How Voice Interface Influences Users' Music Experience : An Exploratory Study Using YouTube Videos

Yuin Jeong[1], Juho Lee[2], Saenanseul Kim[2], Younah Kang[3*]

[1]Management of Technology, Student, Yonsei University, Seoul, Korea

[2]Design Intelligence Major, Student, Yonsei University, Seoul, Korea

[3]Underwood International College, Professor, Yonsei University, Seoul, Korea

Abstract

Background Voice-user interface(VUI), a new interface that enables people to have vocal communication with the system, is recently being commercialized. However, the affects of VUI on users' daily lives and behavioral patterns have not been fully explored.

Methods To identify user needs emerged from the new interface, especially focusing on music service, which is a key service domain of voice interface, we observe 25 AI speaker users' video clips on YouTube by applying a digital ethnography method.

Results From the results of the observation, we classified 29 most common functions of music service found in voice interface. We also found that users, while using VUI, unify multi-functions into a single sentence to activate music retrieval service, struggle to operate music-player options through spoken language, and use melody to communicate with the system.

Conclusions Our study indicates that there is a strong need for service/functions expansion in VUI, which can go beyond the functions of graphic-user interface, and maximizes the strengths of voice-based conversational interface. With these findings, we discuss design implications for music service conveyed via VUI.

Keywords Voice User Interface, AI Speaker, Music Experience, Qualitative Study, Digital Ethnography

This research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (N0001436, The Competency Development Program for Industry Specialist).

*Corresponding author: Younah Kang (kang.younah@gmail.com)

Citation: Jeong, Y., Lee, J., Kim, S., & Kang, Y. (2020). How Voice Interface Influences Users' Music Experience : An Exploratory Study Using YouTube Videos. *Archives of Design Research, 33*(1), 165-177.

http://dx.doi.org/10.15187/adr.2020.02.33.1.165

Received : Jun. 14. 2019 **; Reviewed :** Dec. 02. 2019 ; **Accepted :** Dec. 06. 2019
pISSN 1226-8046 **eISSN** 2288-2987

1. 연구의 배경 및 목적

최근 상용화된 음성-사용자 인터페이스(Voice-User Interface, 이하 VUI)에서의 음악 서비스는 핵심 사업 도메인으로 부상하고 있다. 음악 서비스는 인공지능 스피커(이하, AI스피커)에서 시간당 평균 사용빈도가 가장 높은 서비스로, 88명의 구글 홈(Google Home) 사용자 조사 결과, 입력된 명령어 중 40% 이상이 음악 서비스 관련 기능으로 나타났다(Bentley et al., 2018).

그러나 현재 상용화된 VUI기기로 전달되는 음악 서비스의 기능과 그 인터랙션 방식은 새로운 인터페이스에 대한 사용자의 기대를 만족시키기에는 부족한 수준이다. 기존 그래픽 인터페이스(GUI)에서의 터치와 텍스트로 구동되던 기능이 음성발화에 반응하는 형태로 변화했을 뿐, VUI에서 제공되는 서비스는 음악을 켜고 끄는 등의 단조로운 형태에 머물고 있다(Bentley et al., 2018). 이와 같이 제한된 기능은 새로운 인터랙션으로 나타나는 사용자의 니즈(Needs)를 충족시키지 못하고, 기대와 현실의 격차로 인한 부정적인 사용 경험으로 이어질 가능성이 높다(Luger & Sellen, 2016).

따라서 음성 인터페이스에서의 음악 서비스 사용자 만족도를 증가시키기 위해서는, VUI의 장점을 극대화할 수 있는 방향으로 차별화된 새로운 서비스 및 인터랙션이 제공되어야 한다. 또한 음악은 '소리'와 밀접한 관계를 맺고 있는 서비스라는 측면에서 VUI에서 새로운 형태의 사용자 경험이 창출될 것으로 기대되는 분야이다. 이에 실사용자를 중심으로 인터페이스의 전환에 따라 사용자 경험에 어떠한 변화가 일어나고 있는지를 분석하는 연구 필요성이 증가하고 있다.

그러나 AI스피커에 대한 높은 관심과 관련 사용자 경험 연구 필요성에도 불구하고, 개인 공간에서 사용되는 기기 특성상 AI스피커 사용 경험을 직접적으로 관찰함에는 프라이버시 침해 가능성과 고비용소요라는 현실적인 제약이 존재한다. 또한 크랩트리와 로든(2005)에 따르면 집안 환경에서 이뤄지는 직접 관찰연구(Direct Oberservation)는 연구자의 시선에 영향을 받아 사용자의 일상적인 경험이 왜곡될 가능성이 매우 높다. 이와 같은 직접 관찰연구의 한계는 관련 연구 진행에 장애물로 여겨지고 있으며, 현재 실제 데이터를 기반으로 한 AI스피커의 사용 경험 조사 연구는 그 필요성에 비해 미비한 수준이다.

이에 본 연구는 인터페이스 전환에 따른 음악 서비스 사용자의 니즈를 확인하고 디자인 함의점을 도출함에 그 목적을 두고, 상용화된 VUI인 AI 스피커 사용자를 중심으로 음악 서비스의 사용 맥락과 인터랙션을 관찰하고자 하였다. 또한 그 연구방법으로는 전통적인 관찰연구의 대안적 방식으로 주목받고 있는 '유튜브 활용 디지털 에스노그라피' 방식을 적용하여 AI스피커 사용자의 실제 사용 경험 분석을 위한 탐색적 연구를 진행하였다. '유튜브 활용 디지털 에스노그라피'는 유튜브에 사용자가 업로드한 비디오, 오디오 데이터를 기반으로(Paay, Kjeldskov, Skov, & O'hara, 2013), 디지털 미디어가 어떻게 사용자의 일상에 편입되고 있는지를 관찰 분석하는 방법을 의미한다(Pink, 2016). 특히 최근에는 집안 환경 등 연구자의 접근이 어려운 공간에서의 사용자의 행동과 경험을 분석하는 1차 연구 방식으로 활용되고 있다(Paay, Kjeldskov, Skov, & O'hara, 2012). 본 연구는 연구 진행을 위해 총 25개의 사용자 경험이 담긴 유튜브 영상을 수집하였으며 이를 중심으로 디지털 에스노그라피 연구를 실시하였다.

