https://doi.org/10.5392/JKCA.2018.18.08.092

반자율주행 맥락에서 AI 에이전트의 멀티모달 인터랙션이 운전자 경험에 미치는 효과 : 시각적 캐릭터 유무를 중심으로

The Effect of AI Agent's Multi Modal Interaction on the Driver Experience in the Semi-autonomous Driving Context : With a Focus on the Existence of Visual Character

서민수, 홍승혜, 이정명
연세대학교 정보대학원 UX Lab

Min-soo Suh(zeclix@yonsei.ac.kr), Seung-Hye Hong(seunghye_ux@yonsei.ac.kr), Jeong-Myeong Lee(ux.jmlee@gmail.com)

요약

대화형 AI 스피커가 보편화되면서 음성인식은 자율주행 상황에서의 중요한 차량-운전자 인터랙션 방식으로 인식되고 있다. 이 연구의 목적은 반자율주행 상황에서 음성뿐만 아니라 AI 캐릭터의 시각적 피드백을 함께 전달하는 멀티모달 인터랙션이 음성 단일 모드 인터랙션보다 사용자 경험 최적화에 효과적인지를 확인하는 것이다. 실험 참가자에게 주행 중 AI 스피커와 캐릭터를 통해 음악 선곡과 조정을 위한 인터랙션 태스크를 수행하게 하고, 정보 및 시스템 품질, 실재감, 지각된 유용성과 용이성, 그리고 지속 사용 의도를 측정하였다. 평균차이 분석 결과, 대부분의 사용자 경험 요인에서 시각적 캐릭터의 멀티모달 효과는 나타나지 않았으며, 지속사용 의도에서도 효과는 나타나지 않았다. 오히려, 정보품질 요인에서 음성 단일 모드가 멀티모달보다 효과적인 것으로 나타났다. 운전자의 인지적 노력이 필요한 반자율주행 단계에서는 멀티모달 인터랙션이 단일 모드 인터랙션에 비해 사용자 경험 최적화에 효과적이지 않았다.

■ 중심어 : | 대화형 음성 에이전트 | 멀티모달 융합 | 멀티모달 인터랙션 | 에이전트 체화 | 반자율주행 자동차 | 시뮬레이터 실험 |

Abstract

As the interactive AI speaker becomes popular, voice recognition is regarded as an important vehicle-driver interaction method in case of autonomous driving situation. The purpose of this study is to confirm whether multimodal interaction in which feedback is transmitted by auditory and visual mode of AI characters on screen is more effective in user experience optimization than auditory mode only. We performed the interaction tasks for the music selection and adjustment through the AI speaker while driving to the experiment participant and measured the information and system quality, presence, the perceived usefulness and ease of use, and the continuance intention. As a result of analysis, the multimodal effect of visual characters was not shown in most user experience factors, and the effect was not shown in the intention of continuous use. Rather, it was found that auditory single mode was more effective than multimodal in information quality factor. In the semi-autonomous driving stage, which requires driver 's cognitive effort, multimodal interaction is not effective in optimizing user experience as compared to single mode interaction.

■ keyword : | Conversational AI | Multimodal Fusion | Multimodal Interaction | Agent Embodiment | Semi-autonomous Vehicle | Simulator Experiment |

접수일자 : 2018년 07월 05일
수정일자 : 2018년 07월 26일
심사완료일 : 2018년 07월 27일
교신저자 : 서민수, e-mail : zeclix@yonsei.ac.kr

I. 서 론

자율주행 기술의 발전으로 인해 운전자는 주행 중 다른 콘텐츠를 함께 이용할 가능성이 커졌다. 대화형 음성 에이전트는 이러한 자율주행 환경에서 운전자의 운전 외 콘텐츠 이용을 도와줄 인터랙션 기술로 주목받고 있다. 대화형 음성인식 인공지능이란 음성을 통해 사용자와 상호작용하는 기술로, 시스템의 메시지가 사용자에게 음성으로 전달되는 동시에 사용자의 반응을 유도하는 시스템이다[1][2]. 대화형 음성인식 인공지능이 적용된 에이전트는 사용자에게 에이전트가 실제와 같거나 믿을 만한 것으로 보이게 하며, 더 자연스럽고 빠르게 일을 수행하도록 도와준다[3][4].

대화형 음성 에이전트의 이러한 특성은 자율주행 환경에서의 사용자 경험에 최적화되어 있다. 우선, 운전자가 운전과 운전 외 콘텐츠를 동시에 이용하는 데 들이는 인지적 노력을 줄여줄 수 있기 때문이다. 주행중 발생하는 운전자의 주의 분산은 사고의 위험 요소인데, 특히 운전자가 엔터테인먼트 및 정보 콘텐츠를 이용할 때 주의가 산만해진다. 이 때 음성 에이전트는 인지 과정이 단순하여 운전자의 상황 인식 능력을 향상시키며, 인지 부하를 줄여 사고 위험을 낮추는 역할을 한다[5][6]. 또한, 자율주행 환경에서 운전자에게 기술에 대한 신뢰감과 같은 긍정적인 경험을 줄 수 있다. 자율주행 자동차의 운전 경험을 위해서는 신뢰 요소가 중요하고, 에이전트는 운전자의 신뢰감 형성에 긍정적인 영향을 주는 데 큰 역할을 한다[7].

