

AI 음성 에이전트의 음성 특성에 대한 사용자 반응 연구*

백승주** · 정윤혁***

〈목 차〉

I. 서론	IV. 연구결과
II. 이론적 배경	4.1 조작점검
2.1 목소리 연구	4.2 실험 결과 분석
2.2 AIVA에 대한 호감도	V. 토의 및 향후 연구
2.3 연구문제	4.1 결과 토의
III. 연구방법론	4.2 한계점 및 향후 연구
3.1 파일럿 테스트	참고문헌
3.2 실험설계와 변인측정	<Abstract>

I. 서론

최근에는 인공지능(Artificial Intelligence, AI)이 디지털 기기와 결합하여 음성으로 사용자에게 다양한 서비스를 제공하는 AI 음성 에이전트(AI Voice Agent, AIVA)가 주목을 받고 있다. AI와 관련된 언론보도에서도 AI 음성 서비스가 주요한 키워드가 되고 있다(박종화 등, 2022). 2011년 애플사에서 AIVA 시리(Siri)가 출시되면서 구글, 아마존, 삼성전자, 카카오, 네이버 등 국내외 대표적인 정보통신기업 기업들도 AIVA를 개발 상용화를 추진하고 있다(임종

수 등, 2020). AIVA는 사용자에게 간단한 생활 정보를 제공하는 것부터 검색, 주문, 추천기능까지 사용자 필요와 요구에 맞춰 다양한 형태로 서비스를 제공하고 있다(임종수 등, 2020). AIVA는 음성을 매개로 사용자의 명령에 따라 집안의 가전기기를 제어하거나 사용자 맞춤형 개인화 서비스를 제공하는데, 음악 및 영화 추천 등이 그 대표적인 예라고 할 수 있다(김정현 등, 2018).

AI를 활용한 음성 서비스가 증가하면서 사용자들은 사람과 유사한 AIVA와의 상호작용이 점차 일상화되고 있다(Guzman, 2019). 더욱이

* 본 연구는 고려대학교 미디어학부 특별연구비 지원을 받아 수행된 연구임. 또한 이 논문은 2020년 대한민국 교육부와 한국연구재단의 일반공동연구지원사업의 지원을 받아 수행되었음. (NRF-2020S1A5A2A03041137).

** 고려대학교 미디어학과, monica100@korea.ac.kr(주저자)

*** 고려대학교 미디어학부, beyond@korea.ac.kr(교신저자)

2020년 코로나바이러스(코로나) 감염의 확산으로 사람들은 실내에서 머무는 시간이 많아졌고, 이에 따라 AIVA 사용은 급속히 증가하였다. 날씨/뉴스, 교통과 같은 기본적인 정보를 검색하거나 영상/음악 재생(김정현 등, 2018)을 넘어, 음식 주문이나 쇼핑, 제품 비교 및 추천까지 서비스의 범위가 확장되었고, AIVA와 연동된 홈트레이닝과 엔터테인먼트 서비스도 등장하고 있다(이슬기 등, 2020).

이러한 AIVA의 핵심 구성요소는 에이전트와 사용자를 매개하는 목소리이다. AI 스피커의 목소리의 특징이 사용자와 스피커 관계에 유의미한 영향을 미친다는 실증적 증거가 제시되고 있다(구본권, 2021). 하지만, AIVA 목소리의 역할에 관한 관심에도 불구하고, 여전히 이 분야에 대한 경험적 연구는 부족한 실정이다(Johnson et al., 2020). 인간의 목소리에 대한 연구에 따르면, 화자의 목소리 특징은 그에 대한 청자의 인식에 유의미한 영향을 미친다(박란희 등, 2009). 그러므로, AIVA와 사용자의 관계를 이해하기 위해서는 AIVA 목소리 특성과 이에 대한 사용자의 반응을 탐색하는 것이 필요하다. 본 연구는 목소리 특성(성별, 음도, 발화 속도)이 사용자의 AIVA 평가에 어떤 영향을 미치는지 알아보고자 한다. 나아가, 각 특성 간에 상호작용 효과도 분석하고자 한다

II. 이론적 배경

2.1 목소리 연구

음성은 AIVA가 사용자와의 상호작용에 있

어 기본적인 인터페이스로서 상호 간의 커뮤니케이션에 큰 영향을 줄 수 있다(Jeong et al., 2019). AIVA의 사람 목소리는 사용자와의 사회적 대화뿐만 아니라, 감정적으로 교류를 가능하게 한다(Kreijns et al., 2003). AIVA는 향후 사용자의 감정을 파악하고, 목소리를 통해 자신의 기분과 감정을 표현하는 단계까지 발전할 것(곽노건, 2020)이라는 점에서 AIVA 목소리를 매개로 한 사용자의 상호작용에 대한 연구는 더욱 강조될 것이다.

목소리는 준 언어적 표현(Paralanguage expression) 요소들과 함께 청자에게 전달되기 때문에 준 언어적 차원에서 목소리 분석은 의미 전달의 명확성과 청자의 인식에 중대한 영향을 미칠 수 있다(이유나 등, 2008). AIVA 역시 목소리를 매개로 사용자와 소통하기에 에이전트의 목소리는 앞으로 더욱 중요하게 다뤄질 것이다. 기존에는 청자에게 긍정적 반응을 이끌어 내는 사람 목소리의 특성에 관한 연구들이 수행되었지만, 이를 AIVA 맥락에 적용한 경험적 연구는 부족하고, 따라서 AIVA의 주요한 인터페이스인 목소리의 특성에 대한 연구가 요구된다.

사람은 의사소통 시 메시지 내용을 더욱 정확히 전달하거나 강조하기 위해 비언어적 차원의 커뮤니케이션을 사용한다. 사람 간 커뮤니케이션에는 세 가지 메시지 전달요소, 즉 시각적 요소(55%), 음성적 요소(38%), 내용적 요소(7%)가 있고, 나열된 순으로 전달 효과가 크다(Mehrabian, 1972). 메시지 내용도 의사 전달에 중요한 부분을 차지하나, 표정이나 목소리, 몸짓과 같은 비언어적 차원의 커뮤니케이션이 화자의 이미지나 메시지 의도, 혹은 청자의 화자

에 대한 호감 정도에 영향을 크게 미치는 것으로 볼 수 있다.

비언어적 차원의 커뮤니케이션 요소들을 어떻게 활용하느냐에 따라 청자는 화자의 메시지에 다르게 반응하고 화자에 대한 이미지를 결정한다(이유나 등, 2008). 실제로 사람은 커뮤니케이션에 있어 비언어적 수단(65%)을 통해 메시지를 이해한다(Birdwhistell, 1964). 김우룡 등 (2004)은 언어적인 내용보다 비언어적 표현 요소가 커뮤니케이션에 있어 더 결정적이라 보았다. 화자의 비언어적 요소는 화자에 대한 이미지나 평가, 메시지 의도를 이해하는데 영향을 미쳐, 비언어적인 커뮤니케이션이 메시지 전달에 있어 90%를 차지한다는 연구 결과도 있다(강길호 등, 2001).

비언어적 차원의 커뮤니케이션은 언어적 차원의 표현으로 해결이 되지 않는 부분을 보완해준다. 또한 메시지의 의미를 강화하기도 하고 대체하는 등 언어적 표현 이상을 전달할 수 있다. 비언어적 차원의 커뮤니케이션으로는 시각적 요소와 음성적 요소가 있고, 외모나 목소리가 대표적이다. 목소리의 경우 화자는 청자에게 본래 메시지 이상의 신뢰감을 제공하기도 한다(이유나 등, 2008). 목소리는 사람의 미묘한 감정을 전달해 메시지의 뜻보다 다른 느낌을 주기도 하며 화자에 대한 청자의 인상, 이미지에도 영향을 미친다(박란희 등, 2009). 본 논문은 비언어적 커뮤니케이션으로 음성적 요소, 즉 목소리의 역할을 AIVA 맥락에 적용해보고자 하였다. 목소리 구성요소는 준 언어 혹은 유사언어(Paralanguage)가 있고(오석영, 1997), 준 언어 요소는 발음, 말의 발화 속도, 목소리 크기, 울림, 어조, 억양, 강세, 그리고 음도 등으로 구

성된다. 그중에서 음도와 발화 속도는 가장 많이 연구된 목소리 특징이다(권순복, 2012). 아울러, 로봇 혹은 AIVA 연구에서 목소리의 성별에 관한 연구 역시 광범위하게 진행되어 왔기 때문에(e.g., Nass et al., 1997; Ahn et al., 2022), 본 연구는 대표적인 AIVA의 목소리 특징인 성별, 음도, 발화속도를 가지고 사용자의 반응을 탐색하고자 한다. 본 연구는 다양한 목소리의 특징을 포함하였다는 점에서 목소리 성별에 집중된 기존 AIVA 목소리 연구를 확장하고 있다.

