

# 대화형 가상 현실에서 아바타의 립싱크

이재현 박경주

중앙대학교

hjl101076@cau.ac.kr, kjpark@cau.ac.kr

## Avatar's Lip Synchronization in Talking Involved Virtual Reality

Jae Hyun Lee, Kyoungju Park

### 요 약

가상 현실(VR)에서 사용자와 동일한 아바타를 시각화하는 것은 몰입도를 증가시킨다. 가상 현실 기술이 발달함에 따라 여러 사용자들의 가상 소셜 네트워킹과 말하는 아바타를 포함하는 다양한 어플리케이션이 늘어난다. 보급형 가상현실 기기에서 사용자의 얼굴 전체와 몸 동작 캡처를 위한 디바이스 정보가 부족하기 때문에 대부분의 가상현실 어플리케이션은 대화하는 얼굴과 몸을 나타내지 않는다. 우리는 오디오와 동기화되어 대화하는 가상 얼굴을 위해, 가상현실 어플리케이션을 대상으로 하는, 새로운 방법을 제안한다. 우리의 시스템은 단일 사용자 어플리케이션에서 사용자 자신과 동기화된 아바타를 실험한다. 단일 사용자 환경에서는 자신의 아바타를 미러링 하고 다중 사용자 환경에서 동기화된 대화 파트너를 시각화 하여 동기화된 대화 아바타의 구현한다. 우리는 사실적으로 대화하는 아바타가 동기화되지 않은 말하는 아바타나 보이지 않는 아바타보다 더 영향력이 있다는 것을 사용자 스터디로 검증한다.

### Abstract

Having a virtual talking face along with a virtual body increases immersion in VR applications. As virtual reality (VR) techniques develop, various applications are increasing including multi-user social networking and education applications that involve talking avatars. Due to a lack of sensory information for full face and body motion capture in consumer-grade VR, most VR applications do not show a synced talking face and body. We propose a novel method, targeted for VR applications, for talking face synced with audio with an upper-body inverse kinematics. Our system presents a mirrored avatar of a user himself in single-user applications. We implement the mirroring in a single user environment and by visualizing a synced conversational partner in multi-user environment. We found that a realistic talking face avatar is more influential than an un-synced talking avatar or an invisible avatar.

**키워드:** 인간-중심 컴퓨팅, 가상현실, 얼굴 애니메이션, 아바타, 립싱크

**Keywords:** Human-centered computing, virtual reality, face animation, avatar, lip synchronization

## 1. 서론

가상현실(VR)에서 존재감을 느끼는 중요한 요소 중

\*corresponding author: Kyoung ju Park/chung-ang University(kjpark@cau.ac.kr)

하나는 몰입감이다[1,2]. 기존 연구에서 널리 알려진 바와 같이, 자기 아바타는 제어 가능한 신체 부분을 가상 현실에 도입하여 체화감을 향상시키고, 몰입도를 불러일으킨다[3]. 따라서 사용자와 동기화하여 움직이는 아바타는 사용자의 몰입도를 높이는 데 도움을 줄 수 있다. 최근의 가상현실 아바타 연구는 상체와 손의 시각화에 초점을 맞추고 있다. 말하는 가상현실 어플리케이션이 증가함과 더불어, 사용자 얼굴과 동기화된 말하는 아바타는 점점 더 중요해지고 있다.

가상현실에서 사용자의 아바타 얼굴을 사실적으로 생성하기 위해서는 HMD(Head Mounted Display)를 착용한 사용자의 얼굴 전체를 캡처해야 하는데, HMD가 얼굴의 거의 절반을 덮기 때문에 불가능하다. 그러므로 어플리케이션 개발자들은 현실적인 애니메이션 없이 가상 얼굴을 시각화 하고, 모션 캡처 시스템을 사용하여 정확하게 애니메이션 할 수 있는 신체 부위를 표시한다. 대화형 가상현실 어플리케이션에서는 실제 얼굴의 현실성은 립싱크와 얼굴 방향에 크게 좌우된다는 점 때문에, 말하는 내용을 기반으로 립싱크 함으로써 아바타의 얼굴을 표현하고 얼굴 방향을 연출하는 것이 필요하다.

