

AI 스피커 멀티턴 서비스의 대화 분리감을 위한 청각큐 연구

Study on Auditory Cues for Conversation Separateness in AI Speaker Multi-Turn Services

조수빈

Soobin Cho

서울대학교
융합과학기술대학원
GSCST,

Seoul National University
soobincho0930@snu.ac.kr

이중식

Joongseek Lee

서울대학교
융합과학기술대학원
GSCST,

Seoul National University
joonlee8@snu.ac.kr

요약문

본 연구에서는 AI 스피커의 멀티턴 서비스 사용 시 사용자가 멀티턴 서비스의 대화 분리감을 인지하는지, 어떤 청각큐가 대화 분리감에 어떤 효과를 제공하는지, 그리고 사용자가 어떤 청각큐를 선호하는지 알아보고자 하였다.

이를 위해 조사 참가자 10 명이 20 분간 청각큐를 제공하지 않는 프로토타입과 각각 세 가지 다른 청각큐를 제공하는 세 가지 프로토타입을 사용하도록 한 후 15-20 분간 사후 인터뷰를 진행하였다. 프로토타입에 사용된 세 가지 청각큐는 Auditory Icon, Musicon, 그리고 Voice Alter 이다.

조사 결과, 멀티턴 서비스의 대화 분리감은 약하며, 청각큐가 대화 분리감을 증대시키는 것에 도움을 준다는 것을 파악하였다. 가장 분리감이 강한 청각큐는 특이한 목소리의 Voice Alter 지만, 강한 분리감과는 별개로 자연스러우며 서비스에 잘 어울리는 Musicon 이나 Auditory Icon 에 대한 선호도가 높았다.

본 연구에서 탐색한 내용은 대화 분리감이 중요한 AI 스피커의 멀티턴 서비스 디자인에 활용될 수 있다.

주제어

AI 스피커, 멀티턴 서비스, 대화 분리감, 청각큐

1. 서론

기존 스피커에 인공지능 기능이 더해진 인공지능(AI) 스피커의 보급이 빠르게 이루어지고 있다.[3] 과학 기술정보통신부에 따르면 지난 3 월 이동통신 3 사와 네이버카카오의 AI 스피커 누적 판매대수는 412 만대로 작년 동기보다 106% 증가했다.[2] AI 스피커의 주된 인터페이스는 ‘음성’으로, 사용자와의 대화를 기반으로 구동된다. 음성 사용자 인터페이스 (Voice User Interface, VUI)는 인간의 말에서 의미를

분석하여 인간-기계 상호작용을 가능하게 하는 인터페이스 방식이다[1].

AI 스피커의 서비스는 기기의 제조사에서 제공하는 제조사 서비스와 제조사가 아닌 기업이나 단체, 또는 개인이 제작한 서드파티 서비스로 이루어져 있다. 구글의 구글홈 AI 스피커에서 날씨 정보나 시간 정보가 제조사 서비스에 해당되며, ‘미세미세’ 어플과 디즈니의 ‘겨울왕국 영어이야기’가 서드파티 서비스에 해당된다. 기기의 사용자가 다양한 서비스를 즐길 수 있게 한다는 점에서 서드파티 서비스를 보조하는 것은 사용자 경험적으로도 중요하다.

AI 스피커 서비스를 구분하는 또다른 기준으로는 싱글턴 서비스와 멀티턴 서비스가 있다. 싱글턴 서비스에서는 사용자가 “오케이 구글”이나 “헤이 구글”과 같은 웨이크 단어 이후 하나의 명령만을 입력한다.[5] 싱글턴 서비스에는 시간 정보와 같은 제조사 서비스와 미세먼지 농도를 알려주고 바로 종료하는 ‘미세미세’ 어플과 같은 싱글턴 서드파티 서비스가 모두 해당된다. 반면 멀티턴 서비스에서는 사용자가 웨이크 단어와 함께 서비스를 시작하면 해당 대화형 에이전트와의 대화가 시작된다. 주로 챗봇의 형태를 띠는 멀티턴 서비스에서 사용자는 분명한 시작과 끝을 가진 ‘단한 대화’에 참여하게 된다.[4] 구글홈 AI 스피커의 ‘타로챗봇 라마마’와 ‘인터파크’ 챗봇과 같은 멀티턴 서드파티 서비스가 멀티턴 서비스에 해당된다.