2.2 AIVA에 대한 호감도

호감이란 좋게 여기는 감정 혹은 좋은 감정을 의미한다(Kong et al., 2015). 사전적 의미로 호감도는 어떤 대상에 대하여 좋은 감정을 갖는 정도이다. 기존 연구에서는 호감도를 긍정적인 감정, 우호적 감정, 좋고 나쁨으로 한정된 감정 등 대상에 대해 개인이 느끼는 감정으로 정의해왔다(Kuh, 2015; 이준상 등, 2021). 또한, 더 포괄적 관점에서 호감도는 상품이나 특정 대상에 대한 사용자의 반응이나 인식을 측정하기 위해 사용되기도 하였다. 본 연구는 호감도에 대한 사용자의 포괄적인 반응을 탐구하기 위하여 호감도를 ‘매력도’, ‘신뢰도’, ‘의인화’로 설정하였다. 3가지 하위요소를 설정하는 데는 기존 목소리 연구나 호감도 측정에 대한 선행 연구를 참고하였다(황보명 등, 2015; 권순복, 2015). 본 연구에서 사용자 반응인 호감도는 감정적 차원의 평가인 매력도, 인지적 차원의 평가인 신뢰도, 그리고 의인화 차원의 평가로 느끼는 의인화를 호감도의 하위개념으로 정의하

였다.

본 연구에서 감정적 차원 평가로 채택된 매력도는 소비자 관점에서 요구 및 필요성을 충족한 정도에 대한 평가로 볼 수 있다. 소비자의 욕구 및 기대에 대한 충족 정도 등을 바탕으로 이뤄지는 주관적 평가인 것이다. 본 연구는 사용자가 AI 음성을 듣고 매력을 느낄 때, 감정적 차원에서 긍정적인 평가를 한 것으로 가정한다. 매력도는 기존 연구들에서 호감이나 선호에 대한 하위개념으로 자주 사용되어왔고, 매력 척도는 소비자가 해당 서비스나 제품을 지속적으로 이용할 의지, 혹은 다른 사람에게 추천할 가능성을 보여준다. Gefen et al. (2004)은 사람 사이 정감을 느끼듯이 소비자는 제품 및 서비스를 제공받을 때 적절한 상호작용을 거치거나 정감 및 친밀감 등이 생기면 제품에 대한 호감 및 매력을 느낀다고 말하였다. 즉 매력도는 제품이나 대상, 혹은 서비스에 대한 만족, 감정적 평가 등을 의미한다.

인지적 호감도는 사람의 인지 과정을 통해 만족을 얻는 정도를 의미한다. 감정적 반응과 인지적인 반응은 유기적으로 작용하여 소비자가 제품을 구매하거나 해당 상품을 선택하는데 결정적 역할을 한다. 본 연구는 인지적 호감도로 신뢰도를 사용한다. 신뢰는 호의적인 행동에 대한 기대로 정의되며, 거래나 선택 등의 중요한 의사결정에 있어 선행조건이 된다(Gefen, 2000). 신뢰는 안정성, 믿음 등의 성격을 띠며 신뢰도는 앞에서 말한 호감도의 분류 중 인지적 만족에 해당한다. 사람의 정서와 감정은 의사결정이나 선택에 있어 중요한 영향을 미친다 (Vohs et al., 2008; Jung et al., 2021). 기존의 연구들은 신뢰를 사용자의 반응 중 하나로 간

주하였고, 로봇에 대해서도 사람은 신뢰도를 보일 수 있다(Jung et al., 2021). 신뢰는 소비자 관점에서 상품이나 서비스의 채택에 있어 핵심 요소이며 온라인 에이전트와 사용자 간의 상호작용 및 커뮤니케이션에도 중요한 역할을 한다 (McKnight, 2005; Wang et al., 2005). AIVA에서도 신뢰도는 중요한 요소로 작용한다. 휴머노이드와 같은 사람과 유사한 로봇의 경우, 사용자는 로봇에게 사람과 같은 역할을 기대하고, 신뢰도와 적절한 사회적 상호작용은 사용자가 로봇의 품질을 평가하고 수용하는데 영향을 줄 수 있다(Jung et al., 2021).

의인화는 AIVA에 대한 사용자의 반응을 이해하는 개념으로 활용되어 왔다. AIVA의 경우 사용자가 에이전트에 대해 사람과 같이 상호작용을 하고 사람과 같은 역할을 한다고 인식할 때 사회적 실재감을 느낀다고 말한다(Kim et al., 2014). 의인화는 에이전트의 외형이나 디자인으로도 형성이 가능하지만, 목소리라도 사용자는 AIVA에 의인화를 느낄 수 있다. 사용자는 AIVA를 사회적으로 실재하는 존재로 느끼기 때문에, 에이전트를 단순한 디지털 기기라기보다는 ‘비서’ 혹은 ‘친구’로 인식하며 그에 맞는 역할수행을 기대한다. AI 음성은 사용자에게 친밀감을 제공하고, 사용자와 상호작용하면서 사회적인 실재감을 느끼도록 한다. 이때 사용자는 AIVA를 긍정적으로 평가하고, 호감을 느끼게 된다. 호감도 및 의인화 정도를 로봇에 대한 사람의 평가 하위요소로 둔 선행연구(Breazeal, 2003; Ho et al., 2010)을 기초하면, 의인화는 로봇과의 상호작용, 즉 로봇의 사람과의 원활한 커뮤니케이션이 이뤄질 때 발생한다. 다시 말해 사람은 로봇과 상호작용이 원활히 이뤄질 때

의인화를 느낀다는 것이다. 이에 본 연구는 기존 연구를 바탕으로 사용자가 AIVA와 음성으로 상호소통을 할 때 의인화를 주요한 결과변수로 사용하였다. AIVA와 사용자의 상호작용 및 커뮤니케이션을 의인화의 개념에서 바라보는 것은 사용자의 AI에 대한 태도나 행동을 이해하는 데 도움을 줄 수 있다(임종수 등, 2020). 로봇이나 AIVA의 외형, 디자인, 혹은 시스템의 사람과 유사함으로 인한 의인화 현상은 기존 연구를 통해 많이 설명되어왔다(Nicole et al., 2003; Amin, 2005; Haslam et al., 2014; 최정건 등, 2008). 본 연구는 AIVA의 목소리 특성에 따른 사용자의 의인화 인식을 살펴보고자 하였다.

2.3 연구 문제

사람은 다양한 음역대 및 음도를 지니는데 성별에 따라 음도는 구분될 수 있다(Street et al., 1982; Smith et al., 1975; 박란희 등, 2009). Virtue Nordic 연구는 남성은 평균 80Hz, 중성 평균 145Hz에서 175Hz, 여성 평균은 220Hz 정도의 음도를 지닌다고 하였다. 오경아(2013)에 따르면 남성의 평균 음도는 100Hz에서 150Hz, 여성은 200Hz에서 250Hz 정도의 범위에 해당된다. 목소리의 성별에 따라 청자가 인식하는 화자의 이미지도 다르다. 보통 남성의 목소리는 권위적이고 당차며 영향력이 있는 것으로 인식되었다(이슬기 등, 2020). 여성 목소리의 경우는 서비스 제공 시 신뢰감을 주고 더 선호되었다(Lee et al., 2019). 고영태 (2019)는 목소리 성별은 분야별로 선호가 달라지기도 하는데, 컴퓨터 교육 같은 정보 성격의 분야는 남성 목소

리를, 정서나 사회성을 띠는 분야에서는 여성의 목소리가 선호됨을 밝혔다. 다른 연구에서도 사용자는 메시지 전달에 있어 남성의 목소리를 정보 전달을 위한 용도로, 사회적 의사소통에는 여성의 목소리가 남성보다 효과가 있었다(Furumo et al., 2007).

현재 서비스 중에 있는 AIVA 음성은 대부분 여성 목소리가 기본값으로 설정되어 있다. 사용자는 보통 서비스 제공에 있어 여성의 목소리를 선호하기 때문이다(Lee, 2018). 실제로 삼성 빅스비(Bixbi), 애플 시리(Siri), 아마존 알렉사(Alexa), 구글 에이전트(Google Assistant), KT 지니(Genie), SK텔레콤의 누구(NUGU) 등 대부분의 AIVA는 여성의 목소리가 기본이었다. AI 음성에 대한 성차별 논란이 제기되면서 AIVA에는 성별이나 목소리 종류를 변경할 수 있는 옵션이 제공되고 있다.

본 논문은 AIVA에서도 기존 목소리 연구와 같이 성별에 따라 사용자의 반응에 영향을 주는지 알아보려고 하였다. 특히 본 연구는 호감도를 척도로 사용자 반응 연구를 진행하였다. 앞서 소개한 대로 호감도란 좋게 여기는 감정(Kong et al., 2015)으로서 AIVA 목소리의 특성에 대한 사용자의 반응을 측정하기 위해 사용되었다. 사람 목소리에 대한 연구에서도 호감도는 청자의 반응 측정 요소로써 활용되었다(권순복, 2015; 이준상 등, 2021). 이에 따라 본 연구는 기존 연구를 참고하여 아래와 같은 연구 문제를 설정하였다.

연구문제 1: AI 음성 에이전트의 목소리의 성별은 사용자의 AI 음성 에이전트에 대한 호감도에 어떤 영향을 미치는가?

