

뇌파기반 사용자 맞춤형 음성기반 인터랙션 시스템 개발을 위해

고려해야 할 시스템 설계변수 조사

이준석^{1,2}, 최가영¹, 신종규³, 이지윤^{1,2}, 김상호³, 황한정^{1,2*}

고려대학교 전자및정보공학과¹

고려대학교 인공지능 스마트융합기술 융합전공²

금오공과대학교 산업공학과³

Investigation of system parameters to be considered for the development of a user-customized voice-user interaction system using electroencephalography

J-S. Lee^{1,2}, G-Y. Choi¹, J-G. Shin³, J-Y. Lee^{1,2}, S-H. Kim³, H-J. Hwang^{1,2*}

¹Dept. of Electronics and Information Engineering, Korea university, Sejong, Republic of Korea

²Interdisciplinary Graduate Program for Artificial Intelligence Smart Convergence Technology, Korea University, Sejong 30019, Republic of Korea

³Dept. of Industrial Engineering, Kumoh National Institute of Technology, Gumi 39177, Republic of Korea

*hwanghj@korea.ac.kr

Abstract

To develop a more realistic voice-user interaction (VUI) system, it is required to first define the important factors of a VUI system in terms of the response type of a VUI system, and thereby a VUI system could flexibly change the important factors to provide appropriate responses with the user. In this study, we investigated the two factors used to create VUI responses, i.e., voice type and information quantity, to check they should be necessarily considered when developing a user-adaptive VUI system using electroencephalography (EEG). EEG data were measured from fifty subjects while they repetitively interacted with our developed virtual-VUI system, and the EEG data were used to classify neurophysiological changes caused by different response types of the VUI system using a shallow-convolutional neural network. The performances of all possible combinations of the two VUI parameters were evaluated using a 5×5-fold cross-validation. The mean classification accuracies of the two parameters (voice type and information quantity) were 58.3% and 96.1 %, respectively, showing the importance of information quantity when building a user-adaptive VUI system.

1. 연구 배경

현대사회에서 인간은 대화를 통해 자신의 의사 및 감정을 전달한다. 최근에는 인공지능 기술의 발달로 기계와의 자연어 소통이 가능해지면서 대화형 인간-기계 인터페이스(human-machine interface: HMI) 제품들이 출시되고 있다[1,2]. 특히 삼성의 빅스비(Bixby), 애플의 시리(Siri), 구글의 구글 어시스턴트(google assistant) 등에 적용된 음성기반 인터랙션(voice-user interaction: VUI)은 대화하듯 명령과 응답을 주고받을 수 있기에 별도의 사용법 숙지를 요구하지 않고 누구나 쉽게 사용 가능하다는 장점이 있다.

그러나, 현재의 VUI 시스템은 사용자의 단순명령에 초기 설정된 조건에 따른 획일화된 답변을 제공하는 수준에 그치고 있다. 보다 실용적인 VUI 시스템을 개발하기 위해서는 VUI 시스템과 사용자가 상호작용 도중 인간의 감정 변화를 인지하고, 개인의 선호에 맞는 적절한 응답을 제공할 필요가 있다. 이에 본 연구에서는 VUI 시스템의 답변 종류에 따른 사용자의 신경생리학적 변화를 이용하여 사용자 맞춤형 VUI 시스템 개발 위해 고려해야 할 설계변수를 모색하고자 하였다. 이를 위해, 두 가지 설계변수(음성 종류와 답변 정보량)로 구성된 가상의 VUI 시스템과 상호작용하는 도중에 측정된 뇌파(electroencephalography: EEG) 데이터를 이용하여 답변 종류에 따른 사용자의 신경생리학적 반응의 분류 가능성을 확인하였다. 나아가, 사용자 성별에 따라 동일 답변종류에 따라 신경생리학적 반응이 다른지를 확인하기 위해, 남/여간

답변종류에 따른 분류 정확도 차이도 함께 확인하였다.