제조사 서비스나 싱글턴 서드파티 서비스와 같은 싱글턴 서비스를 사용하다가 멀티턴 서비스와의 ‘단한 대화’에 참여하게 되면 사용자는 분리된 대화를 경험하게 된다. 대화의 분리감을 인지하지 못하고 멀티턴 서비스와의 ‘단한 대화’ 중에 제조사 서비스나 싱글턴 서드파티 서비스를 위한 명령어를 입력하면 사용자와 대화 중이던 멀티턴 서비스에서 오류가

발생하게 된다. 예로, 구글홈 AI 스피커에서 멀티턴 서드파티 서비스인 ‘타로챗봇 라마마’와 대화 도중 “오케이 구글, 날씨 알려줘”와 같은 제조사 서비스를 위한 명령어를 입력하면 ‘타로챗봇 라마마’가 진행되지 못하는 오류가 발생한다.

시각 인터페이스가 없는 AI 스피커의 멀티턴 서비스에서는 위같은 오류를 방지하기 위해 청각큐를 이용하여 대화의 분리감을 증대시킬 수 있다. Yoon et al.[1]의 연구에서는 VUI 연구에서 사용되는 5 가지 청각 피드백을 탐색하여 실험을 설계하였다. 가공된 추상적 소리인 Earcon 과 실제 대상물의 대표적인 소리를 메타포로 이용하는 Auditory Icon 은 듣기 상태가 시작될 때와 끝날 때에만 제공하는 단절음으로 사용하고, 정보 전달을 위한 데이터의 청각 표현인 Sonification 과 음악의 짧은 샘플인 Musicon, 주변적인 청각 소리인 Ambient 는 듣기 상태가 시작해서 끝날 때까지 제공하는 지속음으로 사용하였다.

본 연구는 AI 스피커의 멀티턴 서비스에서 어떤 청각큐를 제공하면 사용자가 대화의 분리감을 인지하는 것에 어떻게 도움을 줄 수 있을지, 그리고 어떤 청각큐를 선호하는지 알아보고자 한다. 이를 위해 조사 참가자 10 명이 다양한 청각큐를 제공하는 프로토타입을 사용하도록 한 후 인터뷰를 진행하였다.

2. 조사 설계

본 조사를 위해 편의 추출 방법을 이용하여 AI 스피커 사용 경험이 있는 인원으로 총 10 명(남 5, 여 5)을 모집하였다. 모든 조사 참가자는 동일한 구글홈 AI 스피커를 이용하여 청각큐를 제공하지 않는 멀티턴 서비스를 사용 후, 동일한 서비스를 세 가지 다른 청각큐를 추가하여 다시 사용하도록 했다. 이후 사후 인터뷰를 통해 질적 연구를 진행했다.

2.1 프로토타입 개발

본 조사를 위해 구글 dialogflow 를 이용하여 운세를 알려주는 AI 스피커 서비스 ‘포춘텔러’의 프로토타입을 개발했다. ‘포춘텔러’는 ‘단한 대화’를 발생시키는 멀티턴 서비스로, 참가자의 개입을 최대화하기 위해 인터랙션 위주의 챗봇 형태로 디자인했다. 위 서비스는 실제 조사에 사용될 구글홈 AI 스피커와 유사한 남성 목소리를 사용했다. ‘포춘텔러’는 그림 1 과 같은 서비스 설계를 사용한다.