연구목적에 따른 주요 연구 질문들은 다음과 같다.
RQ1. 음성-사용자 인터페이스는 사용자의 음악 소비행태에 어떠한 영향을 가져왔는가?
RQ2. 음성인터페이스에서 사용자들은 어떠한 음악 서비스 기능을 사용하는가?
RQ3. 사용자는 어떠한 사용 맥락에서, 어떠한 명령어를 사용하여 음성인터페이스로 음악 서비스를 구동시키는가?
RQ4. 인터페이스의 변화에 따라 음악 서비스에 적용될 수 있는 디자인 함의점은 무엇인가?

2. 문헌연구

2. 1. 음성-사용자 인터페이스 (Voice User Interface)

음성-사용자 인터페이스(이하 VUI)는 사용자와 구두의 언어로 소통하는 상호작용 방식을 지닌 응용프로그램을 의미한다(Porcheron, Fischer, Reeves, & Sharples, 2018). 과거 자동응답서비스 등 제한된 서비스 분야에서 활용되던 VUI는 최근 인공지능 기술의 발전으로 자연어 처리가 가능해짐에 따라 그 영역을 확장해 가고 있다. 2011년 애플에서 출시한 시리를 시작으로 구글 어시스턴트 등 모바일에 탑재된 가상비서 형식의 VUI가 상용화되었으며(Hoy, 2018), 2014년 가정용 인공지능 스피커 아마존 에코의 등장으로 VUI가 일상생활의 영역 전반에 깊이 자리 잡게 되었다. 2017년 기준 약 3억 5천만 명 이상의 사용자가 인공지능 스피커를 사용하고 있으며(Sciuto, Saini, Forlizzi, & Hong, 2018), 국내에서도 SK텔레콤의 누구, 네이버 클로바와 카카오 미니 등의 음성 인터페이스를 탑재한 AI스피커가 출시되어 사용자들에게 날씨 알림, 음악감상, 쇼핑 등의 다양한 일상 서비스를 제공하고 있다.

이처럼 음성형 인터페이스가 주목받고 있는 이유는 전통적인 그래픽 기반 인터페이스와 차별화된 음성 인터페이스만의 장점 때문이다. 선행연구(Cohen, Cohen, Giangola, & Balogh, 2004)에 따르면 음성 인터페이스는 인간에게 가장 자연스럽고 친숙한 소통방식인 대화를 기반으로 서비스를 제공하기 때문에, 인터페이스 사용을 위한 학습시간이 짧으며, 고령자 혹은 어린이 등 기기사용에 익숙하지 않은 사용자도 쉽게 이용할 수 있다는 장점을 지닌다. 또한 두 손과 눈을 사용하지 않아도 되기 때문에 멀티태스킹(Multi-tasking)환경에서 기기를 조작할 수 있는 가능성을 제공한다.

그러나 음성 인터랙션 방식은 기기 구동을 위해 사전에 훈련된 방식이 아닌 자연어의 대화를 기반으로 이뤄지기 때문에 인터랙션 방식이 다변적이며 예측이 어렵다는 한계를 가지고 있다. 코헨 등의 연구(Cohen et al., 2004)에 따르면 인간의 대화는 의사소통 과정에서 사용자가 사용하는 단어선택, 발음, 문장구조 등이 다변적이라는 점에서 기술이 처리하기 어려운 다수의 모호성(Ambiguity)을 지닌다. 또한 대화 속에 내포된 함축어와 사용 맥락, 사회적 의미 등의 상식(Common sense)을 이해하고 이에 적절한 반응을 하는 것은 여전히 인공지능 기술이 풀기 어려운 난제로 남아있다(Weston, Bordes, Chopra, Rush, Joulin, & Mikolov, 2015). 따라서 사용자의 발화 의도에 대한 최선의 대화 전략을 구사하는 형태로 시스템의 수준을 향상시키기 위해서는 실사용자의 VUI 사용 맥락과 인터랙션 방식에 대한 분석이 선행되어야 한다.

2. 2. 인터페이스의 변화에 따른 VUI 사용자 경험 조사연구

VUI에서 사용자가 시스템과 상호작용하는 방식은 인간-인간 간의 소통 방식과 차이가 있으며(Shechtman & Horowitz, 2003) 기존에 사용자들이 시스템과 인터랙션하던 방식과도 다를 수 있다(Myers, Furqan, Nebolsky, Caro, & Zhu, 2018). 일례로 카랏, 할버슨, 혼 그리고 카랏(Karat, Halverson, Horn, & Karat, 1999)의 연구에서는 오랜 기간 시스템과 시각적 인터랙션 방식으로 상호작용해온 사용자들이 새로운 인터페이스인 VUI를 사용하는 과정에서 어색함을 느끼는 것으로 나타났다. 또한 루거와 셀렌(Lugar & Sellen, 2016)의 연구에 따르면 사용자들은 일회성의 정보 검색 수준에 머무르는 단조로운 에이전트의 기능과 인터랙션 방식에 괴리감(Gulf)을 느끼며, 인터페이스의 '대화' 능력이 충분히 전달될 수 있는 새로운 인터랙션 방식에 대한 수요가 증가하고 있는 것으로 나타났다.

이에 HCI분야에서는 그 필요성에 따라 실제 인터랙션 과정에서 사용된 명령어를 중심으로 VUI와 사용자 간의 커뮤니케이션 방식을 분석한 연구들이 주목을 받고 있다. 예를 들어 스키토, 사이니, 포리치, 그리고 홍(Sciuto, Saini, Forlizzi, & Hong, 2018)은 아마존 에코를 이용할 때 사용되었던 음성 호출어 로그데이터를 수집하여 VUI가 어떻게 일상생활에 편입되고 있는지를 분석하였다. 또한 포체론 등은 연구(Porcheron, et al., 2018)에서 실제 집안환경에 VUI 구동을 위한 호출어(Alexa 등)에 반응하는 조건부 녹음기를 설치하고, 대화분석(Conversational Analysis)방법론을 적용하여 사용자와 시스템과의 인터랙션 방식을 미시적으로 분석하였다.