기존의 실감 립싱크 방식은 고정밀 얼굴 모델을 애니메이션하는 것이어서 모델 용량이 크다 보니 모바일VR 장비에서는 초당 80프레임의 실시간 실행이 불가능하다. 따라서, 우리는 립싱크 방법을 단순하게 하고 얼굴 모델 폴리곤 개수도 최소화하는 방법을 제안한다. 이를 위해 우리는 대표적인 몇 가지 발음을 선별하고, 그에 해당하는 얼굴 정지 표정이 아니라 얼굴 애니메이션을 생성하여 데이터로 저장한다. 가상현실 서비스 도중에 소리와 텍스트가 입력되면 매칭되는 얼굴 애니메이션을 선택하고 해당 애니메이션의 길이를 조절함으로써 립싱크 애니메이션을 수행한다.

본 논문은 소리와 텍스트에서 가상의 입술을 애니메이션화하고, 말하는 얼굴을 가상현실에서 제작하는 방법을 제안한다. 우리의 입-애니메이션은 기존의 립싱크 방법의 발음과 발화 시간을 고려하여 활용하고

수정한다. 그 결과 나오는 립싱크 애니메이션은 가상현실 환경에서 아바타의 말하는 얼굴을 만들어낸다. 우리는 가상 얼굴이 자신과 관련된 가상현실 어플리케이션 미러링과 원격 보존 아바타를 포함한 효과와 영향에 대해 조사한다. 또한, 우리는 말하는 아바타를 가지고 있거나 가지고 있지 않은 사용자의 콘텐츠 인식을 평가한다. 이를 통해 말하는 아바타가 사용자의 단일 환경에 대한 몰입도에 얼마나 영향을 미치는지 알아본다.

## 2. 관련 연구

최근에는 가상 환경에서 HMD의 카메라를 설치하여 사용자의 입모양과 아바타의 입모양이 동일하게 움직이는 기술이 개발되었다[4]. 전에는 아바타의 입의 움직임과 사운드를 일치시키기 위하여 Generative Adversarial Networks(GAN)로 표현해야 하는 방법이 시도되고 있다[5]. 또한 Visemes 기반의 음성에 대한 연구 방법은 사운드로 정확한 입술 움직임을 달성한다[6,7,8]. 입 애니메이션 기술들은 아바타를 이용한 채팅이나 화상회의 등에서도 응용 가능한 기술로서 가상현실 개념을 포함하여 가상의 아바타가 마치 직접 사람과 대화를 나누는 등의 실제로 많이 쓰이고 있다[9]. 다른 방법에는 음성 신호에 맞춰 아바타의 입 모양을 발음에 맞게 변화를 시키거나 일부 문장의 뜻을 이해하여 그에 맞는 표정을 짓게 함으로써 상대방에게 더욱 실감나게 의미를 전달할 수 있고 [10,11], 실제 사람을 컴퓨터 그래픽 방법 대신 신경 모듈로 구성하며 오디오와 입의 특징점을 생성하여 순간마다 적절한 3D 입술의 일치로 실제로 말하는 것처럼 보여 지도록 한다[12,13].

## 3. 시스템 개요

우리는 사용자가 제어하는 아바타의 모습을 1인칭으로만 보는 것이 아니라 미러링 되도록 만든다. 3인칭으로 만드는 것은 사용자의 몰입을 감소시킨다. 1인칭으로 사용자가 아바타를 제어하는 것이 모든 실험에서 좋다는 결과가 있다[14]. 1인칭으로는 사용자가 직접 아바타의 얼굴 움직임을 확인할 수 있도록 거울처럼 볼 수 있는 미러링 효과와 함께 실험 환경을 세팅한다.