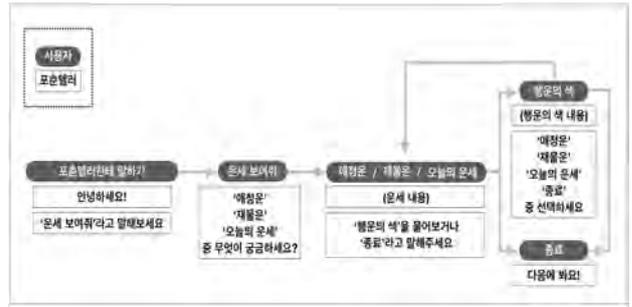


그림 1 ‘포춘텔러’ 설계도

동일한 서비스에 각각 세 가지 다른 청각큐를 추가하여 ‘포춘텔러 이펙트’, ‘포춘텔러 배경음악’, ‘포춘텔러 보이스’의 프로토타입을 개발했다. 본 연구에서는 Yoon et al.[1]의 연구에서 사용된 5 가지 청각 피드백 중 단절음인 Auditory Icon 과 지속음인 Musicon 외에 이질적인 목소리를 추가하여 조사를 진행했다. Auditory Icon 을 사용하는 ‘포춘텔러 이펙트’는 듣기 상태가 시작될 때와 끝날 때 동전이 부딪히는 효과음을 삽입했다. Musicon 을 사용하는 ‘포춘텔러 배경음악’은 듣기 상태가 시작해서 끝날 때까지 지속적으로 동일한 배경음악을 삽입했다. ‘포춘텔러 보이스’는 실험에 사용된 AI 스피커와 나머지 포춘텔러 서비스에서 사용하는 남성 목소리가 아닌 여성 목소리를 사용했다. 네 가지 버전의 포춘텔러 서비스 특징은 표 1 과 같다.

표 2 포춘텔러 각 버전 서비스 특징

	조건	청각큐 내용
포춘텔러	No Control	
포춘텔러 이펙트	Auditory Icon	• 듣기 상태의 시작과 끝에 동전 부딪히는 효과음
포춘텔러 배경음악	Musicon	• 듣기 상태가 시작해서 끝날 때까지 배경음악
포춘텔러 보이스	Voice Alter	• 여성 목소리

2.2 조사 방법

각 참가자가 20 분동안 자유롭게 구글홈 AI 스피커를 이용하며 청각큐를 사용하지 않는 ‘포춘텔러’를 3 회 사용 후 ‘포춘텔러 이펙트’, ‘포춘텔러 배경음악’, ‘포춘텔러 보이스’를 각 한 번씩 사용하도록 하였다. 세 가지 청각큐의 사용 순서 효과를 상쇄시키기 위해 청각큐의 순서는 참가자마다 랜덤으로 배정하였다. 포춘텔러 서비스 사용 시에는 턴 횟수를 제한하지 않았으며, 총 여섯 번의 포춘텔러 서비스 사용 중간에 싱글턴 서비스인 제조사 서비스를 자유롭게 사용하도록 하였다. 자연스러운 AI 스피커 사용 환경을 조성하기 위해 조사 참가자의 필요에 따라 20 분의

시간 내에서 자유롭게 개인의 업무를 보거나 휴식을 취하도록 했다.

모든 참가자는 20 분의 AI 스피커 사용 직후 15-20 분의 사후 인터뷰에 답변했다. 인터뷰에서는 크게 멀티턴 서비스의 인지된 대화 분리감, 세 가지 청각큐의 대화 분리 효과, 그리고 세 가지 청각큐의 선호도에 대해서 물어봤다. 모든 조사 내용 및 인터뷰 내용은 동의 하에 녹음되었다.

