

대화 지속성 암묵적 단서를 고려한 차량 내 음성 인터랙션 구조 연구

남궁기찬
국민대학교 산학협력단 교수

A Study on the In-Vehicle Voice Interaction Structure Considering Implicit context with Persistence of Conversation

Kiechan Namkung
Professor, Industry Academic Cooperation Foundation, Kookmin University

요약 본 연구에서는 차량 내 음성 인터랙션을 사용하는 사용자의 대화 행태를 탐색적으로 살펴보았다. 본 연구의 목적은 시스템과의 음성 인터랙션에서 사용자들이 기대하는 대화 요소를 파악하여 사람 간의 대화와 유사한 음성 인터랙션을 가능하게 하기 위한 구조적 개선점을 제시하는 것이다. 사용자의 차량 내 음성 인터랙션 행태를 관찰하기 위해 맥락 질문법을 통해 자료를 수집하고 개방 코딩을 사용하여 인터뷰 내용을 분석하였다. 이를 통해 음성 인터랙션 기능의 유용성을 탐구할 수 있었으며, 이러한 유용성은 기능에 대한 사용자의 만족도와 사용 지속성을 증가시킨다는 점에서 매우 중요하다. 본 연구는 기술에 대한 사용자의 경험적 요구를 대인관계 모델인 대화의 관점에서 분석하였다는 점에서 의미가 있다고 할 수 있다.

주제어 : 대화, 암묵적 단서, 자동차 인터페이스, 지속성, 구조도, 음성 인터랙션

Abstract In this study, the conversation behavior of users is investigated by using in-vehicle voice interaction system. The purpose of this study is to identify the elements of conversations that the users expect in voice interactions with systems and present the structural improvements to enable the voice interactions similar to those between people. To observe the users' behavior of voice interaction in the vehicle, the data through contextual inquiry are collected and the interview contents are analyzed by using the open coding. We have been able to explore the usefulness of voice interaction features, which are of great importance in that they increase the user's satisfaction with the features and their usage persistence. This study is meaningful in analyzing the user's empirical needs for the technology of interpersonal model from the perspective of conversation.

Key Words : Conversation, Implicit context, In-vehicle interface, Persistence, Structure, Voice interaction

*This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2020R1A6A3A01096470).

*Corresponding Author : Kiechan Namkung(soundux@kookmin.ac.kr)

Received December 8, 2020

Revised January 19, 2021

Accepted February 20, 2021

Published February 28, 2021

1. 서론

사용자와의 인터랙션이 가능한 음성 기반의 인터페이스는 인공지능 기술의 발달에 따라 사용 범위가 확대되고 있으며[1], 사람들은 이러한 기술과의 상호작용을 사회적 기준으로 대하고[2], 이러한 행동에 대해 대인관계(interpersonal) 모델을 적용하려는 경향이 있다[3].

인공지능 연구는 주로 기술적, 구조적 관점에만 집중되어 왔으며, 사용자 경험 디자인영역에서는 자연어 처리 기술의 발전과 함께 챗봇(Chatbot)이나 대화형 에이전트(Conversational agent)를 중심으로 연구되고 있다[4].

음성을 이용한 인터랙션은 사람 간의 대화와 유사한 상호작용을 가능하게 하는 기술로서, 손이 자유롭게 못한 차량의 운전 환경에서 기존의 다른 인터랙션 방법보다 효과적으로 사용될 수 있다. 이러한 유용성을 바탕으로 자동차 제조사들은 음성을 이용하여 차량의 기능을 제어할 수 있는 다양한 서비스를 상용화하고 있다.

음성 인터랙션을 가능하게 하는 음성 인식 기술은 사용자가 발화하는 음성 문장의 의미를 해석하고 상황과 발화 의도를 파악 할 수 있는 자연어 음성 인식 형태로 점차 발전하고 있지만, 현재까지 상용화된 차량 내 대부분 음성 인터페이스 시스템에서 음성 입력은 거의 예외 없이 명시적인 명령으로 제한되고 있다[5].

이렇게 제한적인 음성 입력의 음성 인터랙션 프로토콜은 비교적 단순하게 구성되는데, 보통 <트리거 워드, 질문, 답변>의 형태를 따른다[5].

