

연구발표자료 - 1

사람이 소리의 방향을 구분하는 방법

Interaural Level Difference – 크기 차이

Interaural Time Difference – 시간 차이

Monaural Spectral Cues – “귓바퀴에 의한” 주파수 차이

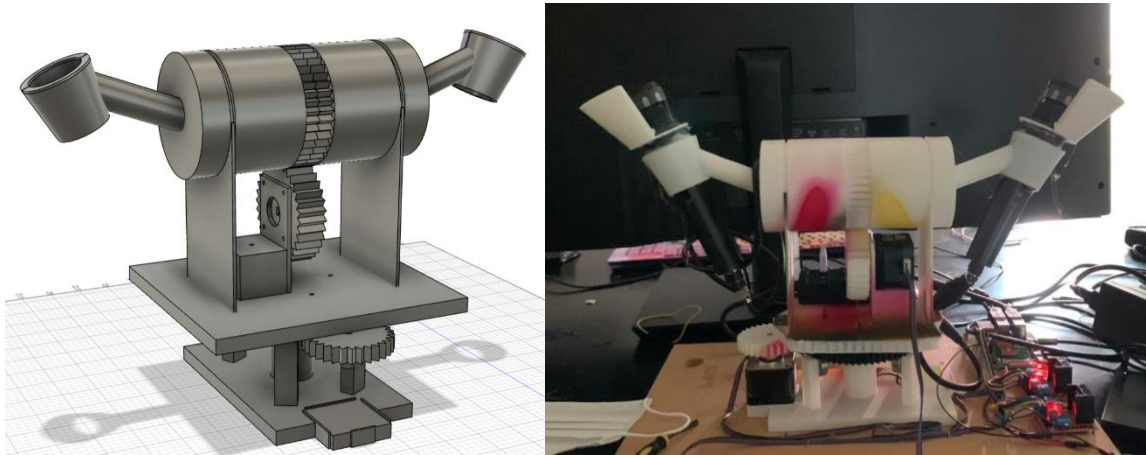
인간이 소리의 방향을 구분하는 방법에는 세 가지가 있습니다. 첫째, Interaural Level Difference(ILD)로, 양쪽 귀에 도달하는 소리의 음량 차이를 이용합니다. 둘째, Interaural Time Difference(ITD)로, 소리가 양쪽 귀에 도달하는 시간 차이를 활용합니다. 셋째, Monaural Spectral Cues로, 귓바퀴에 의한 주파수 변화를 통해 소리의 방향을 파악합니다.

이 연구의 목적은 다양한 귓바퀴 모양과 인공지능 알고리즘을 이용하여 보다 효과적인 방향 탐지 장치를 만드는 것입니다. 수많은 인간과 동물이 하듯, 기계 또한 귓바퀴 모양을 사용한다면 훨씬 적은 수의 마이크로도 효과적인 방향 구분이 가능할 것입니다.

연구 방법은 다음과 같습니다. 먼저, 두 개의 귓바퀴가 각각 접합된 마이크를 사용하여 모든 구 좌표계의 각 단위 방향에서 녹음된 박수 소리를 수집합니다. 그런 다음, 수집된 소리 데이터를 딥러닝 데이터 수집기를 통해 처리합니다. 마지막으로, CNN을 활용한 인공지능 알고리즘을 사용하여 소리 방향 분석의 정확도를 회귀 문제로 접근하여 구합니다.