3. 조사 결과

3.1 인지된 대화 분리감 및 오류 발생 가능성

10 명의 참가자 중 8 명의 참가자(P03, P04, P05, P06, P07, P08, P09, P10)가 청각큐가 없는 ‘포춘텔러’는 대화의 분리감이 강하게 느껴지지 않으며 청각큐가 있는 것이 분리감을 증대시키는 것에 도움을 준다고 답변했다. 나머지 2 명의 참가자(P01, P02)는 청각큐가 없는 ‘포춘텔러’를 사용 시에도 대화의 분리감이 확실히 느껴지며, 청각큐가 있다고 해서 분리감이 크게 증대되는 것 같지 않다고 답변했다. P01 은 “서비스 안에서는 내가 여기에 있다는 걸 잘 인지하고 있는 것 같아요. ...지금 전면에 있기 때문에”라고 답변했다.

20 분간의 AI 스피커 사용 도중 멀티턴 서비스와 싱글턴 서비스를 분리하지 못해 오류를 발생시킨 참가자는 총 3 명(P08, P09, P10)이었다. 동일한 3 명의 참가자는 멀티턴 서비스와 길게 대화 시 약한 분리감 때문에 오류가 발생할 수 있을 것 같다고 답변했다. 반면, 나머지 7 명의 참가자는 멀티턴 서비스의 분리감이 강하지 않더라도 오류를 발생시키지 않을 것이라고 답변했다.

3.2 세 가지 청각큐의 대화 분리 효과

10 명 중 2 명의 참가자(P05, P06)가 Auditory Icon, Musicon, Voice Alter 의 세 가지 청각큐 중 Voice Alter 가 분리감이 가장 강하다고 답변했고, 위 2 명의 참가자를 포함한 6 명의 참가자(P03, P04, P05, P06, P07, P08)가 본 조사에서 Voice Alter 에 여성 목소리가 아닌 특이한 목소리를 사용했다면 Voice Alter 가 가장 분리감이 강할 것이라고 답변했다.

Musicon 이 가장 분리감을 증대시킨다고 답변한 참가자는 4 명(P02, P07, P08, P10)이었으나 이중 2 명(P07, P08)은 특이한 목소리의 Voice Alter 가 사용될 경우 Voice Alter 가 Musicon 보다 분리감이 강할 것이라고 답변했고, 나머지 2 명(P02, P10)의 참가자는 특이한 목소리의 Voice Alter 가 분리감이 가장 클 것이라는 확신은 없다고 답변했다.

Auditory Icon 이 가장 분리감을 증대시킨다고 답변한 참가자는 3 명(P03, P04, P09)이었으나 이중 2 명(P03, P04)은 특이한 목소리의 Voice Alter 가 사용될 경우 Voice Alter 가 Auditory Icon 보다 분리감이 강할 것이라고 답변했다. 나머지 1 명의 참가자(P09)는 Auditory Icon 에 대해 “되게 명확하게 ... 이펙트가 저한테는 애가 시작되는 시그널이구나 라는게 명확했던 것 같아요.”라며, 특이한 목소리의 Voice Alter 가 사용되더라도 Auditory Icon 이 가장 분리감이 강할 것이라고 답변했다.

1 명의 참가자(P01)는 세 가지 청각큐 모두 분리감을 크게 증대시키지 않는다고 답변했다.

3.3 세 가지 청각큐의 선호도

10 명 중 5 명의 참가자(P01, P02, P06, P07, P10)가 세 가지 청각큐 중 Musicon 을 가장 선호한다고 답변했다. 이중 3 명의 참가자(P01, P06, P07)가 Musicon 이 ‘포춘텔러’와 잘 어울린다는 점과 자연스럽다는 점을 이유로 꼽았다. 다른 2 명의 참가자(P01, P10)는 Musicon 이 듣기 상태가 시작해서 끝날 때까지 제공되는 지속음이기 때문에 턴테이킹 포인트를 알려준다는 점을 추가적인 장점으로 꼽았다.