차량 내 자연스러운 음성 인터랙션 경험은 운전자에게 운전 집중할 수 있는 안전한 환경을 제공할 수 있다는 점에서 매우 중요한 요소이다. 하지만 하나의 질문과 답변으로 구성된 차량 내 음성 인터랙션 프로토콜은 인간의 대화 방식과는 차이가 있으며, 사용자들에게 자연스러운 인터랙션 경험을 제공할 수 없다.

인간의 대화 방식은 턴-테이킹(turn-taking)이라는 반복적인 과정을 거치게 되며 이는 대화를 확장 및 지속시키는 역할을 한다[6]. 또, 인간의 대화 속에는 문장 속에 명확하게 표현되지 않은 많은 의도들이 숨어 있는데, 이러한 의도를 담고 있는 대화의 형태를 암묵적 단서(implicit cue)라고 한다[5].

본 연구에서는 차량 내 음성 인터랙션을 인간의 대화 방식 요소인 대화 지속성과 암묵적 단서를 중심으로 탐색하여 구조적인 개선방안을 찾고자 한다.

1.1 음성 어시스턴트의 사용

음성 어시스턴트는 전통적으로 수행하는 하위 역할과 업무를 기준으로 낮은 지위를 갖도록 설계되었으며, 이는 언어 사용 패턴과 성별에 반영되어왔다[6].

하지만, 가상의 자연어 인터랙션이 가능한 차량 음성 인터페이스 시스템을 사용한 연구에서는 실험 참가자들이 음성 어시스턴트에게 동등한 사회적 지위와 높은 수준의 인지 능력을 기대하였으며, 인간과의 대화에서와 유사한 언어 사용 패턴을 보였다[7]. 이러한 패턴들은 응답(response), 턴-테이킹, 맞장구(back-channeling)와 채우기(fillers) 및 망설임(hesitation), 모호한 언어(vague language), 요청 완화(mitigating requests)와 공손함(politeness) 및 칭찬(praise) 등의 6가지 주제로 분류되었다[7].

차량에 탑재된 음성 에이전트와 모바일(mobile) 기반의 음성 에이전트의 복합적 사용 경험을 탐색한 국내 연구에서는 차량에 탑재된 음성 에이전트와의 인터랙션보다 모바일 기반의 음성 인터랙션에 익숙한 사용자들의 멘탈 모델(mental model)을 도출하였다[8]. 또, 이러한 멘탈 모델은 모바일 기반의 음성 인터랙션 대화 방법이 보다 자연스럽기 때문이며, 모바일 음성 에이전트가 자동차 환경에 유입되어 운전 경험의 일부로 사용되는 이유라고 분석하였다[6].

1.2 대화 지속성과 암묵적 단서

인간의 대화 방식은 반복적인 턴-테이킹과 맞장구 등 다양한 형태를 통해 지속된다.

차량 내 음성 인터랙션의 턴-테이킹 복잡도에 따른 사용성을 탐색한 연구[9]는 턴-테이킹의 복잡도가 높은 과업(task)일수록 사용자들이 음성 명령보다 기존의 직접적인 조작 방식을 선호하는 경향을 분석하였다. 하지만 이러한 결과를 초래한 원인은 음성 인식 기술 신뢰도에 대한 우려가 대부분이었다. 사용자들은 기술 신뢰도가 높은 음성 인식 시스템과의 인터랙션에서는 인간과의 대화에서와 유사한 대화 지속성을 보여주었다[7, 10].

인간의 대화는 매우 복잡한 과정으로 언어적, 비언어적 단서들로 가득차 있다[10-12]. 이러한 대화의 단서들은 합의나 의견 불일치, 감정 상태, 정보 등을 성공적으로 인지하고 더 잘 전달할 수 있게 해준다[5].

사용자들은 이러한 신호를 감지하거나 응답할 수 없는 시스템과 상호작용할 때에도 이러한 대화 패턴을 보이며[13, 14], 실제로 기술 신뢰도가 높은 음성 어시스턴트

와의 상호작용에서 사용자들은 수많은 대화 단서를 사용한다는 것이 관찰되었다[5].

이러한 단서들을 인식하고 정확하게 행동할 수 있는 음성 인터랙션 시스템은 높은 수준의 사용자 만족을 달성할 수 있지만[5,15], 음성 인터랙션에서의 암묵적 단서들을 시스템이 해석하기 어렵다는 기술적 문제를 가지고 있다.