#### 4. 아바타 립싱크

가상현실에서 사용자와 동일한 아바타를 시각화 하는 것은 몰입감을 높인다. 가상현실에서 사용자의 동일한 아바타를 제어하기 위해서 HMD의 카메라를 부착하여 사용자의 동작을 기반으로 아바타를 생성한다. 우리는 카메라 없이 HMD와 마이크만으로 사용자와 아바타의 동일한 움직임으로 저비용 환경에서도 몰입감을 높일 수 있는 것을 보여준다.

우리의 방법은 Figure 1에서 보이는 것과 같다. 우선, 사용자의 음성을 STT(speech-to-text)방법[15]으로 변환하여 텍스트로 변환된 음절 데이터를 획득한다. 사용자 음절의 발화 시간을 추정한다. 입력된 데이터가 한국어인지 다른 나라의 언어인지 분석하여 한국어 발음으로 변환한다. 발음에 적합한 아바타의 입 애니메이션을 데이터 셋 중에서 선택하고, 발화 시간에 따라서 립싱크 애니메이션의 길이를 조절한다.

사용자 음성을 STT 방법으로 텍스트로 변환하여 음절을 얻어온다. 우리는 범용성을 위해 한국어 발음으로 변환하였다. 각 나라 언어의 고유한 모국어 발음을 처리하기 위해 한국어 발음으로 변환한다. 또한, 한국어는 한자 소리에서 유래되었기 때문에 한자 또한 가능하다. 한국어 발음은 국제음성기호 표처럼 모든 나라의 언어를 처리할 수 있다. 한국어의 발음은 유성음과 무성음이 존재하기 때문에 [f], [p], [v]와 [b]처럼 구분이 힘든 발음을 정확하게 판단할 수 있다. 한국어의 발음은 발음 기관의 모양을 본떠 만들었기 때문에 마찰음과 발음 규칙이 적지만 발음의 정확성을 보여주어 입모양을 생성하는 것에 있어서 적합하다.

각 음절의 발화 시간을 추정한다. 우리는 음절 개수로 발화 시간을 계산할 수 있다. 음절 발화 시간과 음소 간의 시간 차이를 자동으로 정확하게 획득하는 것은 어렵기 때문에[16], 문장에서 음절 개수를 기반으로 평균 음절 발화 시간을 결정한다. 평균 음절 발화 시간 추정 수식은 다음과 같다:

$$T_M = (T_e - T_s) / N \quad (1)$$

음절의 평균 발화 시간  $T_M$ , 발화 끝 시간  $T_e$ , 발화 시작 시간  $T_s$ , 음절의 개수는  $N$ 이다. 발화 끝 시간과 시작 시간의 차이를 음절의 개수로 나누어 음절의

평균 발화 시간을 추정한다. Figure 2은 각 음절을 한국어 발음으로 변환하고 발화의 시작과 끝 시간을 추정한 결과이다. Figure 2에서 ori는 음절의 원본 데이터, ko는 한국어 발음, vol은 모음, cons는 자음으로써 초성의 자음만 확인한다. start는 단어의 발화 시작 시간 end는 끝 시간으로, 단위는 밀리초(ms)이다. Figure 2의 왼쪽은 hey는 영어 입력을 한국어 발음으로 변환하고 모음과 자음을 얻은 결과이며, 오른쪽은 너라는 한국어 입력에 대한 변환 결과이다.

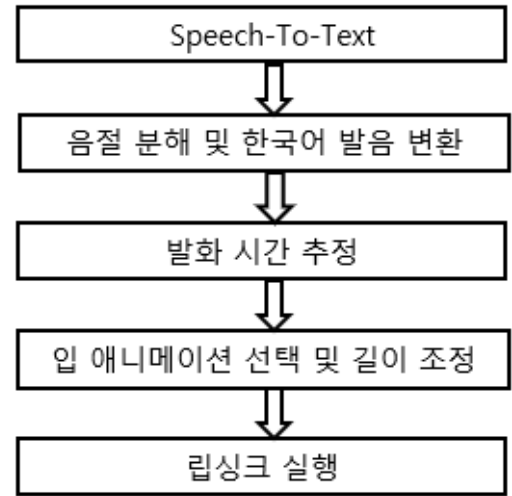


Figure 1. Method structural theory

"ori": "hey",	"ori": "너 ",
"ko": "헤이",	"ko": "너 ",
"start": 50110,	"start": 49430,
"cons": "ㅎ,ㅇ",	"cons": "ㄴ",
"vol": "에,ㅣ",	"vol": "ㅡ",
"end": 50340	"end": 50110

