

기술요약(선행기술조사보고서에 있는 기술요약 말고 이  
부분을 봐주시면 감사하겠습니다 :))

카메라와 마이크 두 개의 검출수단으로 사람의 표정, 행동, 목소리 톤을 검출하고, 검출한 자료를  
분석하여 어떠한 피드백을 줄지 판단합니다. 내부 스탠다드를 기준으로 현재 스피치를 진행중인  
사용자의 감정, 음성, 스피치 내용의 분야별 점수를 환산합니다. 환산된 점수별로 이에 따른 로봇의  
반응을 생성하고, 이것들과 더불어 스피치 상황을 더욱 다양하게 만들어줄 랜덤별 상황으로 묶어서  
사용자에게 출력합니다. 최종적으로 로봇은 아바타 얼굴 표정(display), 목 움직임 (motor), 언어적  
표현(speaker)를 통해서 사용자에게 표현합니다.

## 1. Perception / Recognition

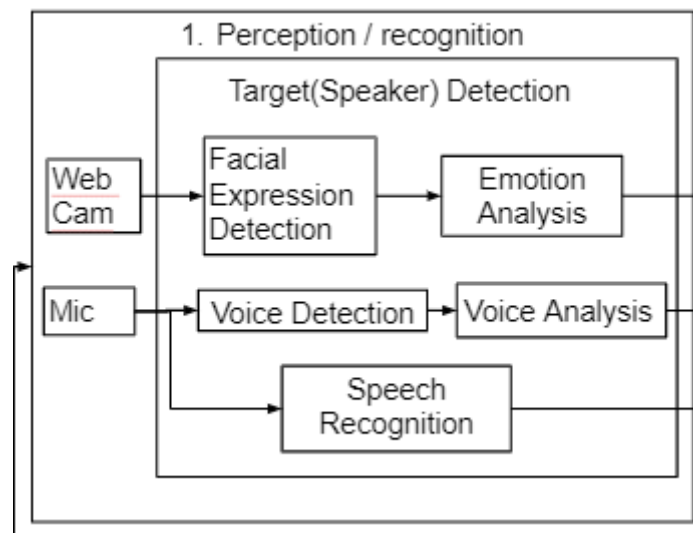
웹캠과 마이크를 통해서 user의 표정, 행동, 그리고 음성을 detect 합니다.

표정 인식의 경우, Affectiva라는 프로그램을 사용하게 되면 눈,코,입과 같은 부분을 포함한 전체적인 얼굴 전체에 34개의 점을 찍습니다. 이것을 바탕으로 눈이 흰 정도, 입꼬리가 올라간 정도와 같은 점의 위치에 따라 Joy, Anger, Sadness, Surprise, Fear와 같이 미리 정해놓은 감정들 중 적합한 감정으로 분석해냅니다.

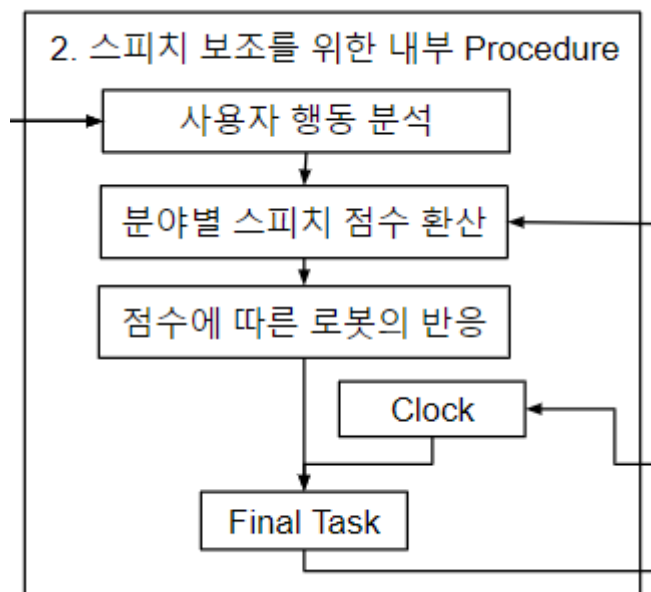
음성의 경우, 두 가지 방향으로 분석한 후 그 다음 모듈로 전달합니다. 첫 번째는 Voice Detection으로, input으로 detect된 사용자의 음성이 얼마나 끊기지 않고 답변을 하는지를 체크합니다.