2. 연구방법

본 연구는 차량 내 음성 인터랙션 구조와 음성 인터랙션에서 대화의 지속성, 그리고 사용자들이 인식하고 있는 암묵적 단서를 조사하기 위해 문화 기술적 접근법(ethnographic approach)의 분석방법으로 맥락 질문법(contextual inquiry)을 사용하였다. 실험 참여자와 함께 차량에 탑승하여 그들의 음성 인터랙션 사용 행태를 관찰하고, 행위 도중 인터뷰를 진행하였다. 인터뷰 기록은 참여자의 동의를 받고 영상 촬영하였으며 인터뷰 내용은 스크립트로 작성하였다.

Table 1. Participants and vehicle information

ID	Gender	Car Type
P1	M	Benz A220 (2020)
P2	M	Hyundai GV80 (2020)
P3	F	Benz A220 (2020)
P4	M	Hyundai GV80 (2020)
P5	F	Benz A220 (2020)
P6	M	Hyundai GV80 (2020)

실험 참가자로 차량 내 음성 인터랙션 사용 경험이 있는 6명을 선정하였으며, 최신 음성 인터랙션 기술이 탑재된 국내외 2개 제조사의 차량을 사용하였다(Table 1).

안전을 고려하여 참가자들은 공터를 지속적으로 주행하도록 하였으며, 주행 중에 평소 사용하던 음성 인터랙션 기능을 자유롭게 사용하도록 하였다. 참가자들은 실험 차량의 음성 인터랙션 기능 작동법을 미리 숙지하였다.

음성 인터랙션의 사용이 제한적인 참가자의 경우 운전자에게 Table 2와 같이 과업을 제시하였으며, 참가자에게 제시하는 과업이 차량에서 지원되지 않는 경우 참가자들의 발화 의도와 예상 시나리오를 인터뷰하였다.

Table 2. Task list

ID	Category	Task
1	Air conditioning	a. Turn on the air conditioner b. Set the desired temperature c. Adjust the rear seat temperature
2	Navigation	a. Set the desired destination b. Cancel the direction
3	Window	a. Open the window b. Close the window
4	Entertainment	a. Turn on the radio b. Set the desired radio channel
5	Others	a. Open the sunroof b. Turn on the heated seat c. Remove the window moisture d. Show me the ambient light menu

3. 결과 및 고찰

3.1 인터뷰 데이터 분석

인터뷰 데이터 분석을 위해 근거이론의 개방코딩(open coding)을 사용하였다[16]. 참여자의 인터뷰 내용 중 불필요한 내용을 제거하고 정리하였으며, 이를 토대로 참여자들이 대화 지속성을 원하는 인터랙션 상황을 찾아보았다. 데이터 분석 결과, 대화의 확장과 지속이 필요한 인터랙션 상황들을 발견하였다.

3.1.1 공조 기능

참가자들은 차량의 공조 기능 작동 시, 음성 인터페이스 시스템의 동작 완료 후 대화의 단절을 느꼈으며, 대화의 확장과 지속이 필요하다고 응답하였다.

“출다고 말했을 때, 26도로 설정하게 되어 있는 시스템 인거 같다. 26도로 온도를 설정한 후에 온도가 괜찮은지 물어봐 준다면, 너무 덥거나 추울 경우 원하는 온도를 다시 명령하고 싶은데... 대화가 끝나버리고 다시 웨이크 업(wake-up) 버튼을 눌러서 온도를 조절해야 하는게 불편하다” (P2)

“너무 더워”라고 말하면 온도가 1도씩 내려가게 설정되어 있는데, 사실 ‘온도를 몇 도로 할까요?’ 라던가 ‘20도로 설정할까요?’ 라는 제안을 기대했다. 사실 1도 차이는 사람이 크게 차이를 느끼기 어려운 온도니까.. 1도씩 낮추더라도 그 이후에 온도가 괜찮은지 물어봐 주면 좋겠는데...” (P6)

“더워”라고 했을 때 적절한 온도를 제안해주거나, 몇 도로 낮출지를 물어볼 것이라고 기대했는데, 무조건 온도가 최저로 설정되고 끝나니 당황스럽다” (P1)

“바람 세기를 약하게 해달라고 했을 때, 최저 단계로 설정이 되고 끝나 버리니 사용하기 불편할 것 같다. 바람 세기 조절 후 마음에 안들면 다시 바꿔야 하는데 다시 물어 봐주면 좋을 것 같다... 설정을 바꿀 때마다 웨이크 업 버튼을 눌러야 하는 것은 너무 불편하다.” (P4)

참가자들은 공통적으로 시스템과의 지속적인 대화를 통해 적절한 차량 내의 환경을 설정하기를 원했다. 특히 참가자들은 시스템의 응답 이후에 설정된 환경에 대한 만족 여부에 따라 관련된 발화 기회를 기대하는 것으로 나타났다. 이러한 기대는 동일한 형태의 사람 간의 대화 속에는 다음과 같이 설정된 환경의 만족 여부를 재확인 하는 대화 지속성을 갖는 암묵적 단서가 포함되어 있기 때문으로 유추된다.

