

# 음성 및 동작 인식 기반의 미디어 제어 애플리케이션 설계 및 구현

이원주\*, 간명해<sup>o</sup>, 강민수\*, 김태완\*, 임정주\*, 강지우\*

<sup>o</sup>인하공업전문대학 컴퓨터정보과,

\*인하공업전문대학 컴퓨터정보과

e-mail: wonjoo2@inhatc.ac.kr, dlawjdown12@gmail.com

## Design and Implementation of Media Control Application Based on Speech and Motion Recognition

Won Joo Lee\*, Kan Myeonghae<sup>o</sup>, Kang Minsu\*, Kim Taewan\*, Im Jeongju\*, Kang Jiwoo\*

<sup>o</sup>Dept. of Computer Science, InHa Technical College,

\*Dept. of Computer Science, InHa Technical College

### ● 요약 ●

본 논문에서는 미디어 플레이어 제어가 어려운 지체 장애인들을 위해 음성과 동작 인식 기반의 미디어 제어 애플리케이션을 설계하고 구현한다. 이 애플리케이션은 사용자의 음성 인식을 위해 먼저 명령어를 정하고, 명령에 매핑되는 키워드 관리하는 데이터 모델을 생성한다. 그리고 이 데이터 모델을 JSON 파일로 정제하여 사용한다. 그리고 키넥트 센서를 활용한 동작 인식은 오른쪽 어깨를 중심으로 오른쪽 손목의 좌표값을 인식함으로써 동작 인식 제어 컨트롤을 실행한다. 오른쪽 어깨를 기준으로 오른쪽 손목의 좌표값으로 현재 팔의 위치를 정하고, 영역 1~4 에 따라 동작을 인식한다.

**키워드:** 미디어 제어(media control), 음성인식(speech recognition), 동작인식(motion recognition)

### I. Introduction



Fig. 1. 연도별 등록 장애인 현황

최근 유튜브, 영화, 드라마 같은 각종 미디어 콘텐츠를 이용하는 사용자가 증가하고 있다. 시각 장애 또는 청각 장애 사용자를 위해 구글 실시간 자막 앱이나 맞춤형 TV 등이 제작되고 있다. 하지만 지체장애인을 위한 서비스는 부족하다. 2017년 12월 기준으로 국내 등록장애인은 약 255만명으로 전국 주민등록 인구대비 4.9%이다. 특히 65세 이상 노년층 장애인은 2016년 43.4%에서 2017년 45.2%로 지속적으로 증가하고 있다. 15개 장애 유형중에 지체장애인 비율은 49.3%로 가장 높다[1].

따라서 본 논문에서는 지체장애인들을 위해 음성과 동작 인식을 기반으로 미디어 콘텐츠 제어를 할 수 있는 애플리케이션을 설계하고 구현한다. 이 애플리케이션은 키넥트 센서의 음성 및 동작 인식 기능을 활용한다[2].

### II. 미디어 제어 애플리케이션 설계 및 구현

본 논문에서는 음성과 동작 인식 기반의 미디어 제어 애플리케이션을 설계하고 구현한다. 이 애플리케이션은 키넥트 센서의 음성 및 동작 인식 기능을 활용한다. 먼저 사용자의 음성 인식을 위해 먼저 명령어를 정하고, 명령에 매핑되는 키워드 관리하는 데이터 모델을 생성한다. 그리고 이 데이터 모델을 JSON 파일로 정제하여 사용한다. 명령어와 키워드를 입력하면 JSON으로 정제해주는 폼을 생성함으로써 JSON 파일 관리를 쉽게 할 수 있다. 이때 인식된 음성이 정제된 JSON 파일과 매핑되면 해당 명령어를 실행한다. 그림 2는 명령어와 키워드를 매핑한 예이다.

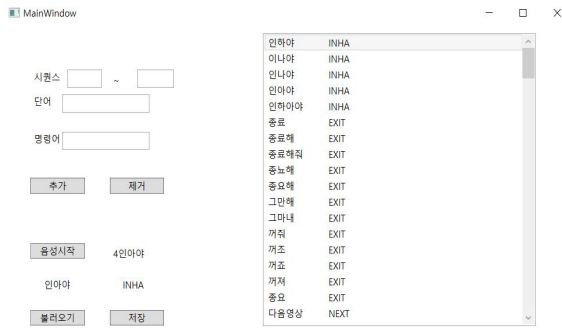


Fig. 2. Mapping example of command and keyword

키넥트 센서를 활용한 동작 인식은 오른쪽 어깨를 중심으로 오른쪽 손목의 좌표값을 인식함으로써 동작 인식 제어 컨트롤을 실행한다. 모션 인식 범위는 그림 3과 같다.

가로 X 세로 Y

(0,0)Shoulder

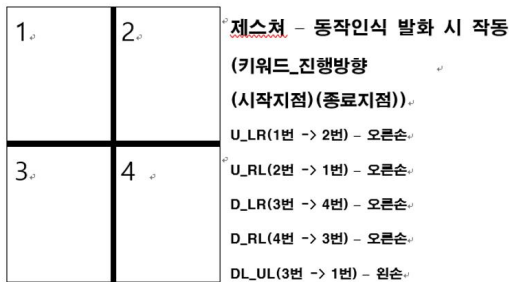


Fig. 3. Motion recognition range

오른쪽 어깨를 기준으로 오른쪽 손목의 좌표값으로 현재 팔의 위치를 정하고 그림 3의 해당 영역 1-4 에 따라 동작을 인식한다. 즉, 오른쪽 어깨를 기준으로 오른쪽 손목이 어깨보다 x값이 작고 y값이 크면 1영역이고, x값이 크고 y값이 크면 2영역이다. 그리고 오른쪽 어깨를 기준으로 오른쪽 손목이 어깨보다 x값이 작고, y값이 작으면 3영역이고, x값이 크고, y값이 작으면 4영역이다.



Fig. 4. Media control application execution

그림 4는 미디어 제어 애플리케이션의 실행화면이다. 그림 4에서 U\_LR 값을 받을 때 오른쪽 팔을 1영역에서 2영역으로 이동하며 원을 계속 그리면 1초씩 후로 계속 돌리고, 오른쪽 팔을 2영역에서

1영역으로 이동하며 원을 계속 그리면 1초씩 전으로 계속 돌린다. D\_LR 값을 받을 때 오른쪽 팔을 3영역에서 4영역으로 이동하면 10초 후로 돌리고, 오른쪽 팔을 4영역에서 3영역으로 이동하면 10초 전으로 돌린다. DL\_UL 값을 받을 때 왼쪽 팔을 3영역에서 1영역으로 이동하면 동영상 재생을 정지한다.

### III. Conclusions

본 논문에서는 음성 및 동작 인식으로 미디어 플레이어를 제어할 수 있는 애플리케이션을 설계하고 구현하였다. 이 애플리케이션은 키넥트 센서의 음성 및 동작 인식 기능을 활용하였다. 향후 연구과제로는 음성 인식의 정확성과 인식률을 높여 사용자의 불편을 최소화하는 것이다. 그리고 스마트폰의 음성 인식 기능과 카메라를 활용하여 모바일 환경에서도 활용할 수 있도록 구현하는 것이다.

### REFERENCES

- [1] "장애인현황 통계정보보고서," 보건복지부, Dec. 2018.
- [2] J. K. Ko, "Kinect Programing," Korea Electronics Association, 2012.