Figure 2. Left Picture: Data in English, Right Picture: Data in Korean

한국어 발음의 입 모양을 하는 방법은 기존에 많이 있는데, 우리는 저비용 가상 현실에서 적용할 수 있게 최소한의 입 모양 개수만을 사용하고자 한다. 또한 발음에 따른 정적인 입모양이 아니라 동적으로 변하는 입 동작 데이터를 구축한다. 따라서, 입모양을 생성하기 위해 한국어의 21개의 모음을 유사한 발음 기호를 Figure 3처럼 9 개로 지정한다. 모음 ‘ㅏ’는 발음 기호 AA, ‘ㅑ,ㅓ,ㅕ,ㅗ’는 AE, ‘ㅓ,ㅕ’는 YA, ‘ㅗ’는 EO, ‘ㅛ,ㅜ,ㅟ,ㅠ’는 OH, ‘ㅜ,ㅞ,ㅟ’는 WA, ‘ㅟ,ㅠ,ㅡ’는 WO, ‘ㅡ,ㅣ’는 EU, ‘ㅣ’는 EE로 9개의 발음 기호로

처리한다. figure 4는 9개의 발음 기호에 따른 구축한 입 모양이다. 발음마다 입 모양의 길이가 서로 다르다. AA는 0.28초, AE는 0.35초, YA는 0.26초, EO는 0.19초, OH는 0.25초, WA는 0.22초, WO는 0.24초, EU는 0.32초, EE는 0.31초로, 총 9개의 입 동작 데이터를 구축한다. 입 동작 데이터를 사용함으로써 9개의 기본 발음이지만 자연스러운 결과를 도출할 수 있다.

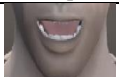
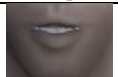
립싱크 애니메이션은 발음기호에 따라 적합한 립 동작을 데이터에서 선택하고, 순음과 비순음에 따라서 립동작 데이터를 수정하고, 발화시간에 따라 길이를 조정함으로써 애니메이션을 생성한다.

입 모양은 모음에 따라서 크게 달라지기 때문에 음절의 중성, 즉 모음에 따른 발음 기호는 Figure 3에서 선택되고, 그에 따른 입 동작을 Figure 4의 발음 기호에 대한 입 애니메이션을 결정한다.

입 동작을 선택한 후에, 순음인 경우에는 입 동작의 첫 움직임의 입모양은 달라지기에 도입 부분의 입 동작을 수정한다. 이를 위해 초성이 순음(‘ㄱ’, ‘ㄴ’, ‘ㅇ’) 인지 분석하고 입을 다물고 있는 정적인 이미지를 애니메이션 도입 부분에 추가한다. 나머지 자음들은 입을 벌린 상태에서 발음을 이어가지만, 3개의 자음은 입을 다물었다 벌리기 때문이다. Figure 5에서 (a)는 순음인 경우, (b)는 비순음인 경우이다. 순음인 경우에는 (a)의 맨 왼쪽처럼 입을 다문 모양이 추가어진다. 자음의 순음 비순음 구분은 립싱크 애니메이션을 구현하는데 있어 자연스러움을 높이기 위해 구분한다.