A : 너무 덥지 않아?

B : 온도를 좀 낮출게. (어때?)

A : 조금 더 낮추자.

3.1.2 창문 여닫기 기능

창문 여닫기를 음성 명령으로 실행해 본 참가자들은 공조 기능과 마찬가지로 미세한 조절을 위한 대화 지속성을 기대하였다.

“차 안에서 창문을 여닫을 때 보통 완전히 열고 닫기 보다는 조금씩 열고, 닫으면서 적당한 환경을 맞추는 경우가 대부분일 것이다.” (P2)

“창문을 ‘조금만 닫아줘’ 라고 했는데 완전히 닫아버리고 끝나버린다.” (P4, P6)

참가자들은 창문을 여닫을 때 일반적으로 완전히 여닫는 행위보다는 조금씩 열고, 닫으면서 적정 환경을 설정한다고 응답하였다. 현재 제공되고 있는 음성 인터랙션 기능은 이러한 조절이 불가능하였지만, 추후 이러한 기능이 제공된다면 일정 높이만큼의 창문 여닫기 동작과 함께 설정값에 대한 확인을 기대하는 것으로 나타났다.

“창문을 조금 여는 기능이 지원된다면, 창문을 조금 연 후에 운전자가 맘에 안들 경우, 바로 이어서 명령을 통해 조금 더 여닫을 수 있으면 좋을 것 같다. ‘조금 더 열 어줘’ 또는 ‘조금 더 닫아줘’ 같은 대화가 이어져야 하지 않을까?” (P3, P5)

3.1.3 기타

공조 및 창문 여닫기 기능 외에도 세부 조절을 통해

사용자가 원하는 환경을 설정해야 하는 경우, 참가자들은 공통적으로 대화의 확장 및 지속을 기대하는 것으로 나타났다.

“라디오 주파수를 운전자가 다 외울 수는 없을 것 같다. 그래서 사람들은 보통 라디오를 켜 후에, 프리셋(preset)에 저장되어 있는 번호를 차례로 눌러서 듣고 싶은 주파수를 찾는데, 라디오를 켜고 다시 명령하여 한 번의 명령에 하나씩 주파수를 말하는 것은 너무 힘든 것 같다. 라디오를 켜달라는 명령 후에 바로 채널을 찾는 명령까지 한 번에 자연스럽게 할 수 있으면 좋겠다” (P1, P2)
“선루프(sunroof)를 열 때, 완전히 열 때도 있지만 조금만 열고 싶은 경우도 있는데 이러한 기능이 지원 안되는 것 같다. 선루프를 연 후에 ‘조금만 닫아줘’ 같은 명령을 하게 될 것 같다.” (P6)

참가자들의 인터뷰에서 사용자가 라디오를 켜는 것은, 그다음 원하는 라디오 채널(channel)을 찾겠다는 행동을 내포하고 있음을 알 수 있다. 따라서 사용자들은 음성 명령을 통해 라디오를 켜는 경우, ‘라디오를 켜니다’라는 시스템 응답 속에 ‘어떤 채널을 틀까요?’라는 암묵적인 단서가 있을 것으로 생각한다.

3.2 인터랙션 구조도

실험을 통해 Fig. 1과 같이 현재 상용화된 차량 내 음성 인터랙션 구조를 도출할 수 있었다.

시스템은 대기(stand by) 상태에서 사용자의 요청(call)으로 웨이크 업 상태로 되어 명령(command)을 기다린다. 사용자의 명령을 청취(listening)한 시스템은 처리(processing)과정을 거쳐 결과(result) 행동을 완료한다. 이처럼 현재 상용화된 차량 내 음성 인터랙션 구조는 시스템과 사용자 측면 모두 선형적이고(linear) 순차적인(sequential) 단순한 구조로 되어있다. 하지만 실험을 통해 차량 내 음성 인터랙션을 사용하는 사용자들은 일상적인 대화의 패턴(pattern)과 유사한 대화의 확장 및 지속을 기대하고 있음을 발견하였으며, 이러한 대화 지속성을 갖는 암묵적 단서는 시스템의 결과 응답 직후에 뒤따를 수 있었다.

