ 3 分, 3 ~ 2 分, 2 ~ 1 分。

再从生理角度看, 无感知状态下的生理数据无变化, 因此 5 分不对应变化点, 变化点 1 可大致对应 4 分, 变化点 2 可大致对应 3 分, 变化点 3 可大致对

应 2 分, 而在实验环境下很难达到极大负面体验的状态, 因此 1 分值也极少有对应的变化点。

综合以上可以发现, 3 个变化点的情况与 5 分制刚好形成对照关系, 即可以认为若某被试的变化素材的分值不对应生理变化点应有分值, 则认为该被试的素材主观分值可信度欠缺, 需要从分值库中排除 (仅排除可信度欠缺项, 不影响该被试的其他评分。)

当变化点多于 3 个时, 则需要查看这些变化点的主观分值, 主观分值最接近 4 分、3 分或 2 分的素材才为变化素材。若无变化点接近 4 分、3 分或 2 分, 则认为该数据无效。

当变化点少于 3 个时, 同样需要查看这些变化点的主观分值, 主观分值接近 4 分、3 分或 2 分的素材为对应分数的跨级素材。生理变化点与主观分值关系见图 5。

4.3 为素材赋值

被试为每段素材都进行了评分, 生理数据对变化素材的关键分进行了验证和排除, 可统计得出每段素材的平均分和每个实验的变化点统计表。实验平均分见表 2, 实验变化点见表 3。

被试情绪状态&主观评分		理论变化点位置	实际变化点位置	变化点不足或多余3个
中常状态☐无感知 过度状态 极大负面体验状态	5分			
	4分	变化点1	变化点1	变化点1 (3.75分) ✗
	3分	变化点2	变化点2	变化点1 (3.75分) ✓
	2分	变化点3	变化点3	变化点2 (2.10分)
	1分			变化点3
		在整数位置小幅度浮动		评分接近4、3或2的点视为该位置的变化点

图 5 生理变化点与主观分值关系
Fig.5 The relationship of the physiological change point and the subjective scores

表 2 实验平均分
Tab.2 The table of the average score

手机（华为 Mate7）							
素材	卡顿占比 1%	卡顿占比 5%	卡顿占比 10%	卡顿占比 15%	卡顿占比 20%	卡顿占比 25%	卡顿占比 30%
平均分	4.82	4.30	4.11	3.32	2.04	1.55	1.08
平板（iPad Air）							
素材	卡顿占比 1%	卡顿占比 5%	卡顿占比 10%	卡顿占比 15%	卡顿占比 20%	卡顿占比 25%	卡顿占比 30%
平均分	4.90	4.45	4.24	3.40	2.55	1.60	1.12

表 3 实验变化点
Tab.3 The table of the change material

手机（华为 Mate7）							
素材	卡顿占比 1%	卡顿占比 5%	卡顿占比 10%	卡顿占比 15%	卡顿占比 20%	卡顿占比 25%	卡顿占比 30%
变化点 1	1 个	9 个	9 个	7 个	2 个		
变化点 2		5 个	7 个	12 个	4 个		
变化点 3			1 个	10 个	15 个	2 个	
平板（iPad Air）							
素材	卡顿占比 1%	卡顿占比 5%	卡顿占比 10%	卡顿占比 15%	卡顿占比 20%	卡顿占比 25%	卡顿占比 30%
变化点 1	1 个	8 个	12 个	4 个	4 个		
变化点 2			9 个	16 个	3 个		
变化点 3			4 个	6 个	14 个	4 个	

4.4 统计变异素材

通过算数平均数方式计算出具有普遍代表意义的变化素材与分数，具体方法为：变化点 1 的素材卡顿比例与对应变化点个数的乘积值累加并除以变化

点 1 的个数总和，得到变化点 1 的卡顿比例，变化点 2、3 的卡顿比例计算方法同上；变化点 1 的素材得分与对应变化点个数的乘积值累加并除以变化点 1 的个数总和得到变化素材 1 的得分，变化素材 2、3 得分计算方法同上。实验变异素材得分见表 4。