연구발표자료 - 2

TSM-1 : 데이터 수집기의 제작과 시행착오



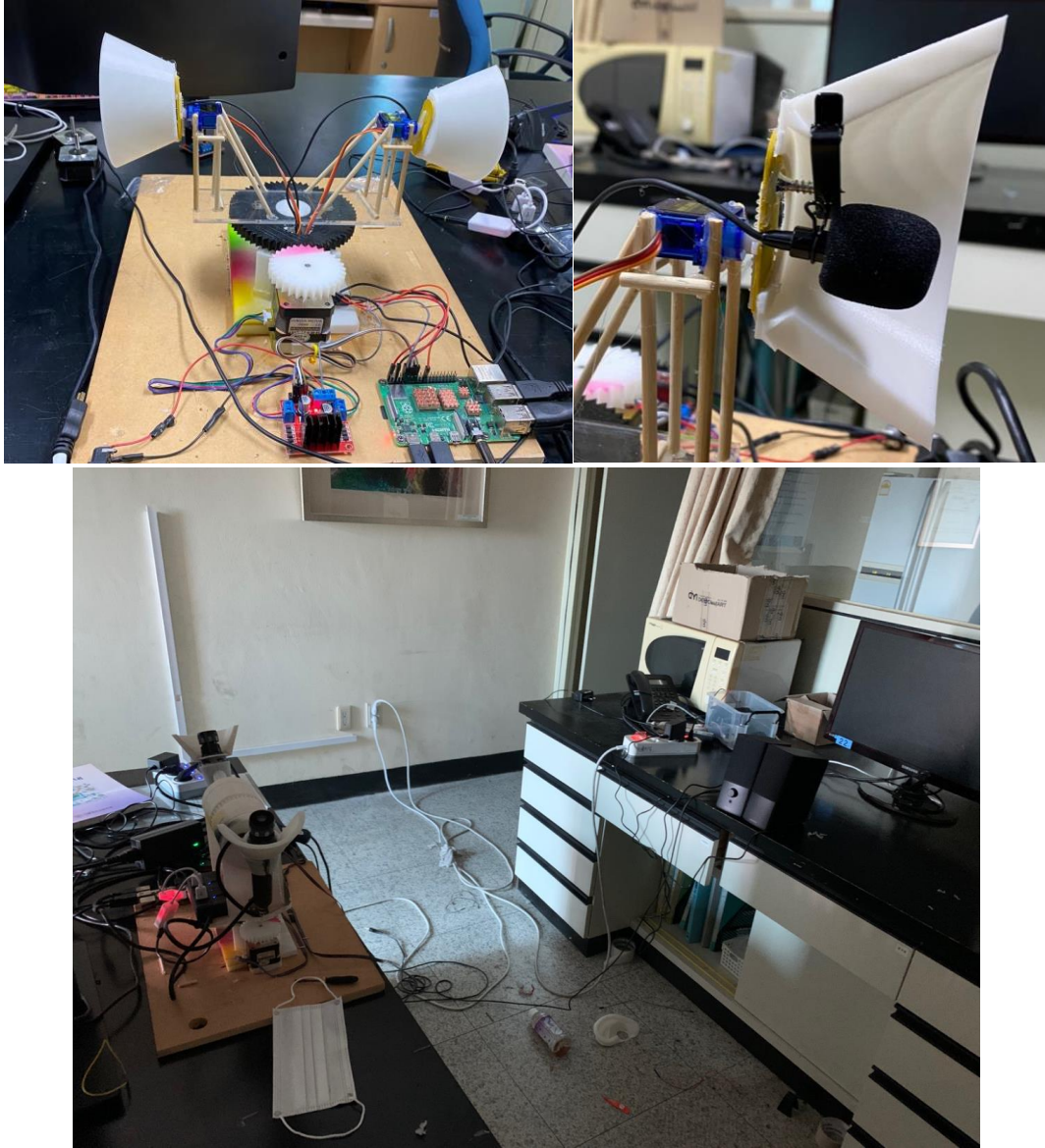
저는 구 좌표계 상 수많은 소리 데이터를 수집하기 위해 데이터 수집기인 TSM (Train Set Maker)를 제작하였습니다. TSM을 제작하는 과정에서 많은 시행착오가 있어, 첫 번째 만든 것을 TSM-1, 최종으로 사용한 것을 TSM-2라고 부르겠습니다.

TSM-1은 소리의 방향을 탐지하기 위해 설계된 장치로, 여러 구성 요소로 이루어져 있습니다. 오른쪽 그림에서는 위에 마이크와 귓바퀴 구조물이 있으며, 우측에는 라즈베리 파이와 두 개의 L298N 스텝 모터 컨트롤러가 보입니다. 귓바퀴의 구조는 가장 간단한, 반으로 잘린 원뿔 각을 사용했습니다.

그러나 TSM-1은 마이크의 큰 무게와 부족한 모터 토크로 인해 회전하지 않았습니다. Polar angle 회전을 담당하던 스텝 모터가 계속된 발열로 인해 열에 약한 3D 프린트된 기어 부속을 녹여 더 이상 작동하지 않는 문제가 발생했습니다. 이를 해결하기 위해 TSM-2는 아예 설계를 변경하였습니다.