음성 외 다른 인터페이스 요소를 함께 활용한 멀티모달 인터랙션(multi modal interaction)은 일반적인 환경에서 대화형 음성 에이전트의 효과를 강화할 수 있다. 모바일 환경에서 음성과 시각 인터페이스를 연계하여 개선했을 때 음성 에이전트 사용자의 학습용이성과 만족도가 증진되었다[8]. 대화형 음성 에이전트의 의인화, 또는 캐릭터 요소를 활용한 청각-시각 복합 인터랙션이 대화형 음성 에이전트의 사용 지속 의도와 만족도 등 사용자 경험을 증진시키는 효과도 나타났다[1][9][10].

그러나, 자율주행 환경에서 대화형 음성 에이전트의 멀티모달 인터랙션 효과 연구는 미흡하다. Okamoto와 Sano(2017)는 비자율주행 환경에서 음성 인터페이스와 시각 인터페이스를 함께 제공하여 대화형 음성 에이전트에 대한 신뢰와 협업감을 향상시킬 수 있음을 연구하였다[11]. 운전자와 차량 간의 협력 관계 형성을 통해 자율주행 기술을 더 쉽게 채택하도록 차량 내 멀티모달 인터랙션이 필요하다는 것이다.

자율주행 기술의 상용화가 확대되고 있는 상황에서 자율주행 단계별로 사용자의 경험을 증진시킬 수 있는 방법에 대한 연구가 필요하다. 또한, 운전 중 콘텐츠 이용이 증가하고 차량 내 인포테인먼트 시스템의 확산 가능성에 따라 콘텐츠 이용시 사용자 경험에 대한 연구가 필요하다. 따라서 이 연구에서는 현재 상용화된 반자율주행 환경에서 엔터테인먼트 콘텐츠를 이용할 때 대화형 음성 에이전트의 효과적인 사용자 경험을 위한 멀티모달 인터랙션 효과를 검증하고자 한다.

II. 이론적 배경

이 연구의 목적은 반자율주행 상황에서 AI가 음성 단일 모드로만 대화 피드백을 전달하는 것보다 화면으로 AI 캐릭터의 시각적 피드백을 함께 전달하는 멀티모달 인터랙션이 사용자 경험 최적화에 더 효과적인지의 여부를 확인하는 것이다. 멀티모달 융합 (multimodal fusion)과 에이전트의 시각적 캐릭터 효과에 관한 HCI 이론을 기반으로 연구 문제를 도출하였다.

1. 멀티모달 인터랙션

멀티모달리티(multimodality)의 사전적인 의미는 '여러 개의 모드, 양상, 감각을 포함하거나 가지는 것'이다[12]. 멀티모달 인터페이스(multimodal interface)는 시스템과 사용자간 의사소통을 위해 음성(speech), 시각(visual), 촉각(tactile) 등과 같은 입출력 모드를 2개 이상 복합적으로 이용하여 인간과 컴퓨터가 상호 작용할 수 있도록 하는 것이다[13][14].

인간은 대화할 때 상대의 음성, 표정, 제스처, 신체접촉 등 복수의 감각 자극을 동시에 인식하고 반응한다. 멀티모달과 단일 모드 인터페이스 간의 차이점은 더 풍

부하고 자연스러운 방식으로 상호작용하는 인간 중심적인 인지적 특성에 있다. 따라서 멀티모달 인터페이스의 목적은 사용자의 지각 및 의사소통 기능을 지원하는데 있다[15]. 이러한 멀티모달 인터랙션은 단일 모드 인터랙션에 비해 빠른 작업을 가능하게 하며 작업 중 오류 감소로 시스템 신뢰도를 증가시킨다[16]. 또한, 고정된 자세에서의 운전과 같이, 특정 인터페이스에 제약이 있는 환경에서 입출력방식을 선택할 수 있고, 둘 이상의 입출력을 융합하여 신뢰성 높은 결과를 얻을 수 있는 장점도 존재한다[14].

멀티모달 인터랙션은 여러 입출력 수단을 활용하여 사용자의 의도와 주변의 상황에 따라 가장 편리한 수단을 제공하기 위해 사용된다[14]. 여러 입출력 정보를 통합하여 의사소통의 효율을 높여 사용자의 상황이나 맥락에 가장 적합한 의사소통을 가능하게 한다.

청각 인터페이스의 경우 입력 정보를 바로 직감적으로 인지할 수 있다는 장점이 있으나, 인지 기억이 낮고 주위 환경에 크게 영향을 받는다는 단점이 있다[17]. 입력 정보를 비교적 오랫동안 인지할 수 있는 시각 인터페이스를 더해 멀티모달 인터랙션의 효과를 반자율주행 자동차 맥락에서 살펴보고자 한다.

2. 에이전트 체화 (Agent Embodiment)

에이전트는 사실상 컴퓨터 알고리듬이므로 구체적인 물리적 형태가 없지만, 에이전트의 체화(embodiment)는 목소리와 같은 청각적 표현 또는 얼굴이나 몸과 같은 시각적 형태의 인터페이스를 갖는 것을 말한다[18]. 특히, 에이전트의 시각적 표현이 있을 경우 없는 경우에 비해 더 긍정적인 사회적 상호 작용을 유도한다[19].

Walker 등(1994)의 초기 연구 이래 인간 얼굴을 인터페이스로 채택한 에이전트 체화 효과 연구는 실증적 연구 결과가 많이 축적되어 왔다[20]. 예를 들어 캐릭터같이 사용자에게 친숙한 대상으로 시각화된 인텔리전트 에이전트의 경우 사용자가 시스템과 직접적인 상호작용을 하며 상대가 친숙하다고 인식하는 효과가 있다[21]. 따라서 이 연구에서는 시각화된 에이전트로 캐릭터를 활용하여 시각화된 에이전트의 유무에 따른 사용자 경험의 차이를 보고자 한다.

3. 사용자 경험 요인

사용자 경험은 시스템을 사용하는 사용자의 만족도에 영향을 미치는 모든 요소를 포함하며, 지속 사용 의도에 긍정적인 영향을 미친다[22][23]. 대화형 음성 에이전트에 멀티모달 인터랙션을 적용할 때, 인지된 실재감, 정보 및 시스템 품질, 지각된 유용성과 용이성 등을 사용자 경험 요인으로 측정할 수 있다. 실재감은 만족도와 지속 의도에 긍정적인 영향을 미친다[24]. 정보 품질과 시스템 품질은 사용자의 만족도에 영향을 미치며, 이는 사용자의 서비스 사용을 촉진한다[25]. 그리고 인지된 용이성은 만족도에 강한 영향을 미치며, 지속 사용 의도와 추천의도에도 유의미한 영향이 있다[26].

III. 연구 모델 및 조사설계

1. 연구 문제

운전자의 음악 청취와 같은 차량 내 엔터테인먼트 과업(entertainment task)에서 캐릭터의 존재유무에 따른 멀티모달 인터랙션의 효과를 확인하기 위한 실험 연구를 진행하였다. 오즈의 마법사(Wizard of Oz; WoZ) 방법을 통해 인공지능 스피커와 캐릭터를 활용하여 운전 시뮬레이터에서 주행 중 음악 선곡과 조정을 위한 상호작용 과업(interaction task)을 수행하게 하고, 사용자 경험 요인인 실재감, 정보 품질, 시스템 품질, 인지된 유용성, 용이성과 지속 사용 의도를 측정하였다.

2. 연구 설계

2.1 연구대상

연구 대상은 운전 면허를 소지하고, 신체에 결격 사유가 없으며 대화형 인공지능 스피커를 사용하는데 지장이 없는 발화능력과 청력을 가진 자로 선정하였다. 총 50명의 참가자가 참여하였고, 1명이 탈락하여 49명을 대상으로 시뮬레이터 실험 데이터를 수집하였다. 참가자 모집 시 운전 경력과 주행 거리, AI 스피커 이용 경험, 연령과 성별 구성이 두 그룹에 유사하게 배정하였다. 최종적으로 대조군인 캐릭터가 없는 그룹에 24명

(남17, 여7), 실험군인 캐릭터가 있는 그룹에 25명(남19, 여6)을 배정하였다.

2.2 변수의 조작적 정의

가. 사회적 실재감(Social Presence)

사회적 실재감은 'AI 스피커와 동반 관계에 있다고 느끼는 정도'로 정의하였으며, Ju와 Sonya(2014), Heerink et al.(2008)의 선행연구를 참조하여 본 연구에 맞게 문항을 조절하였다. 모든 설문 문항은 7점 리커트 형 척도(1=전혀 그렇지 않다, 7=매우 그렇다)로 구성하였다.

나. 정보 품질(Information Quality)

정보 품질은 '사용자가 느끼는 본 연구의 음악 추천 시스템의 정보 품질'로 정의하였으며, Delone과 McLean(2003), Yang, Cai, and Zhou(2005), 김용일과 임재문(2014)의 연구를 참조하여 본 연구에 맞게 문항을 조절하였다.

다. 시스템 품질(System Quality)

시스템 품질은 '사용자가 느끼는 본 연구의 음악 추천 시스템의 품질'로 정의하였으며, Delone과 McLean(2003), Yang et al.(2005), 김용일과 임재문(2014) 등의 연구를 참조하여 본 연구에 맞게 문항을 조절하였다. 3개의 설문 문항은 7점 리커트 척도로 구성되었다.

라. 인지된 용이성(Perceived Ease of Use)

지각된 용이성은 '사용자가 본 연구의 음악 추천 시스템을 이용하는데 많은 노력을 요구하지 않을 것이라고 느끼는 정도'로 정의하였으며, Davis(1989), Davis, Bagozzi, and Warshaw(1989), Venkatesh와 Bala(2008) 등의 연구를 참조하여 문항을 조절하였다.