기존 실험 연구 중 목소리 준 언어 차원의 요소로 가장 빈도 높게 사용된 요소는 ‘음도’였다 (Aronovitch, 1976; Collins et al., 2003; Kim et al., 2005; Smith et al., 2005; Jones et al., 2008; Kwon, 2016). 사람 간 상호소통에 있어 목소리의 준 언어적 요소는 잘 활용하면 메시지 의미 전달이나, 효과적인 커뮤니케이션에 도움을 준다(권순복, 2015). 이런 이유로 사람 간 효과적인 의사소통을 위해 가장 적절한 목소리 높낮이(음도)가 무엇인지에 관한 연구가 폭넓게 수행되었다. 기존 논문들에 따르면 목소리 높낮이는 주로 성별에 따라 기대되는 기준이 존재했고, 음도에 대한 청자의 반응에는 차이가 존재하였다. 여성의 경우는 높은 음도가, 남성의 경우는 비교적 낮은 음도가 선호되는 것을 알 수 있었다. 하보미 등 (2018)은 여성의 경우 낮은 음도는 청자의 부정적인 반응을 야기한다고 하였다. Collins et al. (2003)은 음도에서 평균 여성보다 높은 음도일 때 청자의 선호도가 높아짐을 밝혔다. Jones et al.(2008)도 여성은 평균 여성보다 높은 목소리 주파수가, 남성은 평균보다 낮은 중저음의 목소리가 선호된다고 하였다. Kwon(2016)은 목소리 변조를 이용한 청자의 태도를 본 연구를 진행했는데, 실험을 위한 목소리 구성 요소로는 음도, 강도, 발화 속도가 있었다. 연구 결과 남성은 평균 20대의 남성 목소리보다 조금 낮은 목소리일 때, 여성은 보통 높은 목소리가 선호되나 음도가 일정 음도를 넘어설 경우는 오히려 평균보다 조금 낮은 음도(215.6 Hz)가 선호되었다. 한편, Aronovitch (1976)는 여성은 목소리 높낮이가 의사소통에 영향을 미치나 남성의 경우는 별 다른 영향을 주지 않았다. 본 연구는 사람 목소리가 아닌

AIVA의 목소리 경우에도 음도 변화에 따라 사용자 반응에 변화가 있는지 알아보려고 하였다. 사용자에게 호감을 얻는 AI 목소리 성별에 따라 높낮이는 어떠한지 알아보기 위해 아래와 같은 연구 문제를 설정하였다.

연구문제 2: AI 음성 에이전트의 음도는 사용자의 AI 음성 에이전트에 대한 호감도에 어떤 영향을 미치는가?

목소리의 준 언어 요소가 청자의 태도에 직접적으로 얼마나 영향을 끼치는지 탐구한 연구에서는 발화 속도와 씩이 중요한 구성 요인으로 밝혀졌다. 발화 속도는 화자의 감정과 태도를 반영하기 때문에 의사소통에 있어 중요한 요소로 작용한다(조일환, 2005). 발화 속도는 음도에 이어 목소리 실험 연구에서 빈도 높게 사용되어왔다(Mehrabian et al., 1969; Street et al., 1984; Smith et al., 2005). 발화 속도가 빠를수록 청자에게 긍정적인 반응을 불러일으켰다. 발화 속도가 빠르면 화자의 전문성을 높이고 청자는 화자를 매력적이고 사교적인 존재로 인식하였다(Street et al., 1982). 또한 발화 속도가 빠를수록 청자는 화자를 더 똑똑하고 자신감 있는 이미지로, 효율적이고 전문성 있는 존재로 보았다. Mehrabian (1972)도 발화 속도가 빠를수록 설득력이 높음을 밝혀내었다. Labarbera et al. (1979) 역시 빠르게 말하는 화자가 더 설득력 있고, 청자는 이런 화자를 상대적으로 더 지적이고 객관적이며 전문적이라고 인식하였다. 목소리로 말할수록, 여성의 경우는 좀 더 빠른 발화 속도로 말을 할수록 긍정적인 반응을 보였다. 다른 기존 연구에서도 빠르게 말하는

사람이 비교적 지성적이고 객관적이며 지적 수준이 높다고 인식되며 외향적이고 권위적이라고 여겨진다고 말한 바 있다(조일환, 2005). 이에 따라 본 연구는 AIVA의 발화 속도가 사용자의 호감도 반응에 영향을 미치는지 알아보고자 하였다. 발화 속도는 발화의 시작부터 끝까지의 시간(duration)을 초(sec)로 계산한 것으로 정의하였다(권순복, 2015). 기존 논문들을 참고하여 본 연구는 연구 문제를 아래와 같이 설정하였다.

연구문제 3: AI 음성 에이전트의 발화 속도는 사용자의 AI 음성 에이전트에 대한 호감도에 어떤 영향을 미치는가?

마지막으로 앞서 언급한 세 가지 요소 간에 상호작용이 나타나는지도 추가로 살펴보고자 한다. 하보미 등 (2018)의 경우 음도, 발화 속도, 그리고 강도 간 상호작용을 본 연구를 진행하였고 각 변인 간 상호작용에 따라 다른 결과를 보였다. 앞서서도 언급했듯이, 음도 요소는 성별과 함께 상호작용 효과가 드러났고, 여성의 경우는 높은음도, 남성의 경우는 중저음의 목소리 높낮이에서 청자의 긍정적인 반응이 나타났다. 발화 속도 역시 사람 목소리 연구에서는 성별과 상호작용 효과가 나타났다. Kwon (2016)은 남성의 경우 평소보다 느린 발화 속도로 말하는 것이 선호되며 여성의 경우 평소보다 빠르게 말하는 것이 선호되는 것으로 나타났다. 이를 바탕으로 본 연구는 AI 음성도 성별, 음도, 그리고 발화 속도 간 상호작용 효과가 나타나는지 알아보려고 했고, 아래와 같다.

연구문제 4: 목소리 구성 요소(성별, 음도, 발화 속도) 간에 상호작용 효과가 있는가?

Ⅲ. 연구방법론

3.1 파일럿 테스트

본 실험 전 실험도구를 제작하고 검증하기 위해 파일럿 테스트를 실시하였다. 본 연구에서 음도는 헤르츠(Hz), 발화 속도는 단위시간(초)당 음절(개수)로 단위로 설정하였다. 음성 메시지는 AIVA에서 가장 일반적인 정보서비스인 날씨정보로 구성되었다. 음성 샘플 조사를 위해 Praat과 Adobe Premier가 사용되었고, 음성 샘플은 네이버 파파고 목소리를 바탕으로 하였다. 남과 여 2개의 기본 음성 샘플을 제작하였고, 두 샘플 음성의 발화 속도는 5.68(403자/71초)로, 음도의 경우 여성은 245.752Hz, 남성은 122.474Hz이었다. 이렇게 제작된 음성 샘플에 대해 음도의 높음과 낮음 2가지, 발화 속도의 빠름과 느림 2가지, 성별 여성과 남성까지 고려해 총 8개의 샘플로 조작하여 실험에 사용되었다. 음성 조작에 있어 기본 음성은 음도 15Hz 단위로, 발화 속도는 0.1% 단위로 조절되었다. 조작의 기준은 음도와 발화 속도 차이는 변화를 주되 듣기에 부자연스럽지 않을 정도로 임의적 조절을 하였다. 실험은 실험 대상자들에게 AI 음성을 듣는다는 가정하에 71초 길이의 날씨정보를 듣고 호감도에 대한 질문에 응답했다. 온라인 설문 실험조사를 통해 진행되었고, 실험 대상자들은 8개의 상황에 무작위 배정되었다.

총 109명의 응답자가 참여한 실험에서 음성

샘플에 대한 조작 점검 검정 결과는 다음과 같다. ‘AI 목소리 발화 속도는 빠른 것 같다’라는 질문에 대해 7점 척도로 응답한 결과, 빠른 발화 속도 샘플(M=2.96, SD=1.044)과 느린 샘플(M=2.13, SD=0.941) 사이에 유의미한 차이를 보였다($t=-4.363, p=0.000$). 다음으로 ‘AI 목소리가 높다고 생각한다’란 질문에 높은 음도 샘플의 경우(M=2.93, SD=1.215) 낮은음도 샘플(M=2.41, SD=1.125)에 비해 높다고 평가되었다($t=-2.317, p=0.022$). 마지막으로 성별에 대한 조작 점검은 ‘AI 목소리는 여성의 목소리이다’이란 질문에 대해 여성 음성 샘플(M=4.76, SD=0.751)은 남성 음성 샘플(M=1.19, SD=0.397)과 구분되는 것으로 나타났다($t= 25.26, p=0.000$).

3.2 실험 설계 및 변인 측정

본 연구는 사람 목소리 연구에서 중요하게 여겨지는 준 언어적 요소 중 음도, 발화 속도가 AIVA 목소리에 적용되면 사용자 호감도에 어떤 영향을 미치는지 알아보았다. 사용자 호감도 반응 측정 방법에 있어서는 기존 연구(최지현 등, 2016; Gefen et al., 2004)를 참고하였다. 파일럿 테스트 결과를 바탕으로 본 실험을 설계하였다. 음성 샘플은 네이버 파파고 목소리를 바탕으로 하였는데, 조작 변수는 파일럿 테스트와 마찬가지로 3가지(성별, 음도, 발화 속도) 요소가 고려되었다. 결과적으로 총 8가지 음성 샘플에 대해 사용자 호감도를 측정하였다. 본 실험은 설문 조사회사 마크로밀엠브레인에 의해 2021년 5월 27일부터 5월 31일 동안 온라인으로 진행되었다. 실험 참가자로는 총 516명으로

국내에 거주하는 성인남녀를 대상으로 진행되었고, 연령은 10대부터 50대까지로 평균 연령은 35.44세였다(SD=13.234). 성별 분포는 남성 255명(49.4%), 여성이 261명(50.6%)으로 나타났다.