시스템이 대화 속에서 암묵적 단서를 의미적으로 해석하도록 구현하는 것은 기술적으로 어려울 수 있지만, Fig. 2와 같이 시스템이 결과 응답 후 일정 시간 동안 사용자의 명령을 기다리는 방법을 통해 대화 확장이 가능할 수 있다.

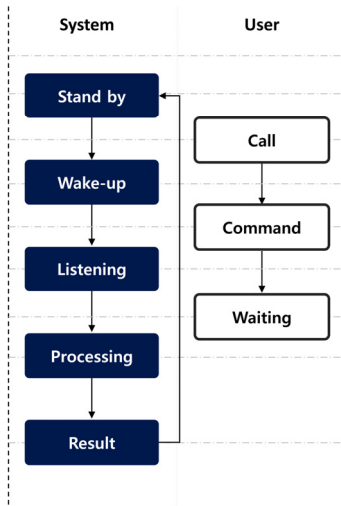


Fig. 1. In-vehicle voice interaction structure (currently)

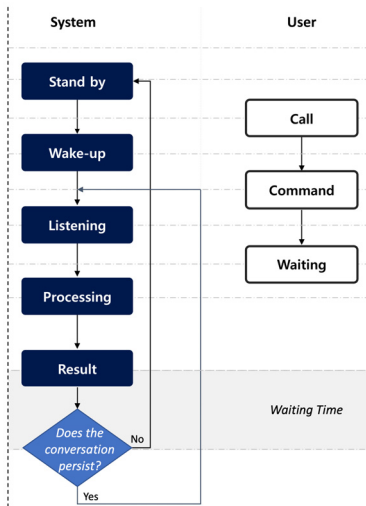


Fig. 2. In-vehicle voice interaction structure (including conversation persistence)

시스템은 결과 응답 후 일정 시간 동안 사용자의 추가 발화를 기다리며, 발화가 없을 시에는 대기 상태로 돌아간다. 반면, 사용자의 발화가 대화의 확장으로 이어지는 경우 추가적인 발화 요청 단계를 거치지 않고, 명령을 처리하게 되는 구조이다.

4. 결론

본 연구에서는 맥락 질문법을 통해 차량 내 음성 어시

스틴트와 사용자의 인터랙션을 정성 조사하였다. 이를 통해 음성 인터랙션 기능의 유용성을 탐구할 수 있었으며, 이러한 유용성은 기능에 대한 사용자의 만족도와 사용 지속성을 증가시킨다는 점에서 매우 중요하다.

본 연구의 결론은 다음과 같다.

첫째, 현재 상용화된 차량 내 음성 인터랙션은 한 번의 명령과 결과 및 응답의 프로세스로 종료되어 대화의 확장 및 지속이 불가능한 구조로 되어있다. 하지만 사용자들은 시스템과의 인터랙션에서도 사람 간의 대화에서 나타나는 암묵적 단서를 통한 대화의 확장 및 지속을 원하고 있었으며, 특히 반복되는 조작으로 적정 환경을 설정해야 하는 경우 이러한 경향이 두드러지게 나타났다.

둘째, 사용자들은 하나의 기능과 관련된 다수의 음성 명령을 발화해야 하는 경우, 매번 음성이나 버튼을 통해 발화 요청을 해야 하는 상황에서 불편함을 느끼고 있었다. 이러한 결과들은 현재 차량 내 음성 시스템이 결과 응답 후 일정 시간 동안 사용자의 추가 발화를 기다리는 구조적 개선이 필요함을 시사한다.

본 연구는 이와 같은 결과를 도출하였지만 다음과 같은 한계점을 갖는다. 진행된 실험은 참가자에 따라 음성 인식 기능의 사용 행태가 주관적일 수밖에 없다. 따라서 보다 많은 수의 참가자들을 확보했다면 실험의 객관성을 더할 수 있었을 것이다. 또, 개선된 구조에서 시스템의 결과 응답 후 입력된 사용자의 명령이 이전 명령과 관련된 명령인지 또는 새로운 화제의 턴-테이킹인지를 구분해야 하는 문제와 대화의 지속이 필요 없는 경우에도 대기 시간 동안 시스템의 마이크가 동작하여 발생할 수 있는 성능 저하나 예상치 못한 입력 오류 등은 본 연구가 가지고 있는 기술적 한계점이라 할 수 있다.