Musicon 청각큐에 대한 부정적인 의견으로, Auditory Icon 청각큐를 가장 선호한다고 답변한 참가자 중 2 명의 참가자(P03, P09)가 Musicon 이 음악을 주로 트는 스피커의 역할과 겹쳐 분리감이 크게 도움이 되지 않는다고 답했다. P03 은 “배경음악도...위낙 음악 듣는 용도로 많이 쓰다 보니까...”라고 답변했으며, P09 는 “배경음악도 사실 음악을 듣다가 와서 그런건지 조금 헷갈릴 수도 있을 것 같은데...”라고 답변했다.

10 명 중 4 명의 참가자(P03, P04, P05, P09)는 세 가지 청각큐 중 Auditory Icon 을 가장 선호했다. 4 명의 참가자 모두 ‘포춘텔러 이펙트’에서 사용된 Auditory Icon 이 실제 소리의 메타포라는 점에서 서비스에 대한 몰입도를 높인다고 답했다. P04 는 “효과음이 약간 몰입감을 좀 더 줬던 것 같아요. 정체성을 부여해서...”라고 추가했으며, P09 는 “포춘텔러 이펙트가...더 많이 사용하게 유도하는 느낌이 있어요. 분위기를 조성해서 그런지...”라고 답변했다.

Auditory Icon 청각큐에 대한 부정적인 의견으로, Musicon 청각큐를 가장 선호한다고 답변한 P02 는 “효과음이 시작할 때랑 끝날 때 계속 나니까...이 소리도 있고 ‘띠링’ 소리도 있어서 소리가 너무 많았어요.”라며 구글홈 자체의 효과음과 더해져 효과음이 너무 많은 느낌이라고 답했다. Musicon 을

선호한다고 답변한 P07 과 선호하는 청각큐가 없다고 답변한 P08 은 Auditory Icon 이 쉽게 “흐려지고”, “잃어버리기 쉬울 것”이라고 답했다.

1 명의 참가자(P08)는 특별히 선호하는 청각큐가 없다고 답변했다.

4. 결론 및 한계

본 연구에서는 사용자가 멀티턴 서비스의 대화 분리감을 인지하는지, 세 가지 청각큐(Musicon, Auditory Icon, Voice Alter)가 대화 분리감에 어떤 효과를 제공하는지, 그리고 위 세 가지 청각큐의 선호도는 어떠한지 살펴보았다.

멀티턴 서비스의 대화 분리감은 약하며, 청각큐가 대화 분리감을 증대시키는 것에 도움을 준다. 가장 분리감이 강한 청각큐는 특이한 목소리의 Voice Alter 지만, 강한 분리감과는 별개로 자연스러우며 서비스에 잘 어울리는 Musicon 이나 Auditory Icon 에 대한 선호도가 높다.

본 연구에서 탐색한 내용은 대화 분리감이 중요한 AI 스피커의 멀티턴 서비스 디자인에 활용될 수 있다. 하지만 본 연구가 조사 참가자 모집을 위해 편의 추출 방법을 사용한 점, 구체적인 특정 서비스만을 조사 대상으로 선정한 점, 그리고 조사를 위해 한 가지만의

Musicon 과 Auditory Icon, Voice Alter 를 선정한 점이 한계로 적용될 수 있다.

참고 문헌

- 1.윤종목. Eyes-free 음성 사용자 인터페이스의 상호작용 사용성을 개선시키는 청각 표상 연구. 서울대학교 융합과학기술대학원 (2019)
- 2.연합뉴스. 초연결 지능형 네트워크 ‘AI스피커’ 412만대 보급 ... 106% 증가. <https://www.yna.co.kr/view/AKR20190701042000017?input=1195m>.
- 3.조규은 and 김승인. 인공지능 스피커(AI speaker) 사례 분석을 통한 고찰. 한국융합학회논문지 9, 8 (2018), 127-133.
- 4.Cathy Pearl. Designing Voice User Interfaces: Principles of Conversational Experiences. O'Reilly Media, Inc. (2016)
- 5.Tips on Designing Conversations for Voice Interfaces. <https://uxdesign.cc/tips-on-designing-conversations-for-voice-interfaces-d4084178cfd2>.