문제점 요약

ㄴ 애니메이션에서 음성이 드러나는 장면은 중요하며 시간과 비용이 오래 걸리는 작업이다.

음성 애니메이션이 부자연스럽게 적용된다면 사람이 말하는 것처럼 자연스럽게 보인다기보다는 동기화되지 않은 로봇처럼, 딱딱한 연설을 하는 것처럼 보일 수 있다.

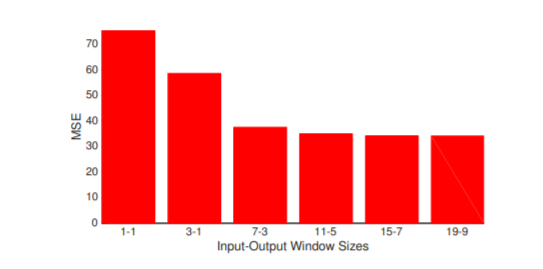
또한, 표정을 제대로 이해하지 못하고 애니메이션을 삽입한다면 움직이는 얼굴이 사람들의 주의를 산만하게 만들 수 있고 이목구비의 위치가 틀어지면 불쾌한 골짜기 효과(출처)를 일으키기도 한다.

다른 언어와 싱크가 맞는 음성 애니메이션을 여러 개 제작하는 방법도 있다.

ㄴ 그러나 이는 많은 비용을 필요로 하며, 모션 캡처&전문 애니메이터들로 구성된 대규모 팀들이 필요하다. 모션 캡처 방식은 캡처의 대상이 되는 배우를 필요로 하고 편집이 복잡하며, 대규모 팀의 핸드 애니메이션은 많은 비용과 시간이 소모된다.

딥 러닝 음성 애니메이션을 사용하면 이를 전부 보완할 수 있다.

음성을 입력하면, 입력된 음성에 맞추어 애니메이션이 생성되기에 비용이 절감되고, 프로그램은 정보를 입력할수록 오차가 줄어든다.

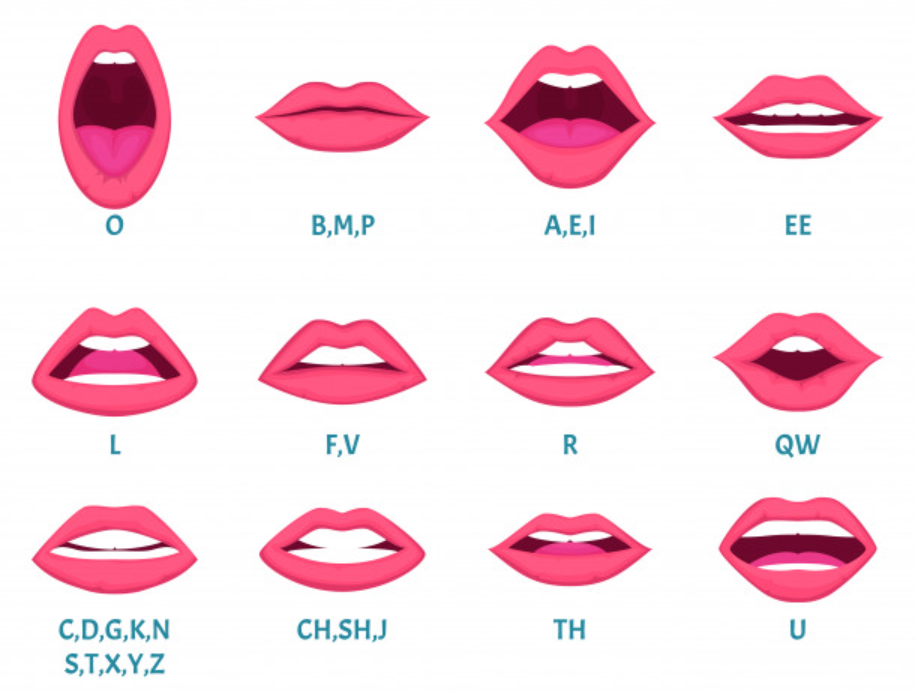


A Deep Learning Approach for Generalized Speech Animation 그림 11의 오차 그래프

후반으로 갈수록 오차 범위의 그래프가 줄어드는 것을 알 수 있다.

이 기술이 필요한 이유

애니메이션 입 모양



이미지 출처: freepik 무료 리소스

2D 애니메이션은 기본적으로 아,에,이,오,우 5종류 정도만 사용하여 단순하게 애니메이팅한다.