연구발표자료 - 3

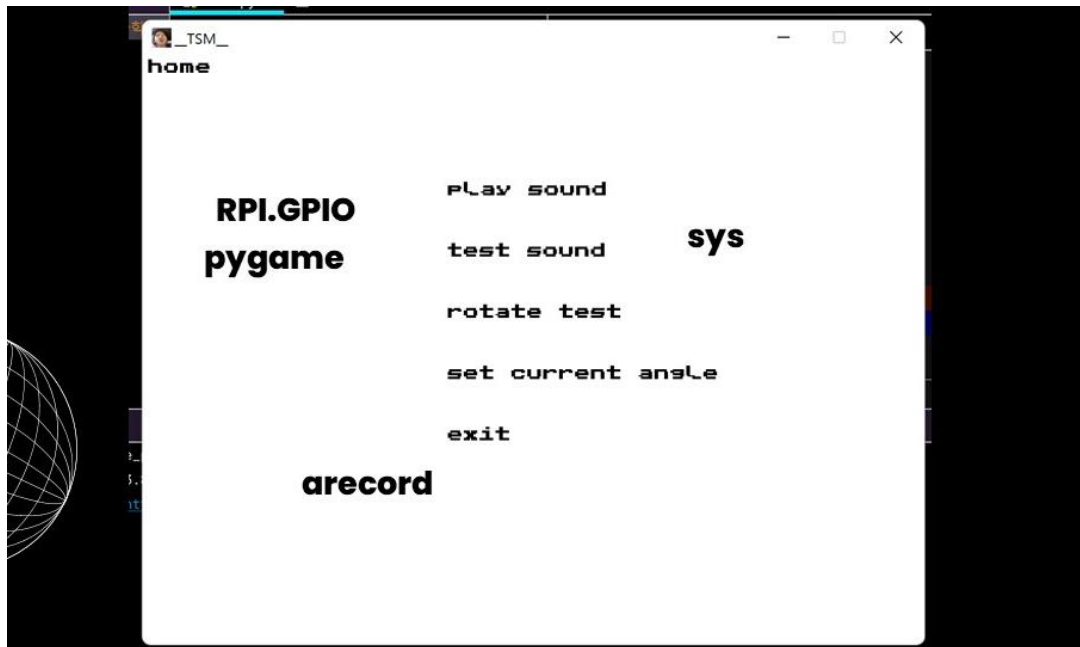
TSM-2 : 연구에 사용된 데이터 수집기



TSM-2의 개선 사항은 다음과 같습니다. 좌측 그림은 앞에서 바라본 TSM-2의 모습을 보여주는데, 회전자가 경량화된 것을 한눈에 볼 수 있습니다. 우측 그림에서는 TSM-2의 마이크가 핀 마이크로 바뀐 것을 볼 수 있습니다. TSM-2는 마이크를 교체하고, 발열이 없는 가벼운 서보모터로 Polar angle 회전을 담당하도록 변경하였습니다.

그러나 TSM-2에서도 서보모터의 알 수 없는 문제로 인해 동작 도중 틱틱거리는 소리와 함께 움직임이 발생했습니다. 이 문제는 때때로 나아졌다가 다시 발생하는데, 원인이 전력 공급 문제인지, 라즈베리 파이를 통한 PWM 제어의 하드웨어적 문제인지 확실하지 않았습니다.