마. 지각된 유용성(Perceived Usefulness)

지각된 유용성은 '사용자가 본 연구의 음악 추천 시스템을 이용시 음악 추천의 효율성이 증가될 것이라고 느끼는 정도'로 정의하였으며, Davis(1989), Davis et al.(1989), Venkatesh와 Bala(2008), 손승혜, 최윤정, 황하성(2011) 등의 연구를 참조하여 본 연구에 맞게 문항을 조절하였다.

바. 사용 의도(Intention of Use)

사용 의도는 '본 연구의 음악 추천 시스템을 계속 사용할 의도'로 정의 되었으며, Davis(1989), Davis et al.(1989), Venkatesh와 Bala(2008), 손승혜 등(2011), 남종훈(2007), Sánchez와 Hueros(2010) 등의 연구를 참조하여 본 연구에 맞게 문항을 조절하였다.

2.3 실험 환경 및 절차

연구과정은 시뮬레이터를 이용한 주행 및 AI 스피커와의 상호작용 실험 후 사후 설문을 실시하고 이후 간단한 인터뷰를 진행하였다.

먼저 참가자는 연구의 목적과 실험 진행에 대하여 간략한 설명을 들은 뒤 시뮬레이터에 착석하였다. 참가자에게 편향이 발생할 가능성을 차단하기 위해, 캐릭터의 존재 유무에 따라 두 그룹으로 나뉘었다는 사전 설명은 하지 않았고, 단순히 운전자가 운전 중 AI 스피커로 음악 콘텐츠를 이용할 때의 주의력과 관련된 연구라고 안내하였고, 설문 완료 후 인터뷰 과정에서 본 목적에 대해 다시 설명하였다.

주행속도 제한없이 시뮬레이터 파일럿 테스트를 진행한 결과, 운전자의 인지적 처리가 AI 스피커와의 소통보다는 주행 과업에만 집중되는 경향이 나타났기 때문에, 본 실험에서는 평균적인 국도 제한속도인 80km/h 근처로 주행하도록 안내하였다. 반자율주행 시스템의 일반적인 운전 조건과 같이, 약간의 전방 주시와 스티어링 휠에 최소한 한 손은 대고 인공지능과 대화하도록 안내하였다. 참가자는 시뮬레이터에 앉은 뒤, 트랙의 반 바퀴 정도 연습 주행을 실시하였고 AI 스피커 사용법을 숙지한 뒤 수차례 연습하여 익숙해진 뒤에 본 실험을 실시하였다. 실험용 시뮬레이터의 스티어링 휠은 로지텍의 G27 레이싱 휠을 이용하였고, 액셀러레이터 및 브레이크 페달은 로지텍의 드라이빙 포스 GT를 이용하여 주행 조작을 하였다. 시뮬레이터 기기로 소니의 플레이스테이션 3와 55인치 FHD TV를 이용하

여 드라이빙 게임인 그란투리스모 6의 인디애나폴리스 슈퍼스피드웨이 트랙에서 실험을 진행하였다. 장비와 세팅 선택 이유는 스티어링 휠의 회전각, 스티어링 휠에 가해지는 포스 피드백(Force feedback) 및 페달의 감각과 가속도 등이 실제 운전 환경과 가까우며 프로 레이스 드라이버 훈련에도 사용되기 때문이다. 또한 선정된 트랙은 원형에 가까워서 큰 주행 능력을 요구하지 않아 참가자로 하여금 인지적 부담을 낮출 수 있기 때문에 선택되었다. LCD 모니터의 경우 차량 내 센터페시아의 네비게이션 화면으로 가정하여, 평균 사이즈인 8인치만큼을 제외한 나머지 공간을 검은 천으로 가려 사용하였으며, 캐릭터는 오즈의 마법사(Wizard of Oz) 방식을 통하여 참가자가 인식하지 못하도록 실험자가 직접 조작하였다. 에이전트와의 친밀도 인식 효과 확인을 위해 캐릭터는 인지도와 호감도가 높은 라이언(카카오)을 활용하였다[36]. 또한, 기억 지속 시간이 낮은 청각 모드를 보완하기 위해 시각 모드를 활용하여 각 발화 상황별로 라이언 이미지 6가지를 사용하였다[37]. 이미지는 4-18프레임 정도의 짧은 애니메이션 GIF 파일이다. 블루투스 스피커를 이용하여, 참가자가 직접 명령을 내리고 노래를 실행하였다. 스피커의 응답어는 네이버 Clova앱에서 지원하는 TTS(Text-to-speech)기능을 이용하여 녹음하여 사용하였다. 명령어와 음악 종류 목록은 [표 1]과 같다. 음악은 음원 스트리밍 서비스인 멜론에서 제공하는 리스트 중에서 파일럿 테스트를 통해 직관적으로 구분이 가능하다고 판단되는 5개의 리스트를 선정하였다. 참가자는 시뮬레이터에서 주행 시작과 동시에 AI 스피커에 명령을 내리고 음악을 청취한다. 음악 선택이나 변경은 명령어 하에서 자유롭게 이용하도록 하였으며 약 10분이 지나면 스피커는 역발화를 통해 음악을 추천하고, 참가자는 해당 추천곡을 듣거나 거절할 수 있었다. 만약 듣는다고 선택한 경우, 추천곡이 끝나면 스피커는 플레이 리스트에 해당 곡을 추가할 의향이 있는지 물었다. 실험이 완료된 후 설문과 인터뷰를 진행하였다.