AIVA 목소리 샘플은 파일럿 테스트에서 검증된 것을 사용하였다. 유형은 성별(여성, 남성)에 따라 2가지 경우, 음도(높음, 낮음)에 따라 2가지 경우, 그리고 발화 속도(빠름, 느림) 2가지를 고려하여 총 8개의 상황을 설정하였다(표 1). 실험 대상자들에게 낯선 정보를 제공하는 AIVA 목소리를 듣게 한 뒤 온라인 설문조사를 시행하도록 하였다. 성별, 음도, 발화 속도에 따른 8개의 실험 집단에 실험 대상자들은 무작위 배정하였다.

<표1> 실험 처치물 구성표

구분	성별	음도	발화 속도
샘플1	여성	높음	빠름
샘플2	여성	높음	느림
샘플3	여성	낮음	빠름
샘플4	여성	낮음	느림
샘플5	남성	높음	빠름
샘플6	남성	높음	느림
샘플7	남성	낮음	빠름
샘플8	남성	낮음	느림

사람 목소리에 대한 기존 연구에서 호감도는 청자의 반응 측정 요소로 이용되어왔다. 황보명 등 (2015) 연구에서는 목소리에 대한 매력도를 측정하고자 하였고, (명료성, 신뢰도, 자상함, 안정감, 부드러움, 상냥함, 쾌활함, 친절함) 등과 같은 하위요소를 측정 문항에 고려하였다. 권순복(2015)은 호감도 파악을 위해 요인을 8가지(명료성, 신뢰도, 자상함, 안정감, 부드러움, 상냥함, 쾌활함, 친절함)로 구분하였다. 이

준상 등 (2021)의 연구는 설문에서 호감도에 대한 문항에 “좋아한다(I like this product)”, “호감이 간다(I am in favor of this product)”, “긍정적이다(I am positive about this product)”, “좋다(I think this product is good)” 등을 항목으로 포함하였다.

AI 음성 에이전트 및 로봇에 대한 기존 연구에서 측정 요소로 사용된 호감도는 로봇의 상호작용, 사람 유사성 정도, 인지된 안전성, 매력도 등으로 설명되었다. Bartneck et al.(2009)는 로봇과의 상호작용이 호감도, 의인화, 활동성 등으로 구분하여 측정 문항을 설정하였다. 사용자의 로봇 및 미디어에 대한 태도를 호감도로 측정하고자 하였고, 사회적 매력도, 외형적 매력도를 구성 요소로 둔 선행연구도 있었다(Lee, et al., 2006; Tung, 2016). 인지된 사람 유사성, 매력도 등이 AIVA에 대한 호감으로 적용된 경우도 있었다(Ho et al., 2010). 대화를 하고 싶은 마음, 즐거움, 사람 유사성, 의인화, 만족스러움 등이 매력도의 특성으로 본 연구도 존재하였다(김상현 등, 2002; 김동환 등, 2019). 이처럼 매력도 및 의인화 등은 AIVA 및 로봇, 미디어 기술 등에 대한 반응 연구에서 측정 요소로 자주 사용되어 왔다.

본 연구는 AI 음성 에이전트의 목소리 특성에 대한 사용자의 호감도를 탐구하고자 하였다. 따라서 본 연구는 기존 목소리 연구 및 AI와 로봇에서 활용된 반응 측정 요소를 참고하여 실험 측정 문항 설계에 적용하였다. 본 연구는 호감도의 하위요소로 매력도, 신뢰도, 의인화 3가지를 두었다. 매력도는 감정적인 반응(Affective reaction)으로 AI 음성 에이전트 목소리에 대한 사용자의 감정적 차원의 평가를 보는 데 사용

되었다. 신뢰도는 인지된 반응(Cognitive reaction)으로 기존 사람 목소리 연구에서도 평가 요소로 자주 활용되어 왔다. 본 연구는 감정적인 반응 요소인 매력도와 구분하여 신뢰도를 인지된 차원의 평가로 보았다. 마지막으로 의인화는 AI 관련 연구에서 ‘의인화’, 혹은 ‘인간 유사성’으로 언급되며 AI에 대한 호감도를 측정하는 데 활용되어왔다. 호감도 측정 요소별로 설문지 문항을 설계하였고, 7점 리커트 척도를 이용하였다. 본 연구의 AIVA에 대한 감정적 및 인지적, 의인화 평가방식은 서비스에 대해 다면적 평가를 도입하였던 기존의 연구와 맥을 같이 한다(김진솔 등, 2020).

선행 연구를 바탕으로 본 연구는 사용자 호감도 측정을 위한 항목을 구상하였다. 호감도에는 AI 음성 에이전트에 대한 감정적 평가, 인지적 평가, 의인화 평가를 확인하고자 3가지 요소(신뢰도, 매력도, 의인화)에 따른 세부 문항을 설계하였다. 본 연구는 선행연구를 바탕으로 본 연구는 사용자 호감도 인식을 파악하기 위한 항목으로 선행연구에서 잦은 빈도로 언급된 요소들 중 대표적 3가지 요소를 호감도 측정 항목으로 구성하였다(권순복, 2015). 각 변인에 대한 측정 항목은 <표 2>와 같고, Cronbach's Alpha값을 통해 신뢰도를 확인하였다.

사용자 호감도 척도 문항 12개의 구성 개념 타당도 확인을 위해 탐색적 요인분석을 실시하였다. 표본 적합도(MSA)는 0.925로 본 자료는 요인분석에 적합하다고 볼 수 있다. Bartlett의 구형성 검정 결과는 $x^2=7051.56$, $p=0.000$ 으로 ‘호감도 척도’ 변수 간의 상관성이 인정되었다. 총 3개의 하위요인이 추출되었고, ‘호감도 척도’의 요인1은 ‘매력도’, 요인2는 ‘신뢰도’, 요

<표 2> AI 음성 에이전트 목소리에 대한 사용자 호감도 측정을 위한 문항

요인	측정항목	회전된 성분행렬	Cronbach's α
매력도	본 AI 목소리를 듣고 좋은 감정이 든다.	0.803	0.942
	본 AI 목소리는 호감 있게 느껴진다.	0.850	
	본 AI 목소리는 매력적으로 느껴진다.	0.823	
	본 AI 목소리가 마음에 든다.	0.775	
신뢰도	본 AI 목소리는 안정감 있게 느껴진다.	0.609	0.916
	본 AI 목소리는 설명력 있게 들린다.	0.788	
	본 AI 목소리는 믿음직스럽게 느껴진다.	0.813	
	본 AI 목소리는 논리적으로 들린다.	0.805	
의인화	본 AI 목소리는 신뢰감 있게 들린다.	0.836	0.921
	본 AI 목소리는 실감나게 들린다.	0.638	
	본 AI 목소리는 실제 목소리 같다.	0.895	
	본 AI 목소리는 사람처럼 느껴진다.	0.911	
	본 AI 목소리와 대화하고 싶다.	0.640	
	본 AI 목소리와 감정적 유대가 느껴진다.	0.796	

인3은 ‘사회적 실재감’으로 명명하였다. 요인분석 결과 설명된 총 분산에서 초기 고윳값의 합계가 모두 1 이상으로 나타났고, 전체 3가지 요인이 전체분산의 79.96%를 설명하였다.

IV. 결과

4.1 조작 점검

본 연구의 실험 측정을 위한 목소리 특성 요소(성별, 음도, 발화 속도)에 따른 음성 보기 조작이 피험자들에게 잘 이뤄졌는지 확인하기 위해 t-test 분석을 시행하였다. 결과적으로 성별, 발화 속도, 음도는 모두 유의한 평균 차이 값을 나타내었다. 성별에 대한 조작점검은 각 집단의 평균이 높을수록 여성의 목소리에 가깝다고 인식함을 의미하는데, 여성 목소리(M=6.59, SD=0.607)가 남성 목소리(M=1.21, SD=0.568)보다 높은 평균값을 보여주었다(t=10.909, p=0.000). 발화 속도 간의 조작점검은 평균이

높을수록 발화 속도가 빠르다고 인식함을 의미하는데, 빠른 발화 속도 집단(M=4.65, SD=1.004)은 느린 발화 속도 집단(M=3.68, SD=1.093)보다 높은 평균값을 보였다(t=10.437, p=0.000). 마지막으로 음도 간의 조작점검은 평균 점수가 높을수록 피험자가 목소리 톤을 높게 인식하는 것을 의미하는데, 높은 음도 집단(M=4.03, SD=1.158)이 낮은 음도의 집단 평균값(M=3.83, SD=0.996)보다 높게 측정되었다(t=12.201, p=0.028). 따라서 성별, 속도, 음도에 대한 실험 조작 점검이 확인되었음을 알 수 있다.

4.2 실험 결과 분석

본 연구는 AIVA의 목소리 특성에 따른 사용자의 반응을 확인하고자 하였다. 연구 문제 1, 2, 3에서는 각 AIVA 목소리 특성과 사용자 호감도 사이 어떤 영향이 있는지 보고자 하였고, 연구 문제에 대한 검증을 위해 t-test 검정을 실시하였다.