Korean Pronunciation	Phonemes	Korean Pronunciation	Phonemes
ㄱ	AA	ㄱ, ㄴ, ㄷ	WA
ㅏ, ㅑ, ㅓ, ㅕ	AE	ㅏ, ㅑ, ㅓ, ㅕ	WO
ㅗ, ㅛ	YA	ㅡ, ㅣ	EU
ㅜ	EO	ㅣ	EE
ㅗ, ㅛ, ㅜ, ㅠ	OH		

Figure 3. Phonemes in Korean vowels

Shape	Phonemes	Shape	Phonemes
	AA		WA

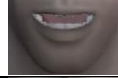
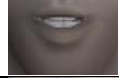


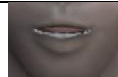
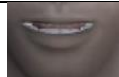
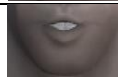
	AE		WO
	YA		EU
	EO		EE
	OH		

Figure 4. Mouth shapes by phoneme

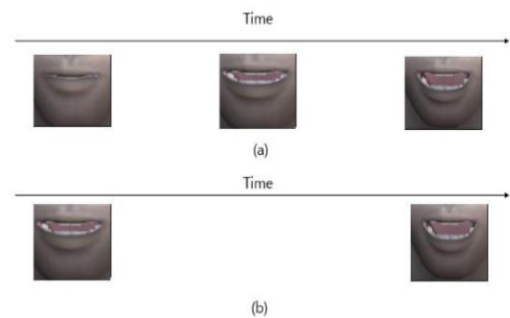


Figure 5. (a) Differences in the mouth shapes among consonants (‘ㄱ’, ‘ㄴ’, ‘ㅇ’) from the left to the right (b) Differences in the mouth shapes between ‘ㅏ’ and ‘ㅑ’

입 동작이 정해진 후에, 이 동작의 길이를 조정한다. 음절의 발화 시간에 따라서 입 동작의 길이를 줄이거나 늘린다. Figure 6은 유니티에서 6음절을 발화하고 난 뒤, 각 애니메이션의 길이를 출력한 것이다. Figure 6은 ‘립싱크의 방식’이라고 말하고, 발화 음성에 대한 애니메이션 길이이다. 발음 기호는 EE-EE-EU-EU-AA-EE 순서로, 애니메이션의 길이는 468ms-465ms-468ms-465ms-221ms-223ms이다. 발음 기호가 같더라도 발화 시간에 따라 애니메이션의 길이가 달라지는 것을 볼 수 있다.

```
[13:24:16] 애니메이션 길이 : 468.5858
UnityEngine.Debug:Log(Object)
[13:24:16] 애니메이션 길이 : 465.7104
UnityEngine.Debug:Log(Object)
[13:24:16] 애니메이션 길이 : 468.5858
UnityEngine.Debug:Log(Object)
[13:24:16] 애니메이션 길이 : 465.7104
UnityEngine.Debug:Log(Object)
[13:24:16] 애니메이션 길이 : 221.2109
UnityEngine.Debug:Log(Object)
[13:24:16] 애니메이션 길이 : 223.5861
UnityEngine.Debug:Log(Object)
```

Figure 6. Difference length of the animation the same pronounce.



## 6. 실험 결과

실험 환경은 Unity 2019.2.1f1 버전을 사용하고 파이썬 3.6과 소켓 통신으로 연동하여 구현한다. 음성 인식은 Google Speech API[17]를 사용한다.



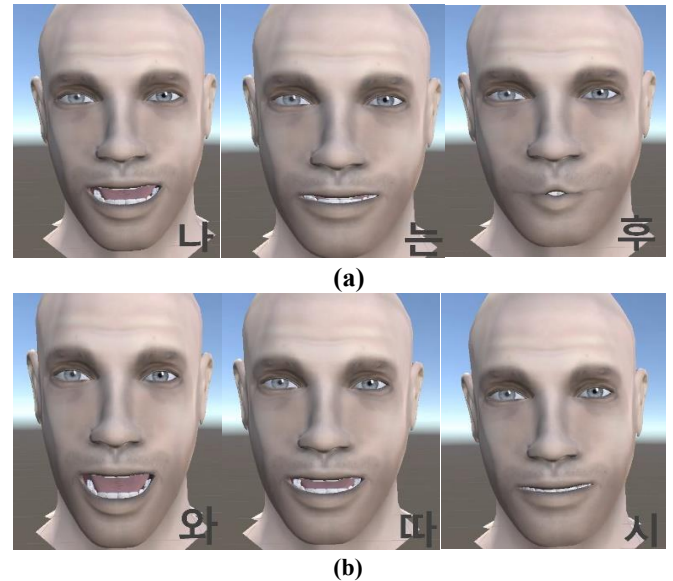
**Figure 7.** A first-person view of the scene that a user controls his/her avatar mirrored view