2. 연구 방법

본 연구에서는 가상 VUI 시스템을 구축하고, 총 50명의 피험자(남성 26명, 여성 24명)가 해당 시스템을 사용하는 동안의 뇌파 데이터를 취득하였다. 뇌파 데이터는 국제 10-20 전극 배치법(international 10-20 system)에 따라 63개의 두피 채널에서 측정되었으며(샘플링율: 1,000 Hz), 접지/기준 전극은 각각 FPz/FCz에 부착하였다.

모든 피험자들은 모니터 화면에 제시되는 문장을 소리 내어 읽음으로써 VUI 시스템에 질문하고, 시스템이 제공하는 답변을 들은 뒤, 3초간 안정상태를 유지하도록 지시받았다. 이후, 화면에 제시되는 감성을 표현하는 형용사마다 점수식 평정척도(numerical rating scale: NRS)를 기반으로 VUI 시스템을 통해 제공받은 답변에 대한 실제 본인의 느낌을 정량적으로 기록하도록 지시하였다. 이를 하나의 시행(trial)으로 간주하고, 음성 종류(어린이/성인여성)와 답변 정보량(간단답변/출처답변) 설계변수를 기반으로 4가지(어린이-간단답변/어린이-출처답변/성인여성-간단답변/성인여성-출처답변) 답변 종류에 대해 각 20회씩, 총 80회 시행을 반복 수행하였다(그림 1 참조).

측정된 원신호 뇌파 데이터는 계산량 감소를 위해 200 Hz로 다운샘플링(down-sampling)을 적용하고, 1-80 Hz 대역에 대한 대역통과필터(band-pass filter)를 적용하였다. 필터링된 데이터를 기반으로 VUI 시스템의 답변종류에 따른 신경생리학적 반응의 분류 가능성을 확인하고자 피험자가 시스템으

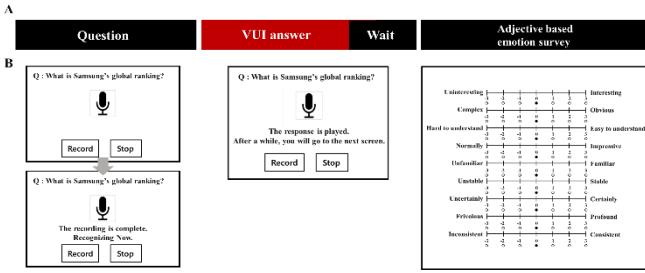


그림 1. 실험 패러다임 및 피험자에게 제공되는 모니터 화면 도식화

로부터 답변 답변을 제공받는 구간에 대해 분할(epoching)하여 개인별 80 시행의 데이터를 추출하였다. 이 때, 필터링 이후에도 $\pm 100 \mu V$ 이상의 값을 가지는 뇌파 채널은 잡음에 오염된 것으로 판단하여 양쪽 측두엽의 6개 채널(FT9, T7, TP9, FT10, T8, TP10)의 데이터는 분석에서 제외하였다. 답변 종류에 따라 모든 시행의 샘플 길이가 불일치하여 업샘플링(up-sampling) 기반의 데이터 증강기법(data augmentation)을 통해 모든 시행의 데이터 길이를 일치시켰다.

추출한 데이터 이용하여 5x5-겹 교차검증 기반 얇은 합성곱신경망(shallow-convolutional neural network: ShallowNet) 모델을 통해 VUI 시스템이 제공하는 답변의 음성타입과 정보량 차이에 따른 분류 성능을 확인하였다 [3]. 또한, 사용자 성별에 따른 차이를 확인하고자 피험자군을 남/여로 구분하여 두 설계변수의 가능한 모든 조합에 대한 이진 분류 성능도 함께 확인하였다.

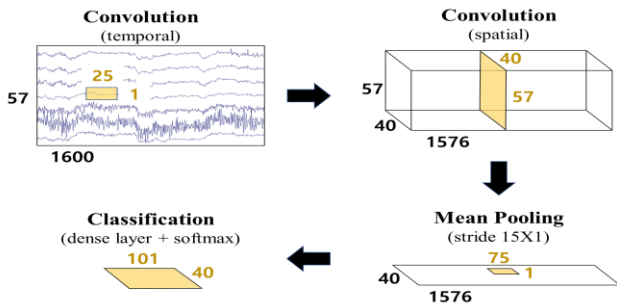


그림 2. 본 연구에서 사용한 딥러닝 모델 도식도

3. 연구 결과

그림 3은 VUI 시스템의 음성과 정보량 차이에 따른 평균 분류 정확도를 나타낸다. 그 결과, 음성 차이(어린이-성인 여성)의 경우 58.3%, 정보량 차이(간단답변-출처답변)의 경우 96.1%의 분류 정확도를 보였다.