연구발표자료 - 4



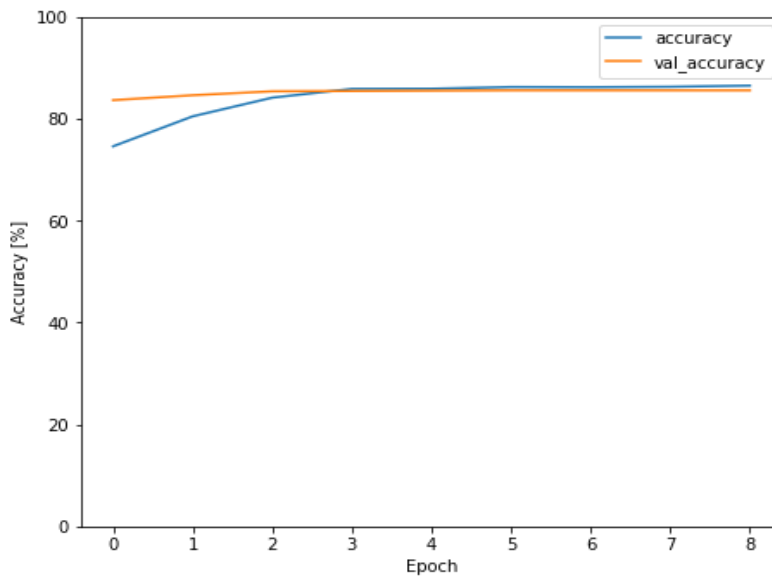
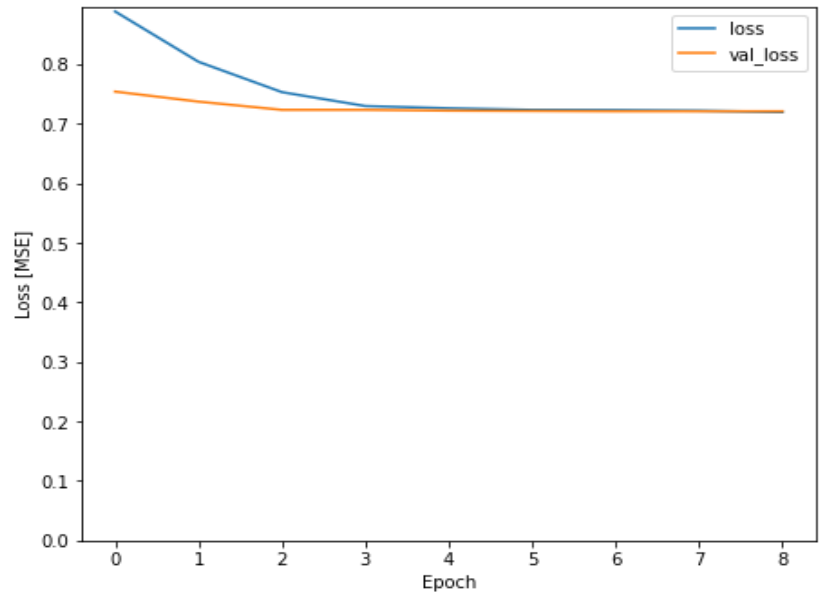
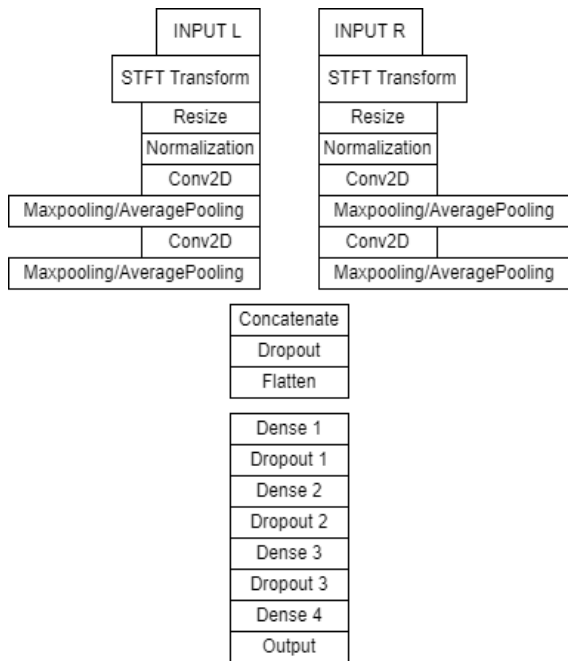
위 사진의 TSM 프로그램 코드를 만드는 과정에서 라즈베리 파이를 이용한 코드 또한 설계하고 탐구하였습니다.

저는 스텝 모터를 이용해 구면 좌표계 상에서 모든 방향의 소리를 녹음할 수 있도록 했습니다. 각각의 스텝 모터는 구면좌표계의 Azimuth angle과 Polar angle을 담당합니다. L298N 모터 컨트롤러를 이용해 스텝 모터의 각도를 제어했습니다.

소리 재생은 pygame의 audio method를 이용해 프로그램의 구성과 소리 재생을 구현했고, 마이크 입력을 통한 소리 녹음을 라즈베리 파이로 제어했습니다.

저는 이 연구를 진행하며, 데이터 수집기를 만드는데 너무 많은 시간을 들여 주객이 전도되었음을 깨달았습니다. 데이터 수집기를 만들어 데이터를 수집하기까지 무려 7개월이나 걸렸기 때문에, 인공지능에 대해 깊이 공부하고 수정해나갈 시간이 부족했습니다.

연구자료발표 - 5



CNN 모델은 좌우 마이크의 입력을 따로 받아 병렬적으로 입력 신호를 처리합니다. 처리 순서는 Input, STFT Transform, Resize, Normalize, Convolution입니다. 이후 좌우 신호를 Concatenate Layer에서 병합하여 Dense Layer를 처리하는 방식으로 구성했습니다.

이렇게 학습시킨 모델로, 귓바퀴의 구조와 딥러닝을 활용하여 약 85.5%의 정확도로 소리 방향을 구분하는 모델을 설계할 수 있었습니다.