그림 1. 실험 환경

표 1. 음성 에이전트 캐릭터 이미지

음악 재생 명령 시	노래 재생 중	주행 시작 시
호출 후 명령 대기 중	잘못된 명령 시	음악 추천 시

IV. 연구 결과 및 분석

1. 연구 대상자의 인구 통계학적 특성

총 49명의 응답자의 연령 평균은 27.9세(표준편차: 5.42)로 25세-29세가 65.4%로 가장 많았다. 평균 운전 경력은 3.30년(표준편차: 2.99)이었다.

2. 측정도구의 타당도 및 신뢰도 분석

설문 문항의 타당도와 신뢰도를 분석하기 위하여 SPSS를 이용하여 요인분석과 신뢰도 분석을 실시하였다. 요인 분석을 통해 타당도가 낮은 요인의 문항이 제

거되었다. 모든 변수들에 대하여 공통성이 0.5 이상이고, 요인적재량 0.5 이상으로 충분한 타당도를 확보하였다. 또한 모든 요인에 대하여 Cronbach's α값이 0.7 이상으로 충분한 신뢰도를 확보하였음을 확인하였다.

3. 각 요인별 멀티모달 방식의 효과 차이

본 항목에서는 캐릭터의 존재 유무에 따라 구분된 두 그룹에 대해 각 요인들인 사회적 실재감, 정보 품질, 시스템 품질, 지각된 유용성, 지각된 용이성, 사용 의도에 차이가 있는지 확인하기 위하여 독립표본 t 검정을 시행하였다.

캐릭터 유-무 집단간 측정 변인 평균 차이를 분석한 결과는 [표 2]와 같다. 시스템 품질 인식 평균값은 캐릭터가 존재하는 그룹이 4.84(sd=1.30), 존재하지 않는 그룹이 4.71(sd=1.59), 사회적 실재감은 캐릭터가 존재하는 그룹이 평균 3.31(sd=1.46), 존재하지 않는 그룹이 3.36(sd=1.41)로 대부분의 사용자 경험 요인에서 시각적 캐릭터가 추가된 멀티모달 방식의 효과 차이는 나타나지 않았다. 반면, 정보 품질의 경우 캐릭터가 존재하는 그룹의 평균이 4.63(sd=1.15), 존재하지 않는 그룹이 5.27(sd=0.93)로 오히려 음성 단일 모드의 인터랙션이 더 효과적인 것으로 나타났다.

표 2. 캐릭터의 존재 유무에 따른 각 집단의 변수별 평균차이

변수	캐릭터 있음 (n=25) 평균 (표준편차)	캐릭터 없음 (n=24) 평균 (표준편차)	t값	p값
사회적 실재감	3.31(1.46)	3.36(1.41)	−0.13	0.90
정보품질	4.63(1.15)	5.27(0.93)	−2.14	0.04*
시스템 품질	4.84(1.30)	4.71(1.59)	0.32	0.75
지각된 유용성	4.72(1.20)	5.13(0.92)	−1.33	0.19
지각된 용이성	6.20(0.63)	6.39(0.54)	−1.12	0.27
사용의도	5.20(1.17)	5.64(0.97)	−1.43	0.16

*p < .05

또한, 지각된 유용성은 캐릭터가 존재하는 그룹이 4.72(sd=1.20), 존재하지 않는 그룹이 5.13(sd=0.92), 지각된 용이성은 캐릭터가 존재하는 그룹이 6.20(sd=0.63), 존재하지 않는 그룹이 6.39(sd=0.54)로 사용 의도는 캐릭터가 존재하는 그룹이 5.20(sd=1.17), 존재하지 않는 그룹이 5.64(sd=0.97)로 만족도 요인과 지속사용 의도에서도 시각적 캐릭터의 효과가 더 낮게 나타났으나 통계적으로 의미있는 수준의 차이는 아니었다.

4. 관찰과 사후 인터뷰 결과

시뮬레이터 실험 중 스피커의 역발화를 통한 음악 추천의 수용 여부와 추천된 음악을 끝까지 듣는지 여부를 기록하였으며, 추천곡의 재생 중단시나 종료된 경우 플레이 리스트에 넣을지에 대해 질문함으로써 3단계에 걸쳐 수용도를 측정하였다. 역발화를 통한 추천음악을 수용한 경우는 84%, 거부한 경우는 16%였으며, 수용한 경우 중 끝까지 음악을 들은 경우는 67%, 중간에 다른 곡으로 변경한 경우는 16%였다. 플레이 리스트에 추천곡을 추가한 참가자는 22%, 거부한 참가자는 61%였다.

실험과 설문 종료 후 연구의 본 목적에 대해 설명한 후 각 참가자에 대하여 인터뷰를 진행하였고, 앞서 실험에서 역발화 추천 수용과 플레이 리스트 추가 여부에 대한 참가자의 의견을 정리하고, 정리된 의견들에 대해서 공통적으로 나타나는 요소들을 중심으로 분류하였다.