본 연구에서 도출된 차량 내 음성 인터랙션의 구조적 시사점을 통해 사람 간의 자연스러운 대화와 유사한 차량 내 음성 인터랙션 기능 구현을 기대한다.

REFERENCES

- [1] K. Namkung. (2020). Research on Emotional Factors and Voice Trend by Country to be considered in Designing AI's Voice - An analysis of interview with experts in Finland and Norway. *Journal of the Korea Convergence Society*, 11(9), 91-97. DOI : 10.15207/JKCS.2020.11.9.091
- [2] B. Reeves & C. Nass. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK:

- Cambridge University Press.
- [3] K. Gelbrich, J. Hagel, & C. Orsingher. 2020. Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence. *International Journal of Research in Marketing*. DOI: 10.1016/j.ijresmar.2020.06.004
- [4] S. Cheon & M. Yeoun. 2020. Categorization of Interaction Factors through Analysis of AI Agent Using Scenarios. *Journal of the Korea Convergence Society*, 11(11), 63-74. DOI : 10.15207/JKCS.2020.11.11.063
- [5] V. Alexandra & F. Adam. (2018, April). Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. (pp. 1-7). New York : ACM. DOI: 10.1145/3173574.3173782
- [6] C. Hannon. 2016. Gender and status in voice user interfaces. *Interactions*, 23 (3), 34-37. DOI : 10.1145/2897939
- [7] R. L. David, L. Clark, A. Quandt, B. Gary, & S. Lee. (2017). Steering the conversation: A linguistic exploration of natural language interactions with a digital assistant during simulated driving. *Applied ergonomics*, 63, 53-61. DOI : 10.1016/j.apergo.2017.04.003
- [8] S. Lee, N. Kim, M. Yang, Y. Yoo, & J. Kim. (2018, February). A Qualitative Study of User Behavior on Interacting with In-vehicle Voice User Interface. *Proceedings of HICK 2018*. (pp. 338-342). Korea : The HCI Society of Korea.
- [9] J. Shin, S. Lee, & J. Kim. (2020, February). Explore the Usability of the Driver's Voice Interface in the Vehicle. *Proceedings of HICK 2020*. (pp. 567-577). Korea : The HCI Society of Korea.
- [10] M. Mizukami, K. Yoshino, G. Neubig, D. Traum, & S. Nakamura. (2016). Analyzing the Effect of Entrainment on Dialogue Acts. In *SIGDIAL Conference*. (pp. 310-318).
- [11] D. Traum & E. Hinkelman. 1992. Conversation Acts in Task-Oriented Spoken Dialogue. *Computational intelligence*, 8(3), 575-599.
- [12] D. Bohus & E. Horvitz. (2014, November). Managing human-robot engagement with forecasts and... um... hesitations. In *Proceedings of the 16th international conference on multimodal interaction*. (pp. 2-9). New York : ACM. DOI : 10.1145/2663204.2663241
- [13] E. Luger & A. Sellen. (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) New York : ACM. DOI : 10.1145/2858036.2858288
- [14] K. Morrissey & J. Kirakowski. (2013, July). "Realness" in Chatbots: Establishing Quantifiable Criteria. *HCI'13: Proceedings of the 15th international conference on Human-Computer Interaction: interaction modalities and techniques*. (pp. 87-96). Springer. DOI : 10.1007/978-3-642-39330-3_10
- [15] L. Yu, K. Hermann, P. Blunsom, & S. Pulman. (2014, December). Deep learning for answer sentence selection. *arXiv*, preprint arXiv:1412.1632.
- [16] S. Elo, & H. Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing*, 62(1), 107-115. DOI : 10.1111/j.1365-2648.2007.04569.x

남궁 기 찬(Kiechan Namkung)

[정회원]



- 2019년 2월 : 국민대학교 경험디자인 과(디자인학박사)
- 2020년 9월 ~ 현재 : 국민대학교 산학 협력단 전임연구교수
- 관심분야 : UX, AUI, VUI
- E-Mail : kc.namkung@gmail.com