우선, 각 목소리 특성 요소 중 음도에 따라

사용자의 호감도 하위 변수에 대한 차이가 있는지 알아보았다. 분석 결과 음도에 따른 매력도($t=-0.448, p=0.654$), 신뢰도($t=-0.412, p=0.680$), 의인화($t=0.179, p=0.858$) 차이는 통계적으로 유의하지 않았다. 다음으로 발화 속도에 따른 사용자의 호감도의 차이를 분석하였는데, 매력도는 통계적으로 유의미한 차이가 있게 나타났다($t=-2.946, p=0.003$). 느린 발화 속도의 평균값은 4.00, 빠른 발화 속도의 평균값은 4.31로 속도가 빠른 음성이 느린 음성보다 상대적으로 매력도가 높게 나타났다. 반면, 발화 속도에 따른 신뢰도($t=-1.754, p=0.080$)와 의인화($t=-1.699, p=0.090$)는 통계적으로 유의미한 차이가 없었다. 성별에 따른 매력도는 유의미한

차이가 나타났는데($t=6.682, p=0.000$), 여성 목소리의 평균값은 4.49, 남성 목소리는 3.81로 여성 목소리에 대해 매력도를 높게 평가하였다. 목소리 성별에 따른 신뢰도 역시 유의미한 차이가 있었는데($t=4.385, p=0.000$), 여성 목소리(4.67)의 신뢰도 값이 남성(4.26)보다 높았다. 하지만 목소리 성별에 따른 의인화는 유의미한 차이가 없었다($t=1.759, p=0.079$). 정리하면 성별은 매력도와 신뢰도에 영향을 주며 여성의 목소리가 남성에 긍정적으로 평가된 것을 알 수 있었다. 반면 성별에 따른 의인화는 통계적으로 유의한 영향을 미치지 않았다. 분석결과에 대한 요약은 <표 3>에 제시되어 있다.

<표 3> 각 목소리 특성 요소에 대한 호감도

구 분		의인화			t(p)
		N	평균(M)	표준편차(SD)	
음도	낮은음도	258	3.69	1.366	0.179(0.858)
	높은음도	258	3.67	1.289	
발화 속도	느린 발화 속도	261	3.59	1.390	-1.699(0.090)
	빠른 발화 속도	255	3.78	1.253	
성별	남성	261	3.78	1.331	1.759(0.079)
	여성	255	3.58	1.316	

구 분		매력도			t(p)
		N	평균(M)	표준편차(SD)	
음도	낮은음도	258	4.13	1.226	-0.448 (0.654)
	높은음도	258	4.18	1.170	
발화 속도	느린 발화 속도	261	4.00	1.264	-2.946(0.003)**
	빠른 발화 속도	255	4.31	1.105	
성별	여성	261	4.49	1.094	6.682(0.000)***
	남성	255	3.81	1.204	

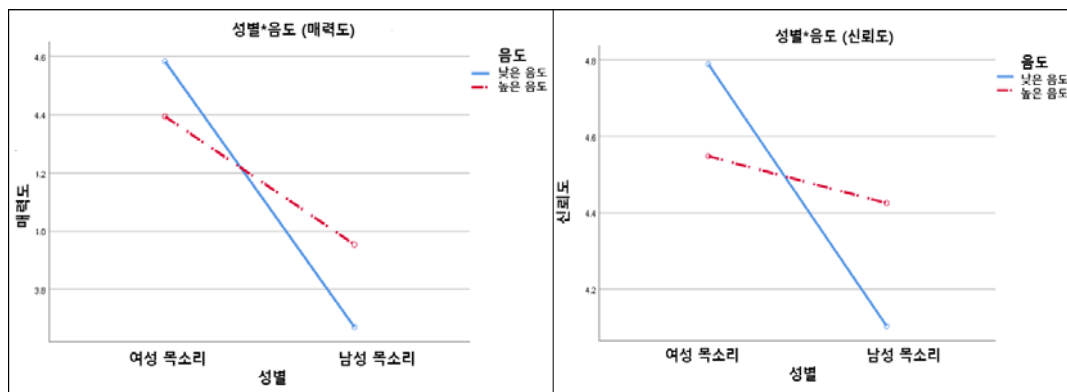
구 분		신뢰도			t(p)
		N	평균(M)	표준편차(SD)	
음도	낮은음도	258	4.45	1.083	-0.412(0.680)
	높은음도	258	4.49	1.051	
발화 속도	느린 발화 속도	261	4.39	1.111	-1.754(0.080)
	빠른 발화 속도	255	4.55	1.014	
성별	여성	261	4.67	.993	4.385(0.000)***
	남성	255	4.26	1.102	

* $p<.05$, ** $p<.01$, *** $p<.001$

다음으로는 연구 문제 4인 AIVA의 목소리 특성 요소 간의 상호작용 효과가 있는지를 확인하였다. 검증을 위해 본 연구는 ANOVA 분석을 시행하였고, 각 목소리 특성 요소 간의 상호작용 효과를 보고자 이원 분산분석(two-way ANOVA)을 실시하였다. 첫째로 매력도를 종속 변인으로 두고 음도와 발화 속도 간의 상호작용 효과 검정을 시행하였다. 음도와 발화 속도 간(음도* 발화 속도) 간의 개체 간 효과 검정 결과는 $F(1, 512) = 0.640, p = 0.424$ 로 유의한 상호작용을 보이지 않았다. 성별과 음도 간(성별*음도)의 상호작용 효과 검정 결과는 $F(1, 512) = 5.503, p = 0.019$ 로 통계적으로 유의한 결과가 도출되었다. 따라서 성별과 음도 간의 상호작용 효과가 있음을 파악할 수 있다. 여성 목소리는 상대적으로 높은 음도 보다 낮은 음도에서 매력도가 높게 나타났고, 남성의 목소리는 상대적으로 낮은 음도 보다 높은 음도에서 높은 매력도가 나타났다. 반면 성별과 발화 속도 간의 상호작용 효과 검정 결과는 $F(1, 512) = 1.080, p = 0.299$ 로 통계적으로 유의한 값을 보여주지 못하였다. 즉, 음도와 발화 속도(음도*발화 속도)와 성별과 발화 속도(성별*발화 속도)에서

는 상호작용 효과가 드러나지 않았다. 이는 성별에 대한 매력도 측정값이 발화 속도에 따라 달라지지 않음을 의미한다. 정리하자면 성별과 음도 간 상호작용 효과가 매력도에 영향을 미치며 여성의 목소리는 비교적 낮은 음도, 남성의 목소리는 비교적 높은 음도에서 매력도가 높게 나타났다.

둘째, 신뢰도를 종속 변인으로 설정하고 각 목소리 특성 요소 간 상호작용 효과를 분석한 결과이다. 먼저 음도와 발화 속도 간의 개체 간 효과 검정 결과는 $F(1, 512) = 0.108, p = 0.742$ 로 유의한 상호작용 효과가 나타나지 않았다. 다음으로 성별과 음도 간(성별*음도)의 상호작용 효과 검정 결과에서는 유의한 값이 도출되었다 ($F(1, 512) = 9.450, p = 0.002$). 여성 목소리는 낮은 음도에서 신뢰도가 높게 측정되었다. 반면, 남성의 목소리는 상대적으로 낮은 음도 보다 높은 음도에서 신뢰도 결과가 높은 것으로 나타났다. 마지막으로 성별과 발화 속도 간의 상호작용 효과 결과는 $F(1, 512) = 3.357, p = 0.067$ 로 통계적으로 유의하지 않았다. 신뢰도 평가에 있어서는 성별과 음도 간 상호작용 효과가 유의한 영향을 미치며, 여성은 낮은 음도일 때, 남



<그림 1> 변수 간 상호작용 효과

성은 비교적 높은 음도일 때 신뢰도를 높인다고 볼 수 있다.

마지막으로 의인화를 종속 변인으로 목소리 특성 요소 간 상호작용 효과를 검정한 결과이다. 음도와 발화 속도 간의 상호작용($F(1, 512)=1.045, p=0.307$), 성별과 음도 간 상호작용($F(1, 512)=2.627, p=0.106$), 성별과 발화 속도 간 상호작용($F(1, 512)=0.097, p=0.756$) 모두 통계적으로 유의하지 않았다.

V. 토의 및 향후 연구

5.1 결과 토의

AIVA는 다양한 서비스를 제공할 뿐 아니라 사용자의 친구이자 비서, 혹은 동료로서 사회적 기능을 수행하고 있다. 이에 따라 사용자와의 의사소통의 주요한 수단인 AI 목소리에 대한 중요성은 점차 커지고 있다. 사람 목소리 특성에 대한 연구는 많이 진행되어 왔으나 AIVA 목소리에 대한 연구는 아직 초기단계이다. 이에 본 연구는 기존 사람 목소리 연구를 AIVA 목소리 연구에 적용하고자 하였다. AI 목소리의 3가지 특성(성별, 음도, 발화 속도)에 따른 AI 목소리에 대한 사용자의 호감도(매력도, 신뢰도, 의인화)를 알아보고, 각 목소리 특성별 상호작용 효과도 탐색하였다.