첫 번째 질문은 실험 후반부에 AI 스피커의 역발화 제안에 대해 수용하거나 거부한 이유에 대한 주제로 시작하였다. 해당 질문은 시각적인 자극인 캐릭터의 존재의 유무와 관계없이 대부분의 참가자가 수용했는데 시각적 자극이 포함된 멀티모달의 효과가 나타나지 않은 것인지 확인하기 위해 질문하였다. 질문 결과 참가자들은 본인이 속한 그룹과 무관하게 음악을 수용한 것으로 나타났으며 음악 수용의 주요 요인은 개인의 음악적 취향으로 나타났다. 거부한 참가자들의 경우 추천 이전 재생된 음악들이 개인적 취향에 맞지 않았기 때문에 추천된 음악도 취향이 아닐 것이라고 판단하여 제안을 거부하였으며, 결과적으로 멀티모달의 효과는 나타나지 않은 것으로 해석된다.

두 번째 질문은 첫 번째 질문에서의 답변으로 주요 요인이었던 콘텐츠 취향을 제외하고 추천 기능 자체에 대한 것이었다. 실험 맥락이라는 한계로 선별된 음악이

참가자의 취향에 맞지 않을 수 있기에 해당 요인을 배제한 채 멀티모달의 효과를 알기 위한 목적이었다. 대부분의 참가자는 시각적 모드의 존재 유무와 관계없이 역발화 추천 기능 자체에 대해서 긍정적으로 평가했다. 이는 앞서 실시한 설문의 통계 분석에서 시스템 품질 요인에서 멀티모달 인터랙션의 효과가 나타나지 않은 점을 뒷받침하는 결과로 해석된다.

세 번째는 캐릭터의 시각적 모드가 존재했던 그룹에 한정한 질문으로 네비게이션 화면에 등장한 시각적 자극이었던 라이언 캐릭터에 대한 의견을 수집하였다. 캐릭터의 존재가 AI 스피커의 에이전트를 더 사람처럼 느껴지게 하는지, 친근감이 느껴지는가를 중심으로 질문하였다. 대다수의 참가자가 캐릭터의 존재 유무는 중요한 점이 아니라고 언급하였으며, 그 중에는 캐릭터의 존재가 친밀도를 올리는 것이 아니라 오히려 주의를 빼앗거나 의식하게 하여 주행에 방해가 된다고 부정적으로 평가하는 응답도 다수 존재했다. 이는 정보 품질 요인에서의 통계 분석 결과와도 일치하는 응답으로, 캐릭터가 존재하는 그룹에서 멀티모달 인터랙션의 효과는 음성 단일 모드에 비해서 효과적이지 않다는 것으로 해석된다.

V. 결론

연구 결과, 처치변수인 캐릭터의 존재유무에 따른 멀티모달 인터랙션의 효과에 대한 두 집단의 통계적 차이는 정보 품질에서만 나타났으나, 대부분의 변인에서 캐릭터가 없는 경우 평균이 더 높았다. 즉, 반자율주행 맥락에서 멀티모달 인터랙션보다 음성 단일 모드 인터랙션이 더 효과적으로 나타났다. 이 결과는 사후 인터뷰에서 캐릭터의 시각적 자극이 주의를 분산시킨다는 응답과도 상통한다.

즉, 반자율주행 단계에서는 멀티모달 인터랙션이 단일 모드 인터랙션에 비해 사용자 경험 최적화에 더 효과적이지 않다. 청각 요소 외 다른 인터페이스 요소를 함께 활용한 멀티모달 인터랙션의 효과를 검증한 다른 사용 맥락에서의 기존 연구 결과와는 달리, 자동차 주행 맥락의 이 연구에서는 멀티모달 인터랙션이 효과가 없는 것으로 나타났다. 이는 반자율주행 단계에서는 여전히 운전자의 인지적 노력이 필요하기 때문에 복합적 자극의 인터랙션이 오히려 방해 요소가 됨을 의미한다.

이러한 결과에 근거하여 현재 반자율주행 환경에서 대화형 음성인식 인터페이스 디자인 전략을 결정하는 데 실무적 의의는 다음과 같다. 차량 내 음성인식 인공지능 시스템을 장착할 때 디스플레이를 통해 에이전트의 시각적 피드백을 표시하는 것은 바람직하지 않으며, 음성 단일 모드가 더 효과적이다. 차량 내 음성인식 인터랙션 방식에 대한 유용성은 높게 인식하는 만큼, 최적의 운전자 경험 만족을 위해 음성 에이전트의 음성 품질과 운전 맥락에 특화된 메시지 설계에 더 초점을 두어야 한다.