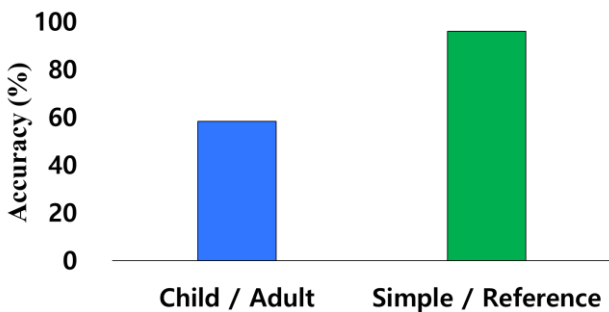


그림 3. VUI 시스템의 답변타입에 따른 평균 분류 정확도

표 1은 사용자 성별에 따른 VUI 시스템 답변 타입 별 평균 분류 정확도를 나타낸다. 그림 2의 결과와 유사하게 음성 종류보다 답변 정보량이 평균 96.5 ± 3.93 %로 더 높은 분류 정확도를 보였지만 성별에 따른 분류 성능 차이는 확인하지 못했다.

표 1. 성별에 따른 답변 타입 별 평균 분류 정확도 (어린이(C)-간단답변(S): CS, 어린이(C)-출처답변(R): CR, 성인여성(A)-간단답변(S): AS, 성인여성(A)-출처답변(R): AR)

Accuracy (%)	Voice		Information	
	CS/AS	CR/AR	CS/CR	AS/AR
Male	60.1	60.2	93.3	100.0
Female	57.2	54.8	92.9	99.8

본 연구에서는 VUI 시스템과 상호작용하는 도중의 답변 종류에 따른 사용자의 신경생리학적 변화를 특징으로 활용하여 사용자에게 영향을 미치는 주요 설계변수를 탐색하였다. 나아가, 사용자의 성별에 따른 감수성의 차이가 존재할 것으로 가정하여, 피험자군을 성별에 따라 구분하고 그에 따른 분류 정확도도 함께 확인하였다. 그 결과, 답변의 정보량 변수에서 평균 96% 이상 높은 분류 정확도를 보였으며, 음성 종류변수는 기회수준(chance level)을 조금 상회하는 분류 성능을 보였다. 이러한 결과는 성별에 따른 분류 결과에서도 동일하게 나타났으며, 현재의 설계변수 조건에서는 성별 간 유의미한 차이는 확인하지 못했다. 현재까지 결과를 토대로, VUI 시스템이 제공하는 답변의 정보량은 사용자 맞춤형 VUI 시스템을 구축할 때 고려해야 할 중요한 변수로 사료된다.

다만, 본 연구에서는 실험 시 획득한 사용자의 실제 답변에 대한 만족도 설문결과를 활용하지 않았기에 현재의 답변 정보량이 가지는 높은 분류성능이 사용자의 어떤 감정변화에 의해 유발되었는지에 대한 객관적인 검증이 필요하다. 이에 후속 연구에서는 설문 결과를 바탕으로 현재의 우수한 분류성능에 대한 신경생리학적 해석을 시도할 예정이다.

4. Acknowledgements

본 연구는 한국연구재단을 통해 과학기술정보통신부의 기초연구사업으로부터 지원받아 수행되었습니다(2020R1A4A1017775).

5.참고 문헌

- [1] Landay, J. A., Oliver, N., & Song, J. "Conversational user interfaces and interactions." IEEE Computer Architecture Letters, Vol. 18, p.8-9, 2019.
- [2] Sugisaki, Kyoko, and Andreas Bleiker. "Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach." Proceedings of the Conference on Mensch und Computer, 2020.
- [3] Schirrmester, Robin Tibor, et al. "Deep learning with convolutional neural networks for EEG decoding and visualization." Human brain mapping, 2017