예를 들어 답변을 해야하는 상황에서 유저가 당황해서 아무 말도 하지 못하는 경우 정적의 빈도 및 시간등을 체크합니다.

두 번째는 Speech Recognition 입니다. Voice Detection과는 다르게, 스피치의 내용적인 측면에서 접근합니다. 특히 그 중에서, 키워드를 인식하는데 초점을 맞추어 스피치의 방향을 파악합니다.



## 2. 스피치 보조를 위한 내부 절차 (1)



Perception / Recognition에서 파악한 사용자의 감정 분석 및 스피치, 음성 인식 내용을 종합적으로 분석하여 합니다.

분석을 통해 현재 사용자의 상태에 따른 객관적인 value값을 계산합니다.

구체적인 예시를 들자면, 사용자가 현재 모의 면접을 보고 있는 상황이라고 가정할 때 사용자가 시선은 면접관을 응시하고 있는지, 고개를 정면으로 하고 있는지, 음성 톤은 적당한지 등을 객관적인 value 값으로 파악하는 것입니다.

또한, 내부 Database의 Standard에는 사전에 미리 입력된 피면접자의 올바른 자세, 표정, 음성등에 대한 허용 기준값과 value값이 지표로 등록되어 있습니다. 이를 스피치 보조를 위한 내부 Procedure로 가져와 기준값으로서 사용자 행동 분석값과 비교하여 분야(감정, 목소리 톤, 스피치 내용)별로 스피치 점수를 환산하게 됩니다.

점수에 따른 로봇의 반응에서는 이전 작업에서 받아온 각 분야에 따른 점수값으로 로봇이 어떠한 반응(모니터 출력, 모터 움직임, 스피커 출력)을 출력할 지 결정합니다.

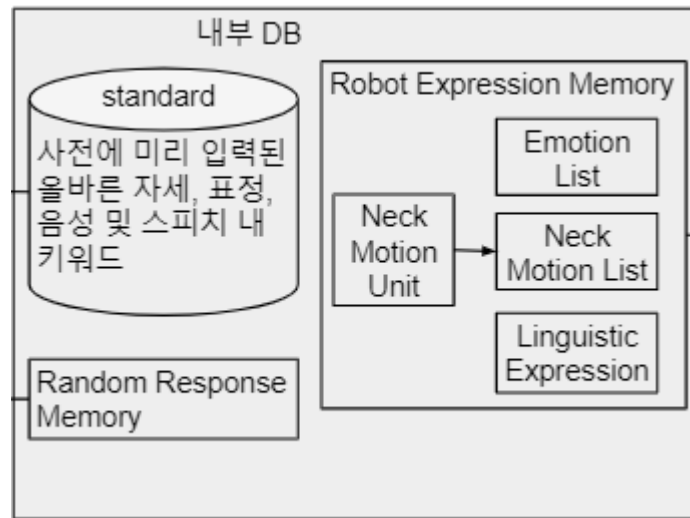
사용자의 스피치에 대한 정상적인 로봇의 반응을 출력하는 것 이외에, 로봇이 랜덤적으로 기대하지 않은 반응을 출력하여 당황스러운 상황 및 긴장감 조성을 연출하도록 내부 Database의 Random Response Memory와 Clock을 사용합니다.

이러한 랜덤 반응을 생성한 이유는 사용자가 스피치를 할 때, 자신이 의도한 사람들의 반응이 출력되면 스피치 실력 향상이 어려울 것이라 판단되었으며 당황스럽거나 난감한 상황의 발생 혹은 이러한 상황으로 인한 긴장감을 조성 하였을 경우에 사용자가 유연하게 대처할 능력을 함양하기 위해 생성하였습니다.