이 연구의 한계점은 다음과 같다. 첫째, 선정된 음악에 대한 참가자의 선호도가 실험에 영향을 미쳤을 가능성이다. 자신이 선호하는 유형의 음악이 나온 참가자는 만족도와 정보, 시스템 품질 면에서 높은 평가를 주고, 그렇지 않은 참가자는 그와 반대로 평가할 수 있다. 실제로 인터뷰 중 한 참가자는 재생된 곡이 모두 자신의 취향에 맞아, 자신의 선호도가 반영되고 있는 줄 알았다고 응답했다. 둘째, 실험 대상의 대부분이 대학생으로 운전 경력이 적은 편이었다는 점이다. 운전 경력이 길어질수록 숙련도가 높아 운전 자체에 대한 인지 부담이 줄어들기 때문에 부가적 행동에 상대적으로 많은 인지적 노력을 들일 수 있다. 따라서 전체 운전자 모집단과 비교하면 다른 결과가 나올 수 있다. 셋째, 시각적 자극이 캐릭터였기에 캐릭터에 대한 남성과 여성의 선호도 차이가 캐릭터에 대한 의견에 영향을 미쳤을 가능성이 있다. 넷째, 시뮬레이터라는 실험 조건 한계 상 실제 차량에서의 운전 환경에서의 도로 상황이나 기상변화, 다른 차량 등 여러 변수를 구현할 수 없었다는 점이다. 실제 환경에서는 이와 다른 결과가 나올 수 있다.

미래에 완전 자율주행 기술이 상용화될 경우 운전 외 콘텐츠 이용과 관련된 사용자 경험의 중요성이 대두될 것이므로 완전 자율주행 환경에서의 인터랙션에 대한 후속 연구가 필요하다. 완전 자율 주행 환경에서는 운전자의 인지 부하가 더 낮아지기 때문에 시각적인 캐릭

터 요소를 방해 요소로 인식하지 않아 사용자 경험에 긍정적인 영향을 미칠 수도 있다. 또한, 자율주행 환경에서의 인터랙션뿐만 아니라 콘텐츠의 범위 확대에 대한 연구도 이루어져야 할 것이다. 사후 인터뷰의 “운전하면서 음악 서비스를 이렇게 이용할 수 있다는 게 재미있었다”, “완전 자율주행일 때 이런 서비스가 더 많으면 지루하지 않겠다” 등의 의견으로 볼 때 자율주행 환경 내 구축된 콘텐츠 자체가 사용자 경험에 긍정적인 영향을 미칠 수 있을 것으로 기대된다.

참 고 문 헌

[1] 김현근, 김지선, 지승도, “다중 에이전트를 이용한 캐릭터 기반 대화형 인공지능 설계,” 한국시뮬레이션학회 학술대회 논문집, pp.75-79, 2017.

[2] M. H. Cohen, J. P. Giangola, and J. Balogh, *Voice user interface design*, Addison-Wesley Professional, 2004.

[3] J. Cassell, “Embodied conversational interface agents,” Communications of the ACM, Vol.43, No.4, pp.70-78, 2000.

[4] S. Macskassy and S. Suzanne, *A conversational agent*, Rutgers University, 1996.

[5] M. Peissner, V. Doebler, and F. Metze, *Can voice interaction help reducing the level of distraction and prevent accidents? Meta-study on driver distraction and voice interaction,* Nuance Communications, 2011.

[6] M. Jeon, B. N. Walker, and T. M. Gable, “The effects of social interactions with in-vehicle agents on a driver′s anger level, driving performance, situation awareness, and perceived workload,” Applied ergonomics, Vol.50, pp.185-199, 2015.

[7] 이지인, 김나은, 김진우, “자율주행 자동차 환경에서의 운전자 경험에 대한 연구,” 한국디지털콘텐츠학회 논문지, 제18권, 제4호, pp.713-722, 2017.

[8] 이혜민, 김승인, “음성인식 기반의 모바일 지능형 개인비서 서비스 사용성 비교,” 디지털디자인학연구, 제14권, 제1호, pp.231-240, 2014.

[9] 홍은지, 조광수, 최준호, “스마트홈 대화형 인터페이스의 의인화 효과,” 한국 HCI 학회 논문지, 제12권, 제1호, pp.15-23, 2017.

[10] 박은영, 이경아, “음성인식을 적용한 멀티미디어 콘텐츠 방안-교육용 유아영어 앱을 위한 사용자 인터페이스를 중심으로,” 브랜드디자인학연구, 제11권, 제1호, pp.111-120, 2013.

[11] S. Okamoto and S. Sano, “Anthropomorphic AI Agent Mediated Multimodal Interactions in Vehicles,” Adjunct Proceedings of the 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp.110-114, 2017.

[12] 이동우, 최은정, 이용희, 박준석, “멀티모달 사용자 인터페이스 기술 동향,” 한국정보기술학회지, 제3권, 제1호, pp.15-20, 2005.

[13] L. Nigay and J. Cautaz, “A design space for multimodal systems: concurrent processing and data fusion,” Proceedings of the INTERACT′93 and CHI′93 conference on Human factors in computing systems, ACM, pp.172-178, 1993.

[14] 장복만, 박성도, 장효경, 장창복, 최의인, “멀티모달 인터랙션 에이전트를 이용한 u-헬스케어 상황인식 보안 프레임워크,” 보안공학연구논문지, 제6권, 제5호, pp.337-350, 2009.

[15] B. Dumas, D. Lalanne, and S. Oviatt, “Multimodal interfaces: A survey of principles, models and frameworks,” Human machine interaction, pp.3-26, 2009.