그러나 실사용자의 발화 데이터를 기반으로 인터랙션 과정을 분석한 연구는 데이터 확보에 상당한 시간과 비용이 소요된다는 점에서 아직 미비한 수준이다. 또한 음악 서비스라는 특정한 서비스 분야에 초점을 맞추어 인터페이스의 전환에 따른 사용자의 사용행태를 분석한 연구는 전무하다는 점에서 관련 연구의 필요성이 증가하고 있다.

3. 연구방법

연구목적에 따라 본 연구는 VUI 사용자들이 유튜브에 업로드한 AI스피커 사용 영상 클립을 중심으로, 질적 연구방법 중 하나인 관찰연구를 진행하였다. 블라이스와 케어른스의 연구(Blyth & Cairns, 2009)에 따르면, 유튜브는 최근 디지털 에스노그라피 분야에서 인간의 행동과 상황을 관찰할 수 있는 새로운 형식의 데이터로 주목받고 있으며, 기타 연구방법에서 얻기 어려운 사용자의 사적인 행동 및 공간에 대한 시청각적 정보를 습득할 수 있다는 이점을 가진다. 또한 사용자의 행동이 연구자의 시선에서 자유롭고, 자발적이고 적극적인 "Think Aloud" 데이터 확보가 가능하다는 점에서 최근 사용자 경험 조사연구 분야에서 주요한 관찰 데이터로 활용되고 있다(Paay et al., 2013).

3. 1. 연구대상

실제 사용 경험을 분석하기 위한 목적에 따라 본 연구는 상용화된 VUI인 국내 AI스피커 사용자(네이버 클로바, 카카오 미니, 구글 어시스턴트 등)의 유튜브 영상물을 연구 대상으로 선정하였다. 기술의 발전 속도와 시기성을 고려하여 2018년 2월 2019년 7월 내에 업로드된 영상으로 대상을 한정하였으며, 분석목적에 따라 단순한 사용자 리뷰 형태의 영상이 아닌 사용자와 AI스피커와의 '실제 인터랙션' 과정이 1회 이상 포함된 유튜브 클립으로 연구 대상을 한정하였다. 또한 음악 서비스 이용 경험이 포함되지 않은 영상물은 분석에서 제외하였다.

분석대상은 질적 연구에서 주로 활용되는 유목적표집법(Purposeful Sampling)을 기반으로(Palinkas, Horwitz, Green, Wisdom, Duan, & Hoagwood, 2015)선정되었다. 대상 선정을 위해 유튜브에서 〈표1〉의 키워드로 관련 동영상 자료를 검색하여 자료를 확보하였으며 이후 분석목적에 따라 스크리닝 과정을 거쳤다.

Table 1 Keywords Used for Data Collection

구분	키워드 (유사어포함)
AI스피커 전반	인공지능 스피커, AI스피커, 스마트스피커
브랜드명	네이버 클로바(Clova, 클로바, 네이버 인공지능 스피커, 브라운, 샐리)
	카카오 미니(카카오 인공지능 스피커)
	구글 어시스턴트 (오케이구글, Google Assistant)
	SK NUGU(누구, SKT 인공지능 스피커, 누구캔들)
	KT 기가지니(지니야, 기가지니)

검색어로 노출된 총 227개의 영상 중 연구목적에 따라 25개의 영상 클립이 최종 분석 대상으로 선정되었다. 선정된 영상물의 길이는 평균 6.42분, 총 160.51분이었으며, 영상에 노출된 참가자 수는 총 35명으로, 한 명의 개인이 영상에 등장하여 스피커와 인터랙션하는 경우(N=17)와 가족 및 지인 등의 다수의 사용자가 스피커와 인터랙션하는 사용자그룹의 영상으로(N=8)구분되었다. 또한 영상에서 사용자들이 사용한 기기는 네이버클로바(N=11), 구글어시스턴트(N=7), 카카오미니(N=8), SKT누구(N=5), KT기가지니(N=1)로 나타났으며, 한 명의 사용자가 한 개 이상의 기기를 사용하는 사례가 포함되었다.

3. 2. 분석방법

수집된 영상물은 관찰연구의 분석 방법 중 AEIOU 프레임워크(Wasson, 2000)를 통해 활동(Activity), 환경(Evaluation), 상호작용(Interaction), 사물(Object), 사용자(User)를 중심으로 분석되었다. 분석을 위하여 사용자가 에이전트와 인터랙션하는 과정에서 사용된 1)명령어 2)서비스에 대한 사용자의 반응인 Think Aloud 데이터 자료들이 전사 과정을 거쳐 기초 자료(Raw Data)형태로 수집되었다. 사용자의 잠재된 니즈를 분석하기 위한 연구목적에 따라 기기의 반응 오류 유무와 관계없이 사용자가 인터페이스에 사용한 명령어를 모두 수집하여, 총 84개의 음악 서비스 관련 명령어가 수집되었다.

수집된 기초 데이터 자료는 질적연구방법인 반복적 비교분석법(Constant Comparative Method)(Glaser, 1965)을 기반으로 분석되었다. 기초 데이터로부터 개념들을 범주화하여 분석하는 반복적 비교분석법의 과정으로 '개방코딩(open coding)', '범주화', '범주확인' 작업을 시행하였다. 또한 삼각검증법(Triangulation)(Morse, 1991)을 기반으로 총 3명의 연구자가 다수(Multiple)의 데이터를 분석하고 다양한 이론을 적용하여 질적연구의 신뢰성과 타당성을 확보하고자 했다.

4. 연구결과

분석 결과, VUI를 활용하여 이루어지는 사용자의 음악 관련 활동은 〈표3〉과 같이 5개의 영역(음악재생, 음악재생관리, 음악추천, 음악검색, 노래부르기)의 총 29개의 하위기능들로 나타났다. 위 영역 중 본 연구목적에 따라 음성인터페이스에서 새롭게 혹은 빈번하게 관측되는 사용자 행동 및 인터랙션 방식을 중심으로 분석 및 서술하였다.