우리는 제안한 방식으로 1인칭 가상 현실 어플리케이션 환경속에서 실험한다. 실험은 사용자 혼자만 들어갈 수 있는 공간으로 이루어진 1인칭 가상현실 어플리케이션으로 Figure 7에서 보이는 것처럼 미러링 방식으로 사용자가 자신의 아바타를 보면서 컨트롤 할 수 있게 한다.

가상 노래방 환경은 개인 또는 상대방과 상호 작용할 수 있는 대화형 환경이다. 사용자는 1인칭 가상현실 미러링 어플리케이션에서 노래를 부를 때 자신의 입모양과 아바타의 입모양이 일치하는지 정확한 판단을 할 수 있다. 가상 노래방 환경은 우리가 제안한 방식을 확인할 수 있기에 적합하다.

우리의 립싱크 기술은 모든 언어를 처리할 수 있다. 자음까지 판단하여 21개의 모음과 4개의 자음으로 판단한다. 자음의 3개는 각각 순음이며 그 외의 자음은 통합하여 처리한다. 다른 나라의 언어를 한국어 발음으로 바꾼다. 한국어의 특성상 모든 언어를 한글로 표현하고 발음할 수 있는 장점이 있다. Figure 8에서는 한국어, 영어, 일본어 세 개의 언어들로 실험한 결과이다. Figure 8 (a)의 왼쪽과 중간은 한국어 ‘나-는’에 대한 발음 기호 ‘AA-EU’ 입 애니메이션을 적용하고, (a)의 오른쪽은 ‘who’를 한국어 발음 ‘후’로 변환한 뒤, ‘OH’ 입 애니메이션을 적용한다. Figure 8 (b)는

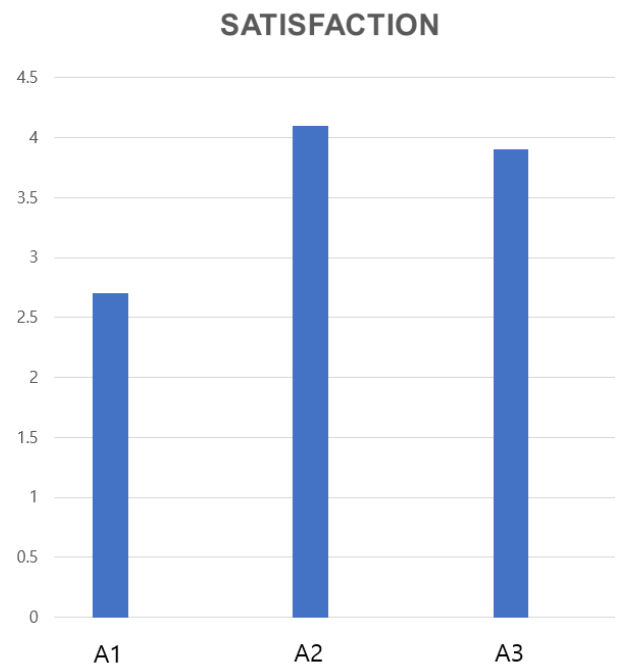
일본어 ‘私’를 한국어 발음 ‘와따시’로 변환하고 ‘WA-AA-EE’ 입 애니메이션을 적용한다.



**Figure 8.** (a) Left, Middle: Korean voice ‘나는’ to text, Right: English voice ‘who’ to text and (b) Japanese voice ‘私’ to text and all voice to Korean pronunciation and apply animation

## 7. 결론

가상 아바타의 얼굴을 표현하는 것이 사용자의 몰입감을 증대하는데 영향을 어떻게 끼치고 있는지를 파악하기 위하여 설문 조사를 수행한다.