[16] S. Oviatt, “Multimodal interactive maps: Designing for human performance,” Human-computer interaction, Vol.12, No.1, pp.93-129, 1997.

[17] 최종우, 박혜선, 김경호, “운전자 정보제공을 위한 모달별 인터페이스,” 대한인간공학회 학술대

회논문집, pp.254-257, 2010.

[18] J. Cassel, "Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents," *Embodied Conversational Agents*, MIT Press: Cambridge, MA, pp.1-27, 2000.

[19] N. Yee, J. N. Bailenson, and K. Rickertsen, "A MetaAnalysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces," Proceedings of CHI San Jose, CA, USA, pp.1-10, 2007.

[20] J. Walker, L. Sproull, and R. Subramani, "Using a Human Face in an Interface," Proceedings of CHI, pp.85-91, 1994.

[21] 조윤주, 최유정, 박수이, "모바일 환경에서의 캐릭터 인텔리전트 에이전트에 대한 신뢰 형성 요인," 디지털디자인학연구, 제8권, 제1호, pp.83-98, 2008.

[22] R. Agarwal and E. Karahanna, "Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage," MIS quarterly, Vol.24, No.4, pp.665-694, 2000.

[23] L. Deng, D. E. Turner, R. Gehling, and B. Prince, "User experience, satisfaction, and continual usage intention of IT," European Journal of Information Systems, Vol.19, No.1, pp.60-75, 2010.

[24] 주영주, 김은경, 박수영, "기업 사이버교육에서 인지적 실재감과 몰입, 만족도, 지속의향과의 구조적 관계," 교육정보미디어연구, 제15권, 제3호, pp.21-38, 2009.

[25] 한상도, 장명복, "정보시스템 품질과 사용자 만족에 관한 연구," 품질혁신, 제2권, 제1호, pp.78-86, 2001.

[26] 김영환, 최수일, "지각된 서비스 품질, 유용성, 용이성이 IPTV 사용자 만족 및 지속적 사용의도에 미치는 영향," 한국콘텐츠학회논문지, 제9권, 제10호, pp.314-327, 2009.

[27] J. J. Choi and S. S. Kwak, "The Impact of Robot Appearance Types and Robot Behavior Types on People's Perception of a Robot," 디지털디자인학연구, 제14권, 제3호, pp.1-11, 2014.

[28] W. H. Delone and E. R. McLean, "The DeLone and McLean model of information systems success: a ten-year update," Journal of management information systems, Vol.19, No.4, pp.9-30, 2003.

[29] Z. Yang, S. Cai, Z. Zhou, and N. Zhou, "Development and validation of an instrument to measure user perceived service quality of information presenting web portals," Information & Management, Vol.42, No.4, pp.575-589, 2005.

[30] 김용일, 임재문, "확장된 기술수용 모델을 적용한 스마트폰 외식정보 서비스 사용자의 수용태도에 관한 연구," 호텔경영학연구, 제23권, 제5호, pp.275-289, 2014.

[31] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS quarterly, Vol.13, No.3, pp.319-340, 1989.

[32] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: a comparison of two theoretical models," Management science, Vol.35, No.8, pp.982-1003, 1989.

[33] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," Decision sciences, Vol.39, No.2, pp.273-315, 2008.

[34] 손승혜, 최윤정, 황하성, "기술수용모델을 이용한 초기 이용자들의 스마트폰 채택 행동 연구," 한국언론학보, 제55권, 제2호, pp.227-251, 2011.

[35] 남종훈, "DMB 의 수용결정요인에 관한 연구," 언론과학연구, 제7권, 제2호, pp.143-188, 2007.

[36] R. A. Sánchez and A. D. Hueros, "Motivational factors that influence the acceptance of Moodle

using TAM," Computers in human behavior, Vol.26, No.6, pp.1632-1640, 2010.
[37] 최종우, 박혜선, 김경호, "운전자 정보제공을 위한 모달별 인터페이스," 대한인간공학회 학술대회논문집, pp.254-257, 2010.
[38] 한국콘텐츠진흥원, *2016 캐릭터 산업백서,* 한국콘텐츠진흥원, 2016.

저 자 소 개

서 민 수(Min-soo Suh) 준회원

- 2016년 2월 : 연세대학교 컴퓨터과학(공학사)
- 2016년 2월 : 연세대학교 일본학(연계학사)
- 2016년 3월 ~ 현재 : 연세대학교 정보대학원 UX 전공

<관심분야> : 자율주행차량, VR, 콘텐츠 UX 등

홍 승 혜(Seung-Hye Hong) 준회원

- 2018년 2월 : 아주대학교 전자공학과(공학사)
- 2018년 3월 ~ 현재 : 연세대학교 정보대학원 UX 전공

<관심분야> : AI, 콘텐츠 UX

이 정 명(Jeong-Myeong Lee) 준회원

- 2016년 2월 : 동국대학교 전자전기공학, 디자인공학(공학사)
- 2018년 3월 ~ 현재 : 연세대학교 정보대학원 UX 전공

<관심분야> : UX리서치, Data-Driven UX