Table 2 29 Music Related Service Features Activated by Users

영역 (5)	하위기능 (29)	명령어 예시
음악재생 (10)	제목으로 재생	라흐마니노프 피아노 협주곡 2번 틀어줘
	가수로 재생	알리 노래 틀어줘 걸그룹 노래 틀어줘
	악기명으로 재생	피아노 곡 틀어줘, 리코더 노래 틀어줘
	가사로 재생	아름다워 사랑스러워 그래 너! (음을 흥얼거리며)
	특정 상황에 적합한 곡 재생	조용한 노래 틀어줘,
	특정 장소와 관련된 곡 재생	카페 노래 틀어줘. 클럽노래 틀어줘. 스타벅스노래 틀어줘
	장르명으로 재생(클래식, 힙합 등)	클래식 틀어줘. 찬송가 틀어줘
	플레이리스트 저장 곡 재생	내 플레이리스트에서 JAZZ 틀어줘
	특정 소리를 재생	소쩍새 소리 들려줘. ASMR 틀어줘
	기타 서비스와의 연동	모모랜드 뿜뿜으로 알람 설정해줘
	상위 기능들의 조합 재생	윤종신의 너를 찾아서 틀어줘 (가수 + 제목)
음악재생 관리 (9)	부분 듣기	하이라이트 부분만 틀어줘
	일시정지 (Pause)	잠깐만
	빨리 감기	빨리 넘겨줘
	간주점프	간주점프해줘
	1절만 재생	1절만 틀어줘
	반복 재생	방금 그 노래 또 들려줘
	다른 음악으로 변경 (Next)	다른 노래 들려달라고
	볼륨/소리크기	소리 완전 줄여줘
	음악 정지(Stop)	음악 꺼줘, 스탑, 멈춰 등

음악추천 (4)	개인 기호에 따른 추천	내가 좋아할 만한 노래 틀어줘
	날씨 관련 추천	비 오는 날 듣기 좋은 노래 틀어줘
	계절 관련 추천	겨울에 어울리는 노래 틀어줘
	기분 관련 추천	위로가 되는 노래 좀 틀어줄래? / 우울한 노래 틀어줘
음악검색 (4)	신곡 검색	임창정 신곡 틀어줘
	영화나 TV콘텐츠에 포함된 수록곡 검색	손예진 나오는 멜로 영화에 나오는 노래가 뭐지
	현재 재생 중인 곡을 검색	야 이거 누구 노래니?
	특정 멜로디로 검색	음음~음~(노래를 흥얼거리며) 이 노래 뭐야
노래부르기 (2)	에이전트와 함께 부르기	팅커벨. I say 팅 you say 커벨. 팅.
	에이전트에 요청하기	클로바 뮤지컬 해줘 /노래해줘 / 랩해줘

4. 1. 인터랙션 과정의 단축 및 기능 간의 통합

연구결과, 음성대화의 인터랙션 방식은 음악 서비스를 사용하는 과정에서 사용자의 인터랙션 단계(Depth)를 단축시키는 것으로 나타났다. 기존 GUI에서 특정 음악을 재생시키기 위해 어플리케이션 실행하는 등의 여러 단계의 인터랙션을 거쳐야했던 것과 달리, VUI에서 사용자는 “A노래 틀어줘”라는 한 마디의 인터랙션으로 A라는 음악을 검색하고 이를 재생시킨다. 또한 터치 인터페이스에서와는 달리 사용자는 다수의 기능을 ‘한 마디의' 문장으로 통합시켜 시스템에 명령을 내린다. 사용자의 발화에서 가장 빈번하게 등장한 동사는 음악재생을 위한 명령어인 “틀어줘”와 “들려줘”(33%)로, 주로 음악검색과 음악추천 등의 관련 기능들과 혼합된 형태로 나타났다. 그중에서도 “알리 노래 틀어줘" 와 같은 ‘곡명/가수명 검색 + 재생기능'의 두 가지 기능이 통합된 명령형 문장이 가장 빈번하게 나타났으며, 세 가지의 기능이 한 마디의 문장으로 실행되는 사례도 나타났다.

“플레이리스트에서(플레이리스트관리) + Jazz노래(음악검색) + 틀어줘(음악재생)”

“임창정(가수검색) + 신곡(곡명검색) + 틀어줘(음악재생)”

또한 영상 관찰 결과 사용자는 이처럼 단축된 인터랙션 과정에 높은 만족감을 느끼는 것으로 나타났다.

“타이핑으로 검색하는 게 아니라 음악을 바로바로 틀어주니까요. 개인적으로 타이핑으로 음악검색을 거의 하지 않게 되었습니다.” -P10

한편 VUI에서 사용자들은 음악과 관련되지 않은 타 분야의 기능을 음악 서비스 기능들과 접목시켜 사용하고자 시도하였다. 일례로 참가자(P7)는 자신이 듣고자 하는 음악으로 ‘알람’을 설정하고자 하였으며, 또 다른 참가자(P19)는 침실에서 음악을 재생하여 오늘의 날씨를 유추하고자 시도하였다. 이는 음악이 사용자의 일상생활에서 중요한 역할을 하며, 다양한 기능 영역들과 접목되어 사용되는 영역임을 보여주는 사례이다.

“카카오, 모모랜드 뿜뿜으로 알람 설정해줘" - P7

“오늘 날씨에 어울리는 노래 틀어줘" -P4

4. 2. ‘음악재생관리’기능 사용의 어려움

일시정지(Pause)와 중지(Stop) 등의 ‘음악재생관리’ 기능은 GUI에서는 빈번하게 사용되는 인터랙션 요소인 반면, VUI에서 제대로 구현되지 않거나 사용자가 관련 기능을 실행시키기 위한 명령어를 사용하는 과정에서 어려움을 겪는 영역으로 나타났다. 특히 AI스피커를 처음 사용하는 것이라고 밝힌 초보자(Beginner) 그룹의 영상(P1, P5, P6, P17)에서는 사용자들이 음성으로 재생 중인 음악을 ‘중지’시키기 어려워하는 현상이 빈번하게 관찰되었다. 음악중지를 위해 사용자가 사용한 명령어는 “꺼줘”, “멈춰”, “그만”, “스탑”, “노래 꺼”, “종료해줘”, “다른 노래” 등 총 9가지로 나타났다. 또한 관련 기능실행을 위해 최소 2회 이상의 명령어가 시도되었으며, 한 번에 음악이 멈춰지지 않는 경우 기기를 두드리거나, 표정을 찡그리거나, 소리치는 등 사용자의 강한 좌절감(User Frustration)이 관찰되었다.