表 4 实验变异素材得分
Tab.4 The table of the score of the change materials

手机（华为 Mate7）				
变化素材	算数平均分	素材区间	分数原始区间	分数修正区间
15%（15%）	3.16	[1% ， 15%)	[5， 4]	[4.82， 3.16)
20%（17.68%）	2.66	[15%， 20%)	[4， 3]	[3.16， 2.66)
25%（23.21%）	1.75	[20%， 25%)	[3， 2]	[2.66， 1.75)
		[25%， 30%]	[2， 1]	[1.75， 1.08]
平板（iPad Air）				
变化素材	算数平均分	素材区间	分数原始区间	分数修正区间
15%（15.34%）	3.30	[1%， 15%)	[5， 4]	[4.90， 3.30)
20%（18.93%）	2.72	[15%， 20%)	[4， 3]	[3.30， 2.72)
25%（23.21%）	1.99	[20%， 25%)	[3， 2]	[2.72， 1.99)
		[25%， 30%]	[2， 1]	[1.99， 1.12]

注：1.原始区间指主观评分 1~5 分划分的 4 个分数区间，即 1~2 分，2~3 分，3~4 分，4~5 分。分数修正区间是综合了所有有效被试实验数据，利用算数平均数方式计算得出的平均分值，该分值更具有普遍性。2.变化素材中括号“()”中的数字为算数平均数，括号“()”外的分数是在实验素材中选择的接近于该平均数的素材值。3.素材区间最后一行最后一个素材不代表 1 分，而是实验素材的最后一个素材。4.所有数据保留两位小数。

5 实验结论

5.1 卡顿占视频时间比结论

在卡顿占视频时间比实验中,无论用手机或平板设备观看,人们都认为卡顿占视频时间比大约为 15% 的素材可以达到 4 分水平,即若想让用户感到比较满意(4~5 分)则视频要保证卡顿时间小于等于视频总长的 15%;卡顿占视频时间比大约为 18% 的素材可以达到接近 3 分的水品,即若想让用户可以坚持观看,视频的卡顿时间要小于等于视频总长的 18% 比较安全;卡顿占视频时间比大约为 23% 时可以达到接近 2 分的水平,即 23% 略高于用户的忍耐极限,视频的卡顿占总时长比的最糟糕情况不要高于 23%。卡顿比例与分值见图 6。

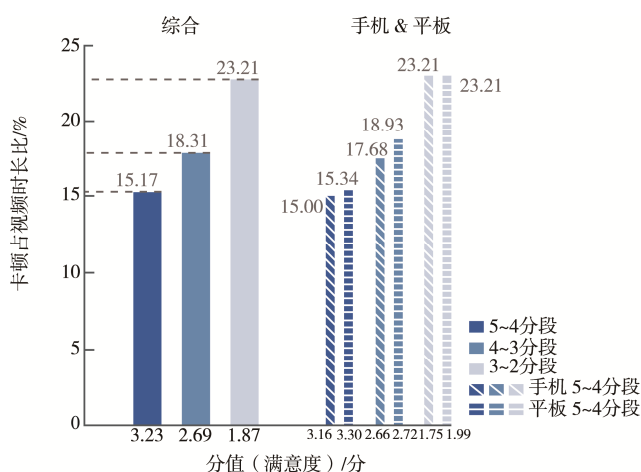


图 6 卡顿比例与分值

Fig.6 The proportion of lag-time and the scores

5.2 设备对比结论

从表 3 和图 6 可以看出,手机(华为 Mate7)和平板(iPad Air)在卡顿占比实验中差异不明显,从实验所准备的素材来看大致的划分等级都在 15%、20%和 25%占比的素材上,但是两者有微小的差异。

同样是 4~5 分区间,平板的卡顿时长占比较手机多 0.34%;在 3~4 分区间内,手机和平板按照实验素材都被划分到卡顿占 20% 的视频素材中去,但是实际占比平板要比手机多 1.25%;但是在 2~3 分区间内,平板和手机的算数平均数值相同。这说明在使用平板时人们的忍耐限度更为宽松,这一点在平均分上也可以得到印证,即在相同的素材下,平板的主观评分总是略高于手机的评分。但是在 2 分档时,两种设备的差距变的不明显,猜想是由于此时频繁的卡顿和越来越高的卡顿占比体验已经变为影响体验的重要因素,超过了由设备带来的不同体验。

6 结语

近年来,对于视频体验人因工程方面的研究越来越多,本研究突破以往单纯运用主观评价的方式,结合了心理学研究中成熟的生理信号成果,为今后的人因工程用户体验研究方法打开了一个新方向。当然本次研究中也存在一些不足,首先,在实验被试的选取上有趋同的嫌疑(都为湖南大学师生),可能会造成结果偏差;第二,没有进一步细化素材样本,将实验中比较接近的 15%~20%部分的素材进一步划分进行实验。但是本次研究获得的实验数据同样可以为移动通信业务提供可靠的参考,让今后的视频业务更加符合用户期望。