**Figure 9.** Average satisfaction for avatar (A1: 2.8, A2: 4.1, A3: 3.9)

아바타는 사용자들이 작업 수행에서 느끼는 관점과 인식을 평가하기 위해 설정한다. 3가지의 아바타를 설정한다. 입모양이 랜덤하게 움직이는 아바타, Face FX 기반 아바타, 본 논문 방법의 아바타이다. 랜덤하게 움직이는 아바타는 발음 기호에 상관없이 무작위의 애니메이션이 진행된다. Face FX 기반 아바타는 기본적인 립싱크 방법이다[18]. 아바타는 다음과 같이 정의한다:

**A1:** 입모양이 랜덤하게 움직이는 아바타

**A2:** Face FX 기반 아바타

**A3:** 본 논문 방법의 아바타

설문 조사는 아바타의 따른 몰입도를 측정하기 위함이다[19]. E1~E3는 사용자가 느낀 인터랙션의 반응 속도를 확인한다. 여기서의 인터랙션의 반응 속도는 사용자가 말하고 나서, 얼마나 빠르게 응답하여 얼굴의 입모양이 변하는 것을 말한다. E4~E5는 사용자가 실험하면서 느꼈던 점을 파악한다. 결과를 도출하기 위해 Likert 5점 척도를 활용한다. 설문지는 다음과 같다:

**E1:** 시작하거나 수행한 작업에 환경이 얼마나 반응이 좋습니까?

**E2:** 환경을 통한 움직임을 제어하는 메커니즘은 얼마나 자연스럽습니까?

**E3:** 환경과의 상호 작용은 얼마나 자연스럽습니까?

**E4:** 가상 환경 경험에 얼마나 몰입했습니까?

**E5:** 할당된 작업이나 필요한 작업에 사용되는 메커니즘에는 집중할 수 있습니까?

우리의 방법을 사용한 아바타는 효과적이다. Figure 9은 아바타의 따른 평균 만족도 결과이다. 남성 6명, 여성 4명을 상대로 총 10명에게 실험한다. A1의 만족도는 2.8점, A2의 만족도는 4.1점, A3의 만족도는 3.9점이다. 본 논문 방법 기반 아바타의 만족도는 랜덤하게 움직이는 아바타보다 높고, Face FX 기반 아바타와 차이는 0.2점으로, 차이는 미세하다.

우리가 제안한 방식으로 립싱크한 아바타의 가상 현실 응용은 효과적이다. HMD로 사용자와 아바타의 머리를 동기화 하는 것에 추가로, 음성 기반 아바타 립싱크로 사용자와 아바타의 입 모양도 동기화함으로

써, 사용자의 몰입감을 높인다. 우리의 기술은 범용성을 위해 각 나라의 언어를 한국어 발음으로 처리하여 모든 나라의 언어가 처리 가능하며 사용자의 음성을 인식하고 발음과 발화 시간까지 계산하는 방법으로 저비용 가상현실에서도 자연스러운 립싱크가 가능한 기술이다.