“음악을 끄는 방법을 아직 잘 몰라요.(기기를 툭툭 치며) 오케이 구글 그만! 그만해!”-P1

“아 이거[노래] 어떻게 끄지? (기기를 만지며) 아휴 껐다 켜야겠다”-P5

한편 음악재생관리 영역에서 사용자들은 다른 장소에서 활용되는 음악 소비 방식을 적용하는 등 VUI에서 보다 적극적으로 음악을 조정하고 관리하고자 하는 시도들이 나타났다. 예를 들어 음악이 재생되는 과정에서 P17은 “간주점프”, “1절만”, “하이라이트 들려줘" 등 노래방에서 빈번하게 활용되는 기능을 AI스피커에서도 시행하고

자 하는 행동들이 관찰되었다. 그러나 해당 명령어들은 아직 AI스피커에서 구현이 되지 않은 기능들로, 이때 사용자들은 관련 기능을 실행하기 위해 모바일 어플리케이션 사용을 병행하는 것으로 나타났다. 예를 들어 P3은 곡의 후렴구를 감상하기 위해 AI스피커에 음성으로 명령을 내리기보다는 핸드폰을 꺼내어 터치 인터페이스를 통해 곡을 조정하고자 했다.

4. 3. '음악추천' 기능에서의 사용자의 기대격차

사용자들은 VUI에서 주로 취향, 날씨, 계절, 기분에 따른 추천 명령을 시행했으며 '여자친구랑 싸웠어, 위로가 되는 노래 좀 틀어줄래' 등의 명령어를 기반으로 VUI에 감정적 지지를 동반한 음악추천을 받고자 하는 것으로 나타났다.

"황사 오는 날..미세먼지 오는 날 듣는 노래도 되나? 황사 오는 날 노래 틀어줘"-P10

"나 여자친구랑 싸웠어, 위로가 되는 노래 좀 틀어줄래?"-P11

이와 같은 음악추천 기능은 사용자의 맥락과 요구를 적절하게 고려하여 맞춤화된 음악을 검색/재생하는 서비스로 사용자를 만족시켜야 한다. 그러나 현재 상용화된 AI스피커의 음악추천기능은 사용자의 기대와 현 기술 수준 간의 격차가 가장 큰 기능 영역으로 나타났다. 기능이 사용된 순간의 전후 과정에서 나타나는 사용자의 표정과 발화내용을 집중적으로 관찰한 결과, 추천기능에 대해 사용자가 부정적인 감정을 표출하는 경우가 빈번하게 관측되었다. 또한 그 원인은 1)스피커의 추천곡이 사용자의 취향에 적합하지 않거나, 2)스피커의 곡 추천 프로세스를 사용자가 알 수 없기 때문으로 나타났다. 일례로 P7는 VUI로 추천기능을 실행한 후 기기에서 흘러나오는 노래를 듣고, 찡그리는 표정과 함께 "방금처럼 취향에 맞지 않는 노래가 나오면 별로 안 쓰겠죠"라고 말하며 VUI의 추천곡이 자신의 취향과 맞지 않다는 점을 강조하였다. 또한 사용자(P4)는 음악추천기능을 실행시킨 후 "이 노래가 도대체 왜 나오는지 알고 싶다"라며 추천 프로세스에 대한 정보를 얻고자 했다.

4. 4. '음악을 듣고 부를 수 있는 존재로 인식되는 VUI

VUI에서 사용된 다섯 가지 기능 분류 영역들 중 '음악재생', '음악재생관리', '음악추천', '음악검색'의 영역은 기존 그래픽 인터페이스(GUI)에서 제공되고 있는 스트리밍 서비스 기능들과 비슷하다. 그러나 시스템에게 노래를 불러 달라고 요청하거나, 사용자와 노래를 주고받고자 시도하는 '노래 부르기' 활동은 음성 인터페이스 사용과정에서 새롭게 관찰된 사용자 행동이다.

"팅커벨, I say (팅), You say (커벨) 팅"-P21

"카카오, 뮤지컬해줘"(노래부르기)-P2

또한 음악 검색 및 재생기능을 사용하는 과정에서 사용자가 노래를 부르거나 멜로디를 흥얼거리는 현상이 관찰되었다. 특정 가사가 나오는 노래를 검색하기 위해 사용자들은 "~라는 가사가 나오는 노래를 틀어줘/검색해줘" 라고 말하는 대신 VUI에게 직접 노래를 부르는 방식을 시도하였다. 이때에는 '틀어줘'라는 요청어가 문장에서 생략되는 사례도 있었으며, 문장이 길어지기 때문에 사용자가 명령어를 끝내기 전에 기기가 반응하여 오류로 이어지는 경우가 빈번하게 나타났다.

"오케이 구글, 자유롭게~ 저 하늘로~(노래 부르며)"-P16

"클로바, 아름다워- 사랑스러워 (노래 부르며)"[사용자] / "그렇게 말씀해주시니 너무 기뻐요"[스피커/오류 발생]-P11

5. VUI에서의 음악 서비스 디자인 함의점

관찰연구 결과를 기반으로, 본 연구는 VUI에서의 음악 서비스를 디자인하는 과정에서 고려해야 할 4가지의 디자인 함의점을 제안한다.

5. 1. GUI와 혼용이 가능한 멀티모달(Multimodal) 방식으로의 음악 서비스 제공

연구결과, VUI는 음악 서비스 사용자들에게 GUI에서의 인터랙션과는 차별화되는 새로운 인터랙션 경험을 제공하는 것으로 나타났다. 인공지능 에이전트가 음악으로 자신의 감정을 위로해주기를 원하고, 에이전트가 멜로디를 이해한다고 가정하고 가사 대신 허밍이나 노래를 직접 부르는 방식으로 특정 곡을 검색하는 인터랙션은 기존 GUI에서의 사용자 인터랙션과는 다른 방식이다. 반면, GUI에서 빈번하게 사용되나 VUI의 명령어에서는 관측되지 않는 서비스 영역들도 확인되었다. 예를 들어 앨범표지로 음악을 선택하거나, 가사를 확인하는 등의 음악 소비행동은 수집된 명령어에서는 확인되지 않았다. 이는 음악이 비록 '소리'와 직접적으로 관계한 분야이지만, 모든 음악 서비스의 세부 기능들이 음성 인터랙션 방식으로 대체될 수는 없으며 여전히 시각적 인터페이스도 음악 서비스 경험을 지원하는 중요한 영역임을 시사한다.