기본적으로 사용자의 반응에 1:1적으로 로봇이 반응하기에 평상시에는 로봇이 스피치에 대한 정상적인 반응을 Final Task로 전달하지만 기존의 설정해 둔 시간(예를 들면 10분 안에서 랜덤적으로 두번의 횟수)이 지났을 경우, Clock으로 Random Response Memory의 값이 스피치에 대한 반응으로 출력될 수 있도록 설정합니다.

최종적으로 로봇의 반응을 결정하고 이를 Expression으로 전달하기 전, 이 블록의 procedure을 통해서 나오는 결과값들은 로봇이 사용자의 반응에 대한 응답으로 분야별로 어떤 표현을 할 지가 설정된 수치값들이 됩니다.

### 3. 스피치 보조를 위한 내부 절차(2)



로봇 내 저장공간이 필요한 모든 구성요소들을 구조적 간결성을 위해 내부 데이터베이스 블록에 모두 포함하였습니다.

각 요소별로 설명을 드리겠습니다.

먼저 standard 의 경우, 사전에 미리 입력된 기준을 의미합니다.

올바른 면접자의 자세를 기준으로 각 부분을 객관적인 value값으로 변환 시켰을 때, 시선 맞춤을 잘 하고 있다는 기준값, 고개는 정면을 잘 바라보고 있다는 기준값 등을 미리 설정하여 사용자 분석이 끝난 value값과 비교합니다.

또한, 스피치 내 키워드의 경우 서비스를 사용하기전 사용자가 본인이 답변하거나 발표하는 내용에 있어서 꼭 들어갔으면 좋을 키워드를 미리 입력해 두는것을 의미합니다. 이것을 바탕으로 1번 블록 내의 Speech Recognition을 비교, 분석합니다.

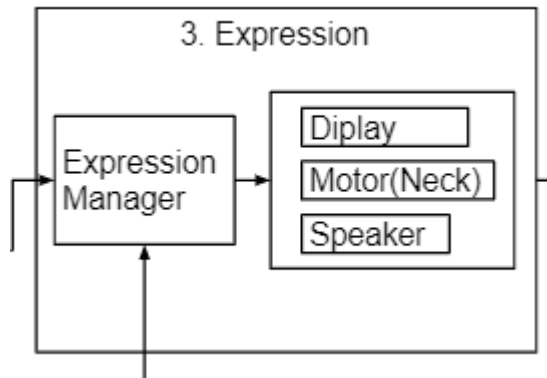
두번째 요소는 Random Response Memory입니다.

해당 메모리에서는 사용자의 스피치 보조상황을 더욱 다채롭게 만들어줄 요소들을 저장합니다. 모션, 얼굴 표정, 음성 감탄사 등을 포함합니다. 이러한 요소들은 2번 블록 내 기존에 결정된 로봇의 반응에 합하여 진 후, Clock을 거쳐 Final Task로 출력됩니다.

세번째 요소는 Robot Expression Memory입니다.

해당 메모리 안에는 감정 리스트, 목 움직임 리스트, 언어적 표현 리스트들이 포함되어 있습니다. 각 리스트들은 사전에 만들어진 것들이며, 이것들을 조합하여 로봇의 반응을 만듭니다. 그리고 그것이 3번 Expression 블록으로 최종적으로 들어가게 됩니다.

#### 4. 표현을 위한 내부 과정 절차



Expression Manager에서 2번의 Final Task로부터 표정, 모션, 언어적 표현에 대한 값을 가져오고 그 값에 따라 내부 Database의 Robot Expression Memory에서 미리 만들어 놓은 로봇의 아바타 표정, 모션, 언어적 표현을 실행할 준비를 한다. 그리고 다음 모듈로 이동하여 Expression Manager에서 로봇에게 출력하도록 결정된 사항들을 Emotion List는 Display에, Neck Motion List는 Motor(Neck)에, Linguistic Expression은 Speaker에 대응되도록 합니다.

이렇게 대응된 요소들은 로봇의 하드웨어적 요소들에 출력되어 사용자와 자연스러운 상호작용이 이루어 질 수 있도록 하며 스피치, 면접과 같은 대화 상황을 계속 이어 나갈 수 있도록 합니다.