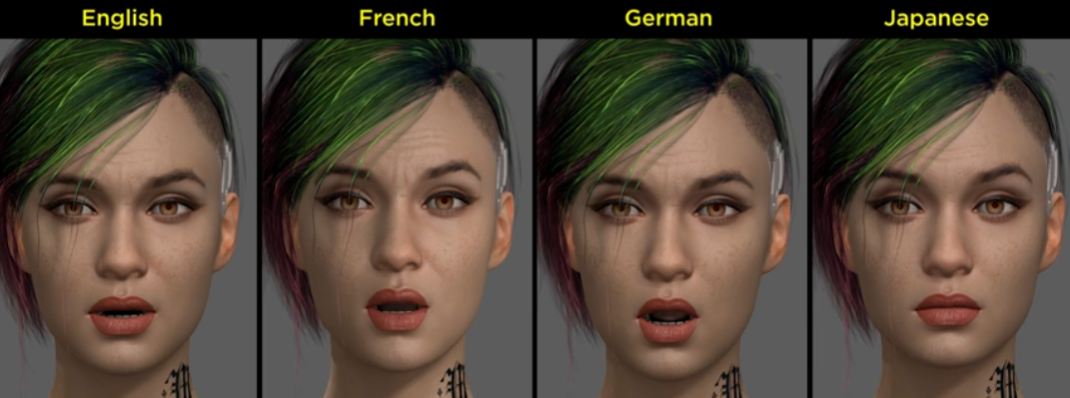
그러나 3D 애니메이션들이 흥행함에 따라 2D보다 훨씬 많고 자연스러운 프레임을 요구하게 되었고, 언어가 다른 국가들로 수출하게 되자 애니메이션의 입 모습과 소리의 싱크로가 맞지 않는 문제가 발생하게 된다.



이미지 출처: 오버워치 공식 유튜브 채널

위 이미지는 오버워치 단편 애니메이션:슈팅 스타 장면 중 한 장면이다.

6분 48초 더빙을 보면 북미 대사를 가정하고 만든 애니메이션과 한국어 더빙의 싱크로가 전혀 맞지 않음을 볼 수 있다.



이미지 출처: JALI Research Inc 공식 유튜브 채널

그러나 딥 러닝 애니메이션을 사용하면, 각 대사를 여러 언어에 맞추어 애니메이팅 할 수 있고, 영상을 보는/게임을 플레이하는 사람들에게 더 자연스러움을 제공한다. 위 이미지는 게임 ‘사이버펑크 2077’에 적용된 음성 애니메이션 시연 장면으로, 말을 할 때 바뀌는 미묘한 얼굴 근육의 움직임까지 재현해내었다.

논문 기술 요약

A Deep Learning Approach for Generalized Speech Animation - 입력 음성과 동기화되는 자연스러운 음성 애니메이션을 자동으로 생성하는 딥 러닝 접근 방식이다. 자연스러운 움직임과 시각적으로 조화로운 효과를 정확하게 포착하는 방식으로 음소 레이블 입력 시퀀스에서 입 움직임까지 임의의 비선형 매핑을 학습하는 슬라이딩 윈도우 예측기를 사용한다. 딥 러닝은 실시간으로 실행되고, 매개 변수를 많이 조정할 필요가 없다. 새로운 음성 시퀀스에도 적용 가능하며 감정적인 음성 생성 방식에 유용하고 기존 애니메이션의 리타게팅(*애니메이션 리타게팅: 같은 스켈레톤(뼈대)을 공유하지 않는 캐릭터임에도 애니메이션을 재사용할 수 있는 기능*) 방식과 호환된다. 이 작업의 중요한 초점 중 하나는 기존 애니메이션 프로덕션 파이프 라인(*동시 제작*)과 통합 가능한 접근 방식을 개발하는 것이다. 딥 러닝 애니메이션은 외국어 입력 및 노래, 다양한 나잇대의 음성 등 광범위한 애니메이션 클립을 통해 시연될 수 있고, 음성 입력 후 애니메이션 생성 방식 뿐만 아니라 실시간 주문형 음성 애니메이션을 생성할 수도 있다.

음성 인식 애니메이션의 교육 단계

1. 레퍼런스 스피커의 오디오 및 비디오 녹화 (균형 잡힌 음계의 문장 모음 사용)
2. 스피커의 얼굴을 추적, 파라메타하여 생성 (음성을 입력할 기준 얼굴 애니메이션 모델)
3. 오디오를 음소 라벨 시퀀스 x로 변환
4. 예측자 h(x)를 훈련시켜 x에서 해당 x로 매핑 (애니메이션 매개 변수)
5. 대상 재설정 함수를 문자 cg모델로 사전 계산 (기존의 대상 재설정 기법 사용)

애니메이션:

1. 입력 오디오를 음소 시퀀스 x로 변환
2. h(x)를 이용해 x에 대응하는 기준면 모델의 애니메이션 파라미터를 예측
3. 기준면 모델에서 대상 cg로 재대상화 (여러 모델로 재사용이 가능)

교육 1~4단계는 한 번만 수행되며, 모든 사례 5단계는 새로운 각 단계하마다 한 번씩 사전 계산되어야 한다.