이에 본 연구는 음악 서비스에서 음성과 그래픽의 영역이 활용될 수 있는 영역의 범주를 구분하고, 두 가지를 혼합하여 사용하는 형태인 '멀티모달 인터랙션'이 인공지능 스피커의 음악 서비스에 적합한 형태임을 강조하고자 한다. 현재 상용화된 대부분의 AI스피커가 스크린이 없는 형태의 음성 단일의 인터랙션 방식을 채택하고 있다는 점에서 사용자의 오류를 최소화하고 잠재 니즈를 충족시키기 위해서는 멀티모달 인터랙션 방식이 보다 적극적으로 활용될 필요성이 있기 때문이다.
따라서 본 연구는 구체적인 VUI 디자인 가이드라인의 한 예시로 구글(Google Developers, 2018)의 멀티모달 스펙트럼(Multmodal Spectrum)에 근거하여 각 인터페이스에서 활용이 유리한 음악 서비스 기능들을 분류 제시한다. 멀티모달 스펙트럼에 따르면 모든 서비스 기능들은 음성 유일(Voice only), 시각 유일의 형태(Visual only), 음성을 우선 사용하되 시각자료의 병행이 가능한(Voice forward) 형태, 음성과 시각이 모두 사용이 가능한 인터모달(Intermodal)의 형태로 구분될 수 있다. 본 연구는 관찰 연구결과에 근거하여 VUI에서 새롭게 나타난 기능을 음성 유일의 영역으로 분류하였으며, 또한 GUI에서 사용되는 기능이나 음성 인터페이스의 사용 경험에서 도출되지 않은 기능 영역은 시각 유일의 기능 영역으로 Figure 1.과 같이 분류하였다.

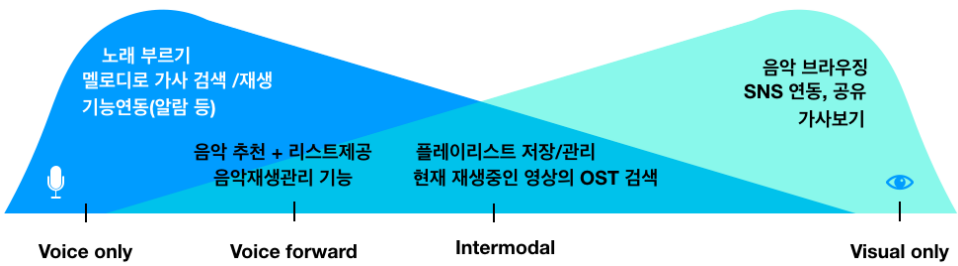

Figure 1 Classifications of Music Service Functions with Multimodal Spectrum

또한 시각과 음성의 두 가지 인터페이스를 혼합 사용했을 때 사용자 편익이 극대화될 것이라 기대되는 기능들은 '음성우선' 혹은 '인터모달'의 영역으로 분류되었다. 예를 들어 현재 VUI에서의 '음악추천' 기능은 사용자가 그 추천 프로세스를 알지 못한다는 점에서 만족도가 낮으며 장기 사용으로 이어지지 못한 측면이 있다. 이를 개선하기 위한 한 방법으로 사용자의 추천 명령어에 즉각적으로 추천곡을 재생하는 기존 방식보다는 몇 가지의 추천곡이 반영된 리스트 형식의 시각정보를 스크린 화면에 제시하고 곡 선택권을 사용자에게 위임하는 형태로의 '음성 우선'에 기반한 서비스 제공 방식이 고려될 수 있다.

5. 2. 기능 안내에서 나아가 언어적 학습과정이 포함된 '명령어' 튜토리얼 제공

VUI는 GUI와 달리 서비스를 사용하기 위한 별도의 학습이 필요하지 않은 가장 자연스러운 인터페이스로 평가되어왔다(Cohen et al., 2004). 그러나 본 연구결과에 따르면 사용자들은 오랜 기간 터치, 클릭 등의 GUI에서의 터치 인터랙션 방식에 익숙해져 있기 때문에 비슷한 기능을 VUI로 실행시키는 과정에서 명령어 선택에 어려움을 겪고 있는 것으로 나타났다. 일례로 '선택(Tap)', '넘기기(Swipe)'와 같은 그래픽/터치 인터페이스에서는

빈번하게 사용되나 동사형의 명령어로는 직관적인 표현방식이 어려운 기능이나, “Stop / Pause the music”와 같이 외국어와 터치 인터페이스에서는 명확히 구별이 가능한 기능이나 한글에서는 “멈춰”라는 한 가지 단어로 두 가지 기능을 모두 표현할 수 있는 경우 등에서 사용자가 명령어 선택에 인지 부하를 경험하는 현상이 관찰되었다. 이는 선행연구(Karat et al., 1999)와 일치하는 결과로, 반복적인 기기 오류로 이어져 사용자의 좌절감을 야기할 수 있다는 점에서 개선이 필요하다.

이에 본 연구는 음성 인터페이스 사용자에게도 그래픽 인터페이스 사용자와 마찬가지로 기기 사용을 위한 ‘훈련과정’이 필요함을 강조하고자 한다. 또한 그 훈련 방식으로 현재 상용화된 AI스피커에서 제공하고 있는 “외출 전 오늘 날씨를 체크해 보세요”, “제가 추천하는 기능을 듣고, 네/그래 또는 아니/아니요라고 말해보세요”와 같은 기능 안내 중심의 매뉴얼에서 나아가, 외국인이 소통을 위해 ‘말을 배우듯’ 에이전트와 사용자 간의 직접적인 인터랙션을 통해 ‘명령어’를 상호 학습하는 언어 학습 형태를 제안한다. 예를 들어 특정 기능을 실행시키기 위해 “음악 꺼줘, 멈춰, 그만” 등 9가지 이상의 명령어가 모두 사용되어 혼란을 야기하는 현상을 사전 방지하기 위해, “재생 중인 음악을 멈추기 위해서는 ‘그만’이라고 말해주세요” 혹은 “재생 중인 음악을 멈추기 위해서는 어떤 명령어를 사용하시겠어요?”와 같은 구체적인 ‘언어적 가이드’가 제공된다면 사용자와 에이전트 간의 인터랙션 빈도를 높이고 오류 가능성을 최소화할 수 있을 것이다.