성별은 매력도와 신뢰도에서 유의한 값을 보였는데, 구체적으로 여성 목소리가 남성의 목소리보다 매력도와 신뢰도에 긍정적인 영향을 미쳤다. 한편 성별은 사용자의 AIVA에 대한 의인화에 별다른 영향을 끼치지 않았다. 이러한

결과를 기존 사람 목소리 연구 결과와 비교하면 다음과 같다. 목소리 연구 결과에서 제시되었던 여성의 목소리가 남성의 목소리에 비해 선호되며 신뢰감을 주는 경향이 있다는 부분은 본 연구 결과에서도 나타났다. 목소리에 대한 사람들의 고정관념은 AIVA 목소리에도 적용될 수 있음을 알 수 있다. 이는 성별에 따른 사회적 역할론(social-role theory)을 통해 설명될 수 있다. 보통 사람들은 성별에 따라 구분된 역할을 기대하는데, 여성의 경우는 주로 비서와 같은 종속적인 역할이 기대되어왔다. 즉, AIVA의 목소리 성별에 따라 사용자의 호감도 평가 차이가 나타난 것은 사용자들이 사회적 역할론에서 제시된 것처럼 성별에 따른 고정관념을 가지고 있기 때문이다. 사용자는 기술이나 서비스 이용에서도 성별에 따라 구분된 역할수행이 존재하며, 사용자는 비서처럼 서비스를 제공하는 AI 에이전트에겐 여성의 목소리가 적합하다고 느꼈을 것이다. 실제로 대부분의 AIVA 음성이 여성의 목소리를 기본으로 설정하고, 아마존과 마이크로소프트 등의 기업이 사람들을 대상으로 AI 목소리 선호도 조사를 한 결과 대다수 사람이 여성의 목소리를 선호하였다(고영태, 2019).

AIVA는 사용자에게 긍정적인 평가를 받기 위해 여성 음성을 기본 목소리로 하고 있다. 하지만 이는 성별에 대한 기대와 성 역할을 고착화하는 문제점을 가진다. 이런 점에서 본 연구는 여성 목소리를 선호하는 결과를 통해 여전히 전통적인 고정관념과 성 역할이 기대되고 있음을 짐작할 수 있었다. 사회 변화와 더불어 전통적인 성 고정관념이 완화되고, 성평등에 대한 인식이 강조되고 있지만, 스마트 서비스 영

역에서도 여전히 성별에 따른 편견 및 고정관념이 유지되고 있음을 확인할 수 있었다.

AIVA의 목소리 높낮이 여부는 사용자의 호감도에 영향을 주지 않았다. 기존 사람 목소리 연구에서 음도는 가장 빈도 높게 사용되었던 준 언어적 목소리 특성 요소(Kim et al., 2005; Smith et al., 2005; Jones et al., 2008)인 점을 고려하면 본 연구 결과는 주목할 만하다. 단순히 음도의 높고 낮음은 사용자의 호감도에 별 영향을 미치지 않았지만, 성별 요소가 함께 고려되었을 때, 유의미한 차이를 보였다. 기존 목소리 연구에서도 음도는 성별과 함께 고려됐고, 커뮤니케이션 영향에 유의한 차이를 보였다. 결국 AIVA의 목소리에서 음도는 독립적으로 사용자의 호감도에 영향을 주지 못함을 알 수 있었다. 사용자가 여성의 목소리를 남성보다 비교적 선호한다는 연구 결과를 참고해보면, 음도가 높더라도 사용자가 목소리에 대한 성별을 파악하지 못할 경우 음도 차이는 사용자의 호감도 변화와는 유의한 관계가 드러나지 않는 것으로 해석할 수 있다.

AIVA의 빠른 목소리는 느린 목소리보다 사용자에게 매력적으로 느껴지는 것으로 분석되었다. 사람 목소리 연구 결과와 비교해보면, 사람 목소리의 경우는 발화 속도가 빠를수록 청자는 화자에게 매력과 사교성을 높게 평가한다는 결과(Street et al., 1982)와 비슷한 맥락이다. 반면, AIVA의 발화 속도에서는 사용자의 의인화 및 신뢰도에 유의한 영향을 볼 수 없었다. 이는 사람 목소리 연구 결과인 말하는 속도가 빠를수록 의사소통의 효율성, 설득력 및 전문성을 갖는 것으로 인식된다(Mehrabian, 1972)는 결과와 차이가 있었다. 즉, 사람 목소리의 경우

목소리가 상대적으로 빠른 것이 신뢰도를 높이나, AIVA의 경우에는 별다른 영향을 주지 못하는 것으로 파악된다. 감정적 차원의 평가인 매력도와는 달리 인지적 차원의 평가인 신뢰도의 경우, 사용자가 AIVA와 상호작용을 하며 순간적이고 본능적으로 긍정적인 감정을 느꼈을지라도(매력도 평가), 사람과 다른 점을 발견하거나, 로봇 같은 목소리 특성을 인식했을 때 사용자는 불편함을 느끼거나 이질감을 느끼며 에이전트에 대한 신뢰를 낮게 평가할 수 있다. 신뢰도 상호작용이 원활히 이뤄질 때 발생하며 언캐니 밸리와 같은 현상은 신뢰도에 결정적인 영향을 미칠 수 있다(Jung et al., 2021). 따라서 AIVA의 음성을 듣고 매력을 감정적으로 느꼈을지라도, 사람과 다른 특성이 발견되면 신뢰와 같은 인지적 차원의 반응에서 사용자는 유의한 영향을 받지 않을 수 있다.

한편, 사람의 목소리의 경우 발화 속도가 빠를수록 호감을 얻는다는 기존 연구가 있으나, 발화 속도는 빠르되 정확한 ‘쉽’과 ‘발음’이 전제될 때 긍정적인 반응이 나타났다(권순복, 2015). AIVA가 사회적 존재이며 의인화 대상으로 여겨지고 있지만, 사실상 현재 수준의 AIVA는 상황에 따라 혹은 문장 길이에 따라 유동적으로 쉽이나 발음 조절에서 부자연스러운 부분이 드러날 수 있다. AI 음성과 사람과의 대화가 더 자연스러워지도록 AI 알고리즘이 개발된다면, 발화 속도 역시 사용자 호감도에 유의한 영향을 미칠 수 있을 것이다. 이를 위해 본 연구에 사용된 목소리 요소뿐 아니라 다양한 목소리 요소를 고려하여 향후 검증할 필요가 있다.

목소리 특성 간 상호작용을 보면, 음도와 발

화 속도 간의 상호작용 효과와 성별과 발화 속도 간 상호작용 효과는 매력도에서 유의한 값을 보이지 않았다. 반면, 성별과 음도 간 상호작용 효과에서는 성별에 대한 매력도는 음도에 따라 차이가 있음이 밝혀졌다. 구체적으로 여성 목소리는 상대적으로 높은 음도 보다 낮은 음도에서, 남성의 목소리는 낮은 음도 보다 높은 음도에서 높은 매력도가 나타났다. 즉, 음도에 대한 매력도 값과 성별에 대한 매력도 값이 발화 속도에 따라 달라지지 않음을 알 수 있었다. 신뢰도를 종속 변인으로 두고 각 목소리 특성 요소 간 상호작용 효과를 본 결과는 성별과 음도 간의 상호작용에서 유의한 값이 도출되었다. 다시 말해, 성별에 대한 신뢰도 값은 음도에 따라 차이가 나타났다. 여성 목소리는 낮은 음도에서 신뢰도가 높게, 남성의 목소리는 상대적으로 높은 음도에서 높은 신뢰도를 보였다. 반면 3개의 목소리 특성 요소는 의인화 정도에 상호작용 효과를 주지 못하는 것으로 볼 수 있다.

기존 사람 목소리 연구 결과를 보면, 성별과 발화 속도 간의 호감도에서 유의한 결과가 나왔다는 점에서 비교가 가능하다. 발화 속도의 경우 남성은 조금 느린 목소리로 말할수록, 여성은 좀 더 빠른 발화 속도로 말할수록 긍정적인 반응을 보였다(Kwon, 2016). 하지만 본 연구에서는 발화 속도와 성별 간의 상호작용에서 유의한 효과를 볼 수 없었다는 점에서 사람 목소리에 대한 청자 반응 연구와 AIVA의 목소리에 대한 사용자 연구에는 차이가 있었다. 성별과 목소리 음도에 대한 상호작용 효과는 기존 사람 목소리 연구에서도 다뤄졌다. 기존 사람 목소리에서는 남성의 경우 보통의 남성 음도 보다 중저음의 목소리가 신뢰감과 안정감을 주

고 선호되는 경향이 있고, 여성의 경우는 낮은 음도 보다 높은 음도가 청자에게 호감을 느낀다고 하였다(Jones et al., 2008; Kwon, 2016). AIVA의 목소리는 여성 목소리일 때 비교적 낮은 음도에서, 남성 목소리는 상대적으로 높은 음도에서 신뢰도와 매력도가 높게 나타났다. 즉, 사람 목소리와 AIVA 목소리는 신뢰도에서 상반된 결과를 보였다.