参考文献:

- [1] 王洗,廖振松,胡玉祥. 4G 视频业务用户感知评估体系研究[J]. 信息通信, 2015(3): 230—231.
WANG Xian, LIAO Zhen-song, HU Yu-xiang. Research on Evaluation System of 4G Video Business User Perception[J]. Information & Communications, 2015(3): 230—231.
- [2] LBARROLA E, LIBERAL F, TABOADA I, et al. Web QoE Evaluation in Multi-agent Networks: Validation of ITU-T G.1030[C]//2009 Fifth International Conference on Autonomic and Autonomous Systems. IEEE Computer Society, 2009: 289—294.
- [3] VERMEEREN A P O S, LAW E L, ROTO V, et al. User Experience Evaluation Methods: Current State and Development Needs[C]. New York: Paper Presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, 2010: 521—530.
- [4] DIRICAN A C, GOKTURK M. Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction[J]. Procedia Computer Science, 2011(3): 1361—1367.
- [5] 吴攀. 基于用户体验的交互式网络视频广告研究[D]. 上海: 上海师范大学, 2012.
WU Pan. Research on Interactive Network Video Advertisement Based on User Experience[D]. Shanghai: Shanghai Normal University, 2012.
- [6] 孟绍兰. 情绪心理学[M]. 北京: 北京大学出版社, 2005.
MENG Shao-lan. Emotional Psychology[M]. Beijing: Peking University Press, 2005.
- [7] 浦江. 全认知情感理论——一种新的心智计算模型[J]. 计算机科学, 2014, 41(7): 15—24.

- PU Jiang. Full Cognitive Affective Theory: a New Mental Computing Model[J]. Computer Science, 2014, 41(7): 15—24.
- [8] 高晶. 正负性情绪调节的生理反应及主观体验研究[D]. 沈阳: 辽宁师范大学, 2009.
- GAO Jing. Research on the Physiological Reaction and Subjective Experience of Positive and Negative Emotion Regulation[D]. Shenyang: Liaoning Normal University, 2009.
- [9] 徐景波, 孟昭兰, 王丽华. 正负性情绪的自主生理反应实验研究[J]. 心理科学, 1995(3): 134—139.
- XU Jing-bo, MENG Shao-lan, WANG Li-hua. The Experimental Study of Autonomic Physiological Responses to Positive and Negative Emotions[J]. Psychological Science, 1995(3): 134—139.
- [10] 刘贤敏. 中国古典音乐诱发情绪的生理活动研究[D]. 南京: 南京师范大学, 2006.
- LIU Xian-min. Research on Physiological Activity of Emotion Evoked by Chinese Classic Music[D]. Nanjing: Nanjing Normal University, 2006.
- [11] 葛燕, 陈亚楠, 刘艳芳, 等. 电生理测量在用户体验中的应用[J]. 心理科学进展, 2014, 22(6): 959—967.
- GE Yan, CHEN Ya-nan, LIU Yan-fang, et al. Application of Electrophysiological Measurement in User Experience[J]. Advances in Psychological Sciences, 2014, 22(6): 959—967.
- [12] HERCEGFI K. Heart Rate Variability Monitoring during Human-computer Interaction[J]. Acta Polytechnica Hungarica, 2011, 8(5): 205—224.
- [13] SONDEREGGER A, SAUER J. The Influence of Laboratory Set-up in Usability Tests: Effects on User Performance, Subjective Ratings and Physiological Measures[J]. Ergonomics, 2009, 52(11): 1350—1361.
- [14] WARD R D, MARSDEN P H. Physiological Responses to Different WEB Page Designs[J]. International Journal of Human-Computer Studies, 2003, 59(1-2): 199—212.
- [15] LIN T, IMAMIYA A, MAO X Y. Using Multiple Data Sources to Get Closer Insights into User Cost and Task Performance[J]. Interacting with Computers, 2008, 20(3): 364—374.
- [16] GANGLBAUER E, SCHRAMMEL J, DEUTSCH S, et al. Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility[C]. Workshop on User Experience Evaluation Methods in Product Development, 2009.
- [17] MANDRYK R L, INKPEN K M, CALVERT T W. Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies[J]. Behaviour & Information Technology, 2006, 25(2): 141—158.