5. 3. 음악적 커뮤니케이션이 가능한 대상으로의 VUI 기능 영역의 확장

음악은 작곡가, 연주자, 감상자 사이에서 메시지와 감정을 전달하는 ‘절대적인 언어’(Ultimate Language)로 기능한다(Baker, Paddison, & Scruton, 2001). MP3, 핸드폰 등의 인터페이스에서 음악을 전달하는 하나의 채널을 담당했던 시스템은 VUI의 등장으로 그 역할이 커뮤니케이션이 가능한, 음악적 표현(Musical expression)을 주고받을 수 있는 존재로까지 확장되고 있다. 관찰 결과에 따르면, 사용자들은 VUI에게 음악을 재생시키는 방식으로 감정적인 지지를 요구하거나, 에이전트와 함께 노래를 부르는 상호작용 방식을 시도하는 것으로 나타났다. 이는 사용자에게 VUI는 운율, 가사가 가진 정서적 메시지를 이해하고 음악으로 커뮤니케이션이 가능한 대상으로 인식되고 있음이 드러난 예라고 볼 수 있다.

따라서 VUI에서의 음악 서비스는 사용자의 명령어에 ‘멜로디’가 포함될 경우 이와 같은 유스 케이스를 민감하게 처리해야 한다. 일례로 상용화된 허밍이나 음으로 노래를 찾는 QBH(Query by Humming)기술 등이 VUI에 적용될 수 있다. 또한 명령어에 ‘우울’ ‘슬픈’ ‘즐거운’ 등의 감정이 나타날 경우, VUI는 음악을 재생하는 재생자(Player)에서 나아가 사용자의 감정을 지지하는 시도를 하는 커뮤니케이션 대상으로 기능해야 할 것이다.

5. 4. 사용 맥락에 대한 정보를 제공하는 요소로의 목소리 크기의 활용

음성 인터페이스를 탑재한 지능형 에이전트는 사회적 존재로 인식되며, 상황과 맥락에 대한 이해를 갖추기로 요구된다. 연구결과, VUI 디자인 과정에서 말소리(Volume)크기는 사용자의 맥락적 정보가 담긴 하나의 도구로 기능할 가능성이 높다는 점이 확인되었다. 사용자가 소곤거리는 목소리로 대화형 에이전트에게 발화하는 속삭임이 그 대표적인 예이다. 본 연구결과 사용자(P6)가 아기를 재우는 상황에서 속삭이는 목소리로 “자장가 틀어줘”라고 발화하고, 아기가 잠든 후에 목소리를 낮추고 “음악 꺼줘”라고 요청하는 현상이 관측되었다. 또한 그 과정에서 AI스피커도 사람과 마찬가지로 속삭임에 내포된 의미를 이해하고 평소보다는 작은 소리로 음악을 재생하거나, 조용한 목소리로 사용자에게 응답해주기를 기대한다고 서술하였다. 이와 같은 점에서 속삭임은 하나의 맥락적 정보로 인식될 수 있으며, 사용자의 명령어에 속삭이는 음성이 포함되는 경우에는 에이전트도 속삭이거나 낮은 목소리로 응답할 필요성이 높다고 해석이 가능하다.

그러나 현재 상용화된 기기는 사용자의 음성을 텍스트로 변환하여 이해하는 STT(Speech to Text)기술을 사용하기 때문에, 사용자의 목소리 크기를 하나의 맥락적 정보로 인식하고 이에 민감하게 반응하는 기능은 미비한 수준이다. 그러나 현재 음성인식기술은 사용자의 목소리 크기를 인식할 수 있는 수준으로 발전하였으며, TTS에서도 발화 운율 모델링(Speech prosody Modeling)기술이 향상됨에 따라(Petrushin, Tsirulnik, & Makarova,

2010) VUI의 목소리 크기가 상황에 따라 변화할 수 있는 가능성이 높아지고 있다. 실제로 최근 아마존의 AI스피커 알렉사(Alexa)가 일부 국가를 대상으로 사용자의 속삭임에 속삭임으로 반응하는 기능을 업데이트하였다는 점은 음성형 인터랙션이 목소리 크기에 대한 반응을 포괄하는 방향으로 확장되고 있음을 시사한다. 그러나 아직 국내 AI스피커에서는 아직 이와 같은 기능이 구현이 되고 있지 않다는 점에서 향후 관련 기능에 대한 검토가 필요할 것이다.

6. 결론

본 논문은 음성 인터페이스의 도입이 가져온 음악 서비스 사용행태 변화를 관찰하기 위한 목적으로, 유튜브 영상의 시청각정보를 기반으로 상용화된 VUI인 AI스피커 사용자들의 행동을 관찰 분석하였다. 연구결과, VUI에서 사용되는 총 5개 영역, 29개의 음악 서비스 세부 기능이 도출되었으며 VUI로 새로운 음악 서비스 사용자 니즈가 나타나고 있음이 확인되었다.

본 연구는 VUI의 핵심 도메인 분야인 음악 서비스에 집중하여 인터페이스 전환에 따른 새로운 사용자 니즈(Needs)를 확인하고 디자인 함의점을 제공한다는 점에서 연구 의의를 지닌다. 또한 기존 인터넷 댓글과 음성 녹음 데이터를 중심으로 이뤄진 VUI 사용자분석방법에서 나아가 시각적인 정보를 포함한 유튜브 데이터로 VUI 연구 영역을 확장한다는 점에서 사용자 경험 조사 연구 분야에 시사점을 제공할 것으로 기대된다.

그러나 유튜브 콘텐츠라는 사용자들이 직접 업로드한 영상물을 관찰 분석했기 때문에 사용자가 영상을 기획하고 편집하는 과정에서 실사용과는 다른 형태의 인터랙션이 포함되었을 수 있으며, 극단적 사용자 그룹을 대상으로 했다는 점에서 연구 한계점을 지닌다. 따라서 '자연스러운' 실생활에서의 음성 인터페이스 사용 경험을 분석하기 위해서 확대된 그룹의 사용자를 대상으로 실제 집안환경에서의 사용과정을 관찰 분석하는 후속 연구가 필요할 것이다. 또한 다양한 그룹의 사용자 영상을 수집할 수 있는 유튜브 데이터의 장점을 활용하여 국내뿐 아니라 국내 외 인공지능 스피커 사용자의 사용 행태를 비교 분석하는 후속 연구를 제언한다.