또한, 본 연구는 AIVA에 대한 실무적인 시사점에 제공한다. 연구결과는 AIVA 생산자나 마케터가 AIVA의 최적화된 목소리에 대한 관심을 가질 필요가 있음을 보여준다. 현재까지는 성별에 따른 선호도에 기반하여 여성 목소리를 기본 목소리로 제공하고, 남성 목소리 혹은 연예인의 목소리로 변경가능한 옵션을 제공하는 수준이다. 연구결과가 보여주듯이, 다양한 목소리 특징이 사용자의 반응에 유의미한 영향을 미칠 수 있음에도, 상용화된 AIVA의 목소리에 대한 선택은 제한적이다. AIVA 생산자나 마케터는 AIVA 목소리에 대한 사용자 반응에 대한 엄밀한 실험을 통해 AIVA 목소리에 대한 사용자 선호를 파악하고, 이를 반영할 필요가 있다. 나아가, 고도의 AI기술을 접목하여 제공하는 정보의 종류나 사용상황, 사용자의 특징을 고려하여 AIVA가 상황에 최적화된 목소리를 제공하는 것도 실무적으로 고려해 볼 필요가 있다.

5.2 한계점 및 향후 연구

본 연구 결과는 다음과 같은 한계점을 가지고 있다. 본 연구는 성별 요소에서 현재 관심이 증가하고 있는 중성 목소리를 두고 실험 연구를 하지 못했다는 한계를 가진다. 실제로 AIVA

목소리는 중성이 적합하다는 주장도 있을 뿐만 아니라(고영태, 2019), 최근 Virtue Nordic 연구팀은 실제 사람들의 목소리를 기반으로 “Q”라는 중성의 목소리를 개발하였다. 본 연구의 결과가 음도가 낮은 여성 목소리일 때와 음도가 높은 남성 목소리일 때 호감도가 높게 나타난 것은 사용자가 중성에 가까운 목소리에 호감도를 보인 것이라고 볼 수 있다. 따라서, 향후 연구에서는 사용자의 AIVA의 중성 목소리에 대한 반응을 면밀히 살펴볼 필요가 있다.

둘째, 본 연구는 다양한 호감도 측정 하위요소 중 대표적인 요소인 3가지 요소만을 가지고 실험을 진행하였다. 기존 목소리 연구에서는 화자 목소리 특성에 대한 청자 반응 측정 요소로 명료성, 신뢰도, 자상함, 안정감, 부드러움, 상냥함, 쾌활함, 친절함(황보명 등, 2015; 권순복, 2015) 등을 제시하였다. 또한 AIVA 및 로봇에 대한 호감도 측정 요소로는 상호작용, 사람 유사성 정도, 인지된 안전성, 매력도, 호감도, 의인화, 활동성 등이 언급되기도 하였다. 또한 매력도는 구체적으로 사회적 매력도, 외형적 매력도로 구분할 수 있으며 이를 구분하여 진행한 선행연구들도 존재하였다(Lee et al., 2006; Tung, 2016) 독립변수 역시 마찬가지로 목소리의 준 언어적 특성 요소로 강도, 발음, 발화 속도, 목소리 크기, 울림, 어조, 억양 강세, 쉼, 음질(오석영, 1997; 이도영, 2001) 등이 다양한 요소들이 존재한다. 보다 폭넓은 연구 결과를 도출하기 위해 기존 독립변수와 종속변수로 사용된 요소보다 많은 변수항목 들을 가지고 폭넓게 연구 분석할 것을 제안한다.

셋째, 본 연구는 AIVA의 외형적인 형태를 고려하지 않고 목소리 특성에 집중하여 사회적

실재감에 대한 반응을 밝혀내지 못했다. 기존의 연구들은 주로 AIVA의 외형이나 이미지에 초점을 두고 의인화 현상을 다뤘었다. 특히, 본 연구는 사용자와 인터랙티브한 대화 과정을 실험 설계로 두지 않았다는 점에서 사용자가 AIVA에 대해 의인화를 느끼기에 부족한 점이 있었다. 사용자가 일방적으로 정보를 제공받는 시나리오로 실험이 이뤄졌기 때문에 사용자가 AI 에이전트의 목소리만을 듣고 에이전트에 대해 사람과 비슷함, 원활히 상호작용을 하고 의인화 현상을 느끼기에는 한계가 있었다. 목소리 외에 AIVA의 다른 특성(외형)과 인터랙티브한 상황을 포함한 실험 설계로 본 연구결과를 재검증할 필요가 있다.

마지막으로 실험 설계에 있어 시나리오의 부족과 목소리 종류가 적다는 한계가 있었다. 본 연구는 날씨 정보를 전달하는 정보제공 중심의 메시지만이 실험에 사용되었다. AIVA는 다양한 서비스를 수행하며 어떤 역할을 하느냐에 따라 사용자는 AIVA에 대해 다른 반응을 보일 수 있다. 또한 에이전트가 제공하는 메시지의 방식이 단순한 정보 중심 성격이 보이는지, 흥미성을 띠는 가벼운 주제 위주인지에 사용자가 AIVA에 대해 느끼는 반응 또한 다를 것이다. 다시 말해 사용자는 목소리 준 언어적 차원 특성의 차이 외에도 메시지의 성격이나 내용에 의해 AIVA에 대한 다른 반응을 보일 가능성이 있다. 실제로 Ahn et al.(2022)의 연구에 따르면 AI가 상품을 추천할 때 실용적 특징이 두드러진 상품의 경우에는 남성 목소리를, 즐거움에 관련된 상품을 추천할 때는 여성 목소리를 선호한다는 것을 밝혀냈다. 따라서 향후 연구에서는 AIVA가 제공하는 정보 혹은 서비스의 특징

을 포함하여 목소리 특징에 따른 사용자 반응을 탐색할 필요가 있다.

참고문헌

- 강길호, 김현주, “커뮤니케이션과 사람,” 서울: 한나래, 1995.
- 고영태, “너는 아니? 중성 목소리... AI가 여성 목소리인 이유,” Mar. 2019, <KBS News>, <https://news.kbs.co.kr/news/view.do?ncd=4191793>.
- 구본권, “[ESC]왜 음성비서는 여성 목소리일까,” April 2019, <한겨레>. Retrived from http://www.hani.co.kr/arti/specialsection/esc_section/889562.html
- 권순복, “준 언어적 구성 요소를 통한 매력적인 목소리 분석과 호감도에 관한 실험 연구,” 언어치료연구, 제24권, 제1호, 2015, pp. 157-167.
- 김광모, 최희원, 권송일, “사회적 실재감이 온라인 커뮤니티 지속사용의도에 미치는 영향,” 한국콘텐츠학회논문지, 제14권, 제2호, 2014, pp. 131-145.
- 김동환, 최상욱, “소셜 로봇의 자기의식 표현이 로봇의 호감도에 미치는 효과 연구,” 정보과학회논문지 제46권 제7호, 2019, pp. 653-663.
- 김상현, 오상현, “고객재구매의도, 결정요인에 관한 연구,” 마케팅연구, 제17권, 제2호, 2002, pp. 25-55.
- 김우룡, 장소원, “비언어적 커뮤니케이션론,” 서울: 나남출판, 2004.
- 김정현, 최준호, “대화형 에이전트의 추천 전략이 음성쇼핑경험에 미치는 영향에 관한 연구,” 사이버커뮤니케이션학보, 제35권, 제4호, 2018, pp. 5-35.
- 김진솔, 손새아, 김희웅, “운동에 영향을 끼치는 피트니스 앱 요인 연구,” 정보시스템연구, 제29권, 제4호, 2020, pp. 1~24.
- 박란희, 이시훈, “목소리 구성 요소의 커뮤니케이션 효과에 관한 연구,” 스피치와 커뮤니케이션, 제11권, 2009, pp. 293-327.
- 박종화, 김민성, 김정환, “언론은 인공지능(AI)을 어떻게 다루는가?: 뉴스 빅데이터를 통한 한국과 미국의 보도 경향 분석,” 정보시스템연구, 제31권, 제1호, 2022, pp. 175-195
- 오경아, “남성이 여성 음역대로 노래 부를 수 있을까,” <중앙일보>. Mar, 2013, Retrived from https://jhealthmedia.joins.com/article/article_view.asp?pno=8252.
- 오석영, “설교에 있어서 비언어적 커뮤니케이션의 효과성에 관한 연구,” 목원대학교 신학대학원 석사학위논문, 1997.
- 이슬기, 남영자, “인공지능 스피커 성별이 마음챙김 명상 사용자에게 미치는 영향,” 한국디지털콘텐츠학회 논문지, 제21권, 제9호, 2020, pp. 1645-1652.
- 이유나, 허경호, “발표상황에서 발표자의 비언어적 요소가 발표자의 이미지 평가 및 메시지 인지도에 미치는 영향,” 한국소통학보, 제10권, 2008, pp. 38-72.
- 이준상, 박준홍, “제품디자인 혁신성이 호감도와 구매 의도에 미치는 영향 -블루투스