## 참고문헌

- [1] K. Kiltner, R. Groten, and M. Slater. "The Sense of Embodiment in Virtual Reality". Presence: Teleoperators and Virtual Environments, 21(4):373–87, 2012.
- [2] M. Slater and A. Steed. "A virtual presence counter. Presence: Teleoperators and Virtual Environments", 9(5):413–434, 2000.
- [3] M. Parger, J. H. Mueller, D. Schmalstieg, and M. Steinberger, "Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality", Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology-VRST '18, pp. 1–10, 2018. E. Catmull. A tutorial on compensation tables. In Computer Graphics, volume 13, pages 1–7. ACM SIGGRAPH, 1979.
- [4] K. Olszewski, J.-J. Lim, S. Saito, and H. Li, "High-Fidelity Facial and Speech Animation for VR HMDs", ACM Trans Graph, Vol.35, No. 6, 2016
- [5] K. Vougioukas, S. Petridis, and M. Pantic. "Realistic Speech-Driven Facial Animation with GANs", Computer Vision and Pattern Recognition, pp. 1–16, 2019.
- [6] T. Frank, M. Hoch, and G. Trogemann, "Automated lip-sync for 3d-character animation", 15th IMACS World Congress on Scientific Computation Modelling and Applied Mathematics, August 1997.
- [7] L. N. Hoon, K. A. A. Rahman, and W. Y. Chai, "Framework development of real-time lip sync animation on viseme based human speech", Jurnal Teknologi, Vol. 75, No. 4, pp. 43–48, 2015.
- [8] S. H. Park, S. H. Ji, D. S. Ryu, and H.G. Ch, "A new cognition-based chat system for avatar agents in virtual space", Proceedings of the 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry Article No. 13-VRCAI '08, 2008.
- [9] Y. LEE, D. Terzopoulos, and K. Waters, "Realistic Modeling for Facial Animation", In: Computer Graphics (SIGGRAPH Proc.), pp 55–62, 1995.
- [10] H. C. Yehia, T. Kuratate, and E. Vatikiotis-Bateson, "Linking facial animation, head motion and speech acoustics", Journal of Phonetics, 2002.
- [11] R. Kumar, J. Sotelo, K. Kumar, A. de-Brebisson, and Y. Bengio, "Obamanet: Photo-realistic lip-sync from text". In NIPS 2017 Workshop on Machine Learning for Creativity and Design, 2017.
- [12] S. Suwajanakorn, S. M. Seitz, I. Kemelmacher-Shlizerman, "Synthesizing Obama: Learning Lip Sync from Audio", ACM Transactions on Graphics (TOG), Vol. 36, No. 4, July 2017. D. Roth, J. L. Lugin, J. Büser, G. Bente, A. Fuhrmann, and Marc E. Latoschik, "A simplified inverse kinematic approach for embodied vr applications", In in Proceedings of the 23rd IEEE Virtual Reality (IEEE VR) conference, 2016.
- [13] D. Roth, J. L. Lugin, J. Büser, G. Bente, A. Fuhrmann, and Marc E. Latoschik, "A simplified inverse kinematic approach for embodied vr applications", In in

Proceedings of the 23rd IEEE Virtual Reality (IEEE VR) conference, 2016.

- [14] D. Medeiros, R. K. dos Anjos, D. Mendes, J. M. Pereira, A. Raposo, and J. Jorge, "Keep my head on my shoulders!: why third-person is bad for navigation in VR", Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology-VRST '18, pp. 1-11, 2018.
- [15] D. Bijl, H.-T. Henry, "Speech to Text Conversion". U.S. Patent. No. 6,173,259, 9 January 2001.
- [16] M. A. Siegler, "Measuring and Compensating for the Effects of Speech Rate in Large Vocabulary Continuous Speech Recognition", Thesis, Carnegie Mellon University, 1995
- [17] J. Adorf, "Web Speech API", KTH Royal Institute of Technology, 2013.
- [18] J. Busby, Z. Parrish, J. V. Eenwyk, "Mastering Unreal technology. Volume II: Advanced level design concepts with Unreal Engine 3" in, Sams Publishing, 2010.
- [19] B. G. Witmer, M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire", Presence: Teleoperators and Virtual Environments, Vol. 7, No. 3, pp. 225-240, 1998.

## 〈 저 자 소 개 〉

### 이 재 현

- 2011~2013.11 경북고등학교 졸업
- 2014~2018.06 중앙대학교 컴퓨터공학부 학사
- 2018.08~2020.08 중앙대학교 소프트웨어 대학 석사
- 2020.08~ 회사 근무
- <https://orcid.org/0000-0002-6354-5337>



### 박 경 주

- 2005.08 University of Pennsylvania, Computer and Information Science, Ph.D
- 2005.09~2007.02 삼성전자 무선통신연구소
- 2007.03~2017.02 중앙대학교 첨단영상 대학원 교수
- 2017.03~현재 중앙대학교 소프트웨어대학 교수
- <https://orcid.org/0000-0003-0198-4729>