References

1. Baker, N., Paddison, M., & Scruton, R. (2001) "Expression", *The New Grove Dictionary of Music and Musicians* (2nd ed.) London: Macmillan Publishers Limited. doi:10.1093/gmo/9781561592630.article.09138
2. Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B., & Lottrjdge, D. (2018). Understanding the Long-Term Use of Smart Speaker Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 1-24. doi:10.1145/3264901
3. Blythe, M., & Cairns, P. (2009, April). Critical methods and user generated content: the iPhone on YouTube. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1467-1476). ACM. doi:10.1145/1518701.1518923
4. Cohen, M. H., Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). Voice user interface design. Addison-Wesley Professional.
5. Crabtree, A., & Rodden, T. (2004). Domestic routines and design for the home. *Computer Supported Cooperative Work (CSCW), 13*(2), 191-220.
6. Cunningham, S. J., Reeves, N., & Britland, M. (2003, May). An ethnographic study of music information seeking: implications for the design of a music digital library. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries* (pp. 5-16). IEEE Computer Society.
7. Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems, 12*(4), 436-445.

8. GMR (2018, April 24). Global Music Report: Annual State of the Industry. Retrieved from https://www.ifpi.org/downloads/GMR2018.pdf
9. Google Developers. (2018, May). Design Actions for the Google Assistant beyond smart speakers (Google I/O '18) [Video file]. Retrieved from https://www.youtube.com/watch?v=JDakZMlXpQo
10. Hargreaves, D. J., & North, A. C. (1999). The functions of music in everyday life: Redefining the social in music psychology. *Psychology of music, 27*(1), 71–83.
11. Holland, S., Mudd, T., Wilkie-McKenna, K., McPherson, A., & Wanderley, M. M. (Eds.). (2019). *New Directions in Music and Human-Computer Interaction*. Springer.
12. Hoy, M. B. (2018). Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical reference services quarterly, 37*(1), 81–88. doi:10.1080/02763869.2018.1404391
13. Karat, C. M., Halverson, C., Horn, D., & Karat, J. (1999, May). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In P*roceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 568–575). ACM. doi:10.1145/302979.303160
14. Luger, E., & Sellen, A. (2016, May). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). ACM. doi:10.1145/2858036.2858288
15. Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing research, 40*(2), 120–123.
16. Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018, April). Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 6). ACM. doi:10.1145/3173574.3173580
17. Paay, J., Kjeldskov, J., Skov, M. B., & O'Hara, K. (2012, May). Cooking together: a digital ethnography. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 1883–1888). ACM.
18. Paay, J., Kjeldskov, J., Skov, M. B., & O'hara, K. (2013, September). F-formations in cooking together: A digital ethnography using youtube. In *IFIP Conference on Human-Computer Interaction* (pp. 37–54). Springer, Berlin, Heidelberg. doi:10.1145/2212776.2223723
19. Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544.
20. Petrushin, V. A., Tsirulnik, L. I., & Makarova, V. (2010). Whispered speech prosody modeling for TTS synthesis. In *Speech Prosody 2010-Fifth International Conference*.
21. Pink, S. (2016). Digital ethnography. *Innovative methods in media and communication research*, 161–165.
22. Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018, April). Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*(p. 640). ACM. doi:10.1145/3173574.3174214
23. Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June). Hey Alexa, What's Up?: A Mixed-Methods Studies of In-Home Conversational Agent Usage. *In Proceedings of the 2018 on Designing Interactive Systems Conference 2018* (pp. 857–868). ACM. doi:10.1145/3196709.3196772
24. Shechtman, N., & Horowitz, L. M. (2003, April). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281–288). ACM. doi:10.1145/642611.642661
25. Wasson, C. (2000). Ethnography in the field of design. *Human organization*, 377–388.
26. Weston, J., Bordes, A., Chopra, S., Rush, A. M., van Merriënboer, B., Joulin, A., & Mikolov, T. (2015). Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698*.

음성 인터페이스에서의 음악 서비스 사용행태 및 인터랙션에 관한 탐색적연구 : 유튜브 영상 분석을 기반으로

정유인[1], 이주호[2], 김새난슬[2], 강연아[3*]

[1]연세대학교 기술경영학협동과정, 학생, 서울, 대한민국

[2]연세대학교 디자인인텔리전스 전공, 학생, 서울, 대한민국

[3]연세대학교 언더우드국제대학, 교수, 서울, 대한민국

초록

연구배경 최근 시스템과 음성형의 대화가 가능한 음성-사용자 인터페이스(Voice User Interface)가 상용화됨에 따라 시스템과 사용자 간의 새로운 인터랙션 요소들이 등장하고 있다.

연구방법 이에 본 연구는 음성 인터페이스에서의 핵심서비스 도메인인 음악 서비스를 중심으로, 디지털 에스노그라피 방식을 적용하여 인터페이스 전환에 따른 새로운 사용자 니즈(Needs)를 탐색하고 4가지의 디자인 함의점을 제안하였다.

연구결과 사용자가 유튜브에 생성한 25개의 인공지능 스피커 사용 경험영상을 관찰 분석한 결과, 총 5개 영역의 29개의 음악 서비스 관련 기능들이 음성형 인터페이스에서 나타났다. 또한 그 중에서도 음성 인터페이스의 등장으로 1) 인터랙션 기능의 단축 및 기능 간의 통합 2) 음악재생기능에서의 새로운 인터랙션 요소의 등장 3) 멜로디/속삭임의 활용의 변화가 확인되었다.

결론 이러한 결과는 음성인터페이스로 전달되는 음악 서비스를 디자인하는 과정에서 그래픽-사용자 인터페이스에서 사용되던 기능 영역을 넘어 음성 대화라는 인터페이스의 장점을 극대화할 수 있는 새로운 서비스와 인터랙션 요소가 필요함을 시사한다.

주제어 음성-사용자 인터페이스, 음악 서비스, 인공지능스피커, 질적연구, 디지털 에스노그라피

*교신저자 : 강연아 (kang.younah@gmail.com)