- 스피커를 중심으로,” 한국정보통신학회논문지, 제25권, 제2호, 2021, pp. 228-233.
- 임종수, 최진호, 이혜민, “AI 미디어와 의인화: AI 음성 대화형 에이전트의 의인화 평가척도 개발 연구,” 한국언론학보, 제64권, 제4호, 2020, pp. 436-470.
- 최정건, 김명석, “로봇 디자인에서 의인화 기법의 활용 평가 방법에 관한 연구,” HCI2008 학술대회 발표논문집, 2008, pp. 976-980.
- 최지현, 조동욱, 정연만, “음성 분석을 이용한 청자가 호감을 느끼는 목소리에 대한 규명,” 한국통신학회논문지, 제41권, 제1호, 2016, pp. 122-131.
- 하보미, 허명진, “음도, 발화 속도, 강도가 화자의 말소리 인식에 미치는 영향,” 언어치료연구, 제27권, 제3호, 2018, pp. 45-54.
- 황보명, 한의진, “목소리로 연상되는 이미지에 관한 연구,” 언어치료연구, 제24권, 제4호, 2015, pp. 249-257.
- Amin, M. E., “Social Science Research: Conception, Methodology and Analysis,” Makerere University Press, Kampala, 2005.
- Aronovitch, C., “The Voice of Personality: Judgments and the Relation to Voice Quality and Sex of Speaker,” *Journal of Social Psychology*, Vol. 99, No. 2, 1976, pp. 207-220.
- Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S., “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots,” *International Journal of Social Robotics*, Vol. 1, 2009, pp. 71-81.
- Birdwhistell, R. L., “Communication without Words,” *L'aventure humaine*, Paris, Societd d'Etudes Litttraires et Artistiques, 1964.
- Breazeal, C., “Emotion and Sociable Humanoid Robots,” *International Journal of Human-Computer Studies*. Vol. 59, No. 1, 2003. pp. 119-155
- Collins, S., and Missing, C., “Vocal and Visual Attractiveness are Related in Women,” *Animal Behaviour*, Vol. 65, No, 5, 2003, pp. 997-998.
- Tung, F. W., “Child Perception of Humanoid Robot Appearance and Behavior,” *International Journal of Human-Computer Interaction*, Vol. 32, No. 6, 2016, pp. 493-502.
- Gefen, D., and Straub, D., “Consumer trust in B2C e-Commerce and the Importance of Social Presence: Experiments in e-Products and e-Services,” *Omega*. Vol. 32, 2004, pp. 407-424.
- Guzman, A. L., “Voices in and of the Machine: Source Orientation toward Mobile Virtual Assistants,” *Computers in Human Behavior*, Vol. 90, 2019, pp. 343 - 350.
- Lee., H. E., “Why do Voice Activated

- Technologies Sound Female?: Sound Technology and Gendered Voice of Digital Voice Assistants,” *Korean Journal of Communication & Information*, Vol. 90, 2018, pp. 126-153.
- Kong, H. S., Song, E. J., and Kang, M. S., “Consumer Interest Areas and Attention to Bluetooth Product Advertising Design,” *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 19, No. 10, 2015, pp. 2241- 2246.
- Haslam, N., and Loughnan, S., “Dehumanization and Infrahumanization,” *Annual Review of Psychology*, Vol. 65, 2014, pp. 399-423.
- Ho, C., and MacDorman, K., “Revisiting the Uncanny Valley Theory: Developing and Validating an Alternative to the Godspeed Indices,” *Computers in Human Behavior*, Vol. 26, No. 6, 2010, pp. 1508-1518.
- Johnson, C. D., Bauer, B. C., and Singh, N., “Exploring Flow in the Mobile Interface Context,” *Journal of Retailing Consumer Services*, Vol. 53, 2020, 101744
- Jones, B., Feinberg, D., Debruine, L., Little, A., and Vukovic, J., “Integrating Cues of Social Interest and Voice Pitch in Men’s Preferences for Women’s Voices,” *Biology Letters*, Vol. 2, No. 4, 2008, pp. 192-194.
- Jung Y., Cho, E., and Kim, S., “Users’ Affective and Cognitive Responses to Humanoid Robots in Different Expertise Service Contexts,” *Cyberpsychology, Behavior and Social Networking*, Vol. 24, No. 5, 2021, pp. 300-306.
- Furumo, K., and Pearson, J. M., “Gender-based Communication Styles, Trust, and Satisfaction in Virtual Teams,” *Journal of Information, Information Technology & Organizations*, Vol. 2, 2007, pp. 47-60.
- Kim, M., and Na, E., “Effects of Nonverbal Communication in National Assembly Candidates’ Broadcasted Speech on Viewers Depending on Involvement,” *Korean Journal of Broadcasting and Telecommunication Studies*, Vol. 19, No. 3, 2005, pp. 42-103.
- Kreijns, K., Kirschner, P. A., and Jochems, W., “Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: A Review of the Research,” *Computers in Human Behavior*, Vol. 19, No. 3, 2003, pp. 335 - 353.
- Kwon, S., “Characteristics of the Auditory Evaluation of Good Impression Using Speech Manipulation Scripts,” *Phonetics and Speech Sciences*, Vol. 8, No. 4, 2016, pp. 131-138.
- Labarbera, P., and Maclachlan, J.,

- “Time-Compressed Speech in Radio Advertising,” *Journal of Marketing*, Vol. 43, No. 1, 1979, pp. 30-36.
- Lee S., Ratan R., and Park T., “The Voice Makes the Car: Enhancing Autonomous Vehicle Perceptions and Adoption Intention through Voice Agent Gender and Style,” *Multimodal Technologies and Interaction*, Vol. 3, No. 20, 2019.
- McKnight D. H., “Trust in Information Technology,” In *The Blackwell Encyclopedia of Management, Management Information Systems*, Vol. 7 (pp. 329-331), Malden, MA: Blackwell, 2005.
- Mehrabian, A., *Nonverbal Communication*, Chicago: Aldine Antherton, 1972.
- Mehrabian, A., Williams, M., “Nonverbal Concomitants of Perceived and Intended Persuasiveness,” *Journal of Personality and Social Psychology*, Vol. 13, No. 1, 1969, pp.13, 37-58.
- Meet Q, Meet Q: The First Genderless Voice. 01, Mar. 2019, Available: <https://www.youtube.com/watch?v=jasElteA3Ag>
- Mori, M., “The Uncanny Valley. Energy,” *IEEE Spectrum*, Vol. 7, No. 4, 1970, pp. 33-35.
- Shechtman, N., and Horowitz, L. M., “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*, 2003.
- Street., R., Brady., R. “Speech Rate Acceptance Range as a Factor of Evaluative Domain, Listener Speech Rate and Communication Context.” *Communication Monograph*, Vol. 49, 1982, pp. 290-308.
- Gefen D, Karahanna E, and Straub D. W., “Trust and TAM in Online Shopping: An Integrated Model,” *MIS Quarterly*, Vol. 27, No. 1, 2003, pp. 51-90
- Smith, B., Brown, B., Strong, W., and Rencher, A., “Effects of Speech Rate on Personality Perception,” *Language and Speech*, Vol. 18, No. 2, 1975, pp. 145-152.
- S. Y. Kuh., “Behavioral Intentions of Luxury Brands: Structural Relations among the Influences Focusing on the Theory of Reasoned Action,” *Journal of Industrial Economics and Business*, Vol. 28, No. 6, 2015, pp. 2789-2815.
- Street, R., and Hopper, R., *A Model of Speech Style Valuation*, London: Edward Arnold, 1982.
- Jeong, Y., Park, D. E., Yoon, J., and Jang, M. K. “Exploring Effects of Dialect on User Perception of Conversational Agents,” *Journal of Digital Contents Society*, Vol. 20, No. 7, 2019, pp. 1439-1446.
- Vohs, K. D., Baumeister, R. F., Twenge, J. M., et al., “Making Choices Impairs

Subsequent Self-control: A Limited Resource Account of Decision Making, Self-regulation, and Active Initiative,” *Journal Personality and Social Psychology*, Vol. 94, No. 5, 2008, pp. 883-898.

Wang, W, and Benbasat I., “Trust in and Adoption of Online Recommendation Agents,” *Journal of the AIS*, Vol. 6, No. 3, 2005, DOI: 10.17705/1jais.00065

백 승 주 (Paek, Seung Ju)



고려대학교 역사학 학사와 미디어학 석사를 취득하였다. 현재 HS에드에 재직하고 있으며, 주요 관심분야는 스마트 미디어, 미디어 경영 등이다.

정 윤 혁 Jung, Yoon Hyuk)



현재 고려대학교 미디어학 부 교수로 재직하고 있으며, 주요 관심분야는 디지털 미디어 사용자, 미디어 경영, 가상 세계 등이다.

<Abstract>

AI Voice Agent and Users' Response

Beak, Seung Ju · Jung, Yoon Hyuk

Purpose

As artificial intelligence voice agents (AIVA) have been widely adopted in services, diverse forms of their voices, which are the main interface with users, have been experimented. The purpose of this study is to examine how users evaluate vocal characteristics (gender, voice pitch, and voice pace) of AIVA, depending on prior research on human voice attractiveness.

Design/methodology/approach

This study employed an experimental survey which 516 participated in. Each participant was randomly assigned into one of eight situations (e.g., male - higher pitch - faster pace) and listened a AIVA voice sample, which introduce weather information. Next, a participant answered three consequence factors (attractiveness, trust, and anthropomorphism).

Findings

The results reveal that female voices of AIVA were perceived as more attractive and trustworthy than male voices. As far as voice pitch goes, while lower-pitch voices were preferred in female voices, higher-pitch voices were preferred in male voices. Finally, faster voices of AIVA were more attractive than slower voices.

Keyword: Artificial Intelligence, Voice Agent, Voice Pitch, Voice Pace, Attractiveness, Trust, Anthropomorphism

* 이 논문은 2022년 1월 21일 접수, 2022년 3월 25일 1차 심사, 2022년 4월 22일 게재 확정되었습니다.