

1. 발화자의 감정을 반영한 3D 가상인간 자동 수화 프로그램

설명 및 작동방식:

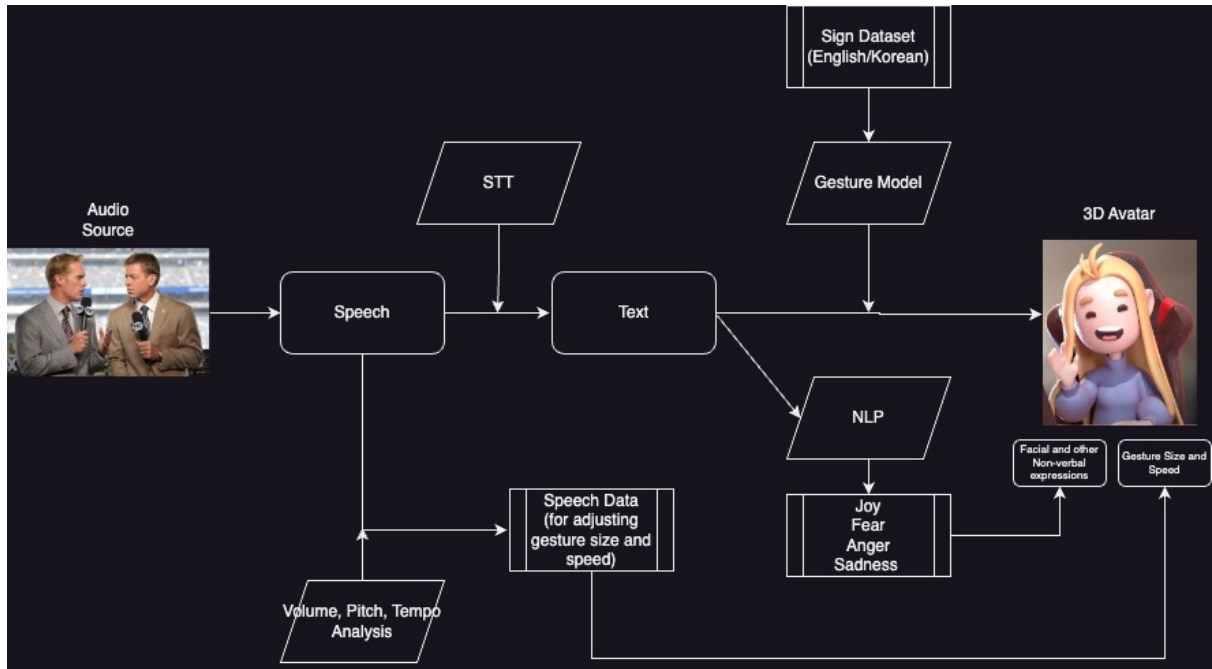
본 발명은 입력받은 음성과 텍스트로부터 발화자의 감정을 분석하여 이를 3D 가상인간의 자동 수화에 반영하는 프로그램임.

1. 발화자의 음성을 입력받고, 이를 Speech-To-Text를 통해 텍스트로 변환하는 한편, 음성 자체의 **음량과 성조, 빠르기를 분석하여 수화 제스처에 반영될 데이터를 얻어냄**. 구체적으로, 음량이 클수록, 성조가 높을수록 제스처는 커지고, 음성이 빠를수록 제스처의 속도도 빨라짐.
2. 변환된 텍스트는 일차적으로 기존의 Sign Language Translation Systems를 통해 3D Avatar Animation Module에 반영됨. 이때의 수화는 기본적인 의미를 전달할 뿐이고, 감정에 관한 내용은 반영되지 않음. 그러므로 Natural Language Processing을 통해 **발화내용을 기반으로 발화자의 감정을 추출해내어 4개의 카테고리(joy, fear, anger, sadness)로 분류함**.
3. 기존의 SLTS로 얻어낸 3D 가상인간 수화에, 1에서 얻은 음성 데이터와 2에서 얻은 감정 카테고리 데이터를 반영함. 음성 데이터는 수화 제스처의 크기와 속도에 반영되고, 감정 카테고리 데이터는 가상인간의 표정 및 비언어적 표현(눈의 모양, 눈썹의 움직임, 입의 움직임, 고개 기울임 및 고개 젓기)에 반영됨.

종래기술 및 차별성:

음성 또는 텍스트를 3D 가상인간의 수화로 변환하는 프로그램은 기존에 존재하나(HandTalk 등), 이들은 감정을 나타내는 단어("Sad", "Angry" 등)를 입력받을 시 이를 가상인간의 표정에 반영하는데에 그치고, **발화내용 그 자체에서 감정을 분석하여 반영**하거나 발화자의 음량과 성조 등을 수화에 반영하는 경우는 없었음. 수화는 자막과 달리 발화자의 감정을 담을 수 있다는 측면을 고려하면, 음성과 텍스트에서 감정을 분석하여 이를 수화에 반영할 수 있는 기술은 청각장애인의 콘텐츠 접근성에 있어서 종래기술보다 효과적일 것으로 예상되며, 이는 스포츠 중계, 교육, 음악과 엔터테인먼트 등 다양한 분야에서 활용될 수 있음.

도면:



2. 실시간 피드백 기반 음성데이터 자동수집 프로그램

설명 및 작동방식:

본 발명은 STT 모델 학습용 음성데이터 수집 과정에서 상용 STT 모델을 활용하여 참여자에게 누락된 단어나 음절에 대해 즉각적인 피드백을 주어 수정하게 함으로써, 데이터 수집 프로세스를 자동화, 간소화하는 프로그램임.

1. 원시 텍스트를 입력받아 문장단위로 분할함.
2. 분할 텍스트를 발화자가 보는 화면에 띄움. 발화자는 텍스트를 읽고 발화함. 발화음성은 녹음되어 저장됨.
3. 발화음성을 상용 음성인식 모델(정확도 낮음)로 검토하여, 누락된 단어나 음절이 있을 시 이를 발화자 화면에 표시함. (텍스트는 “오월의 하늘은 티끌도 없다”인데 발화음성은 “오월의 하늘은 끌도 없다”인 경우, 문장에서 ‘티’를 다른 색깔로 표시)
4. 위의 발화자는 자신의 음성을 다시 재생하여 듣고, 모델의 피드백대로 자신의 음성을 재 녹음함.
5. 발화자에 의해 확정된 음성은 텍스트와 매칭되어 데이터셋에 저장됨.

종래기술 및 차별성:

종래의 음성데이터 수집 방식은 녹음 단계에서 발화자의 발화를 감독하고 오류사항을 수정하는 감독자를 필요로 하는 경우가 많고, 감독자를 두지 않을 시 이후 녹음 검수와 전사(labeling) 단계에서 많은 인력과 비용을 필요로 하는 경우가 많음. 해당 프로그램을 사용 시, 발화자가 단독으로 자신의 음성을 녹음하고 상용 STT 모델로부터의 피드백을 기반으로 자신의 발화를 수정하므로, 추가적인 감독이나 전사를 필요로 하지 않음. 이에 따라 원시 텍스트에 대한 음성데이터 수집 비용을 감축하는 효과가 기대됨.

도면:

