SMS 메시지에 따른 모바일 3D 아바타의 입술 모양과 음성 동기화

Lip and Voice Synchronization with SMS Messages for Mobile 3D Avatar

윤재홍, 송용규, 김은석, 허기택
동신대학교

Youn Jae-Hong, Song Yong-Gyu, Kim Eun-Seok, Hur Gi-Taek
DongShin Univ.

요약

모바일 3D 엔진을 탑재한 단말기의 등장과 모바일 콘텐츠 시장의 성장에 따라 3D 모바일 콘텐츠 서비스에 대한 관심이 고조되고 있다. 모바일 3D 아바타는 개인화된 모바일 기기 사용자의 개성을 표출할 수 있는 가장 효과적인 상품이다. 그러나 3D 아바타 표현 방법은 PC기반의 가상환경에서 캐릭터의 얼굴 표정 및 입술 모양 변화에 따른 자연스럽고 사실적인 표현에 대한 연구들이 주로 이루어 졌다.

본 논문에서는 모바일 환경에서 수신된 SMS 메시지를 3D 모바일 아바타에 적용하여 입술 모양 및 음성과 동기화시키는 방법을 제안한다. 제안된 방법은 수신된 메시지 문장을 음절단위로 분해하여 모바일 3D 아바타의 입술 모양과 해당 음성을 동기화시킴으로써, 모바일 아바타의 자연스럽고 효과적인 SMS 메시지 읽기 서비스를 구현할 수 있도록 해준다.

Abstract

There have been increasing interests in 3D mobile content service with emergence of a terminal equipping with a mobile 3D engine and growth of mobile content market. Mobile 3D Avatar is the most effective product displaying the character of a personalized mobile device user. However, previous studies on the method of expressing 3D Avatar have been mainly focused on natural and realistic expressions according to the change in facial expressions and lip shape of a character in PC based virtual environments.

In this paper, we propose a method of synchronizing the lip shape with voice by applying a SMS message received in mobile environments to 3D mobile Avatar. The proposed method enables to realize a natural and effective SMS message reading service of mobile Avatar by disassembling a received message sentence into units of a syllable and then synchronizing the lip shape of 3D Avatar with the corresponding voice.

I. 서 론

모바일 콘텐츠 및 무선인터넷 시장의 성장에 따라 모바일 싸이월드, 모바일 미팅, 모바일 TTS(Text To Speech) 등 기존 방식과는 달리 자발적인 콘텐츠 생성, 양방향 실시간 커뮤니케이션 등 모바일 서비스 본연의 특성을 그대로 살리고 있어 모바일 콘텐츠 서비스의 전형을 제시하고 있다. ARM9 프로세스 그래픽 연산 가속칩을 탑재한 고성능 3D 게임전용 폰의 등장과 함께 이동통신사들도 3D 모바일 게임 지원에 나서 차세대 모바일 게임 서비스를 제공하고 있다.

이렇게 게임 폰과 콘텐츠, 이를 뒷받침하는 이동통신상의 지원 서비스 등 3박자가 맞물리기 시작하면서 차세대 3D 모바일 게임 시장이 활발하게 떠오르고 있다. 단말기 성능의 업그레이드는 과거 메모리에 대한 한계와 내부 동작 속도 때문에 느린 게임 환경을 제공할 수밖에 없었던 문제에서 외장 메모리 지원, 강력한 2D/3D 그래픽 가속 기능 지원 등의 성능에 힘입어 3D 게임까지의 영역으로 확대되었다[1].

3D 모바일 콘텐츠의 개발 환경은 하드웨어의 가속기능과 단말기 메모리 가용량 확대로 인하여 콘텐츠의 용량이 커졌지만, 여전히 PC환경에 비해 콘텐츠의 용량에 제한을 받는다. 따라서 3D 캐릭터 애니메이션의 저용량화를 위한 연구의 필요성이 커지고 있다. 또한 3D 모바일 게임에서 단말기의 특성상 LCD 사이즈가 작기 때문에 3D 캐릭터가 중심이 된다.

모바일 상에서 3D를 구현하기 위해서는 모바일 기기에 적합한 3D 엔진과 API가 준비되어야 하며, 휴대용 단말기의 VM 위에서 3D엔진은 게임이나 각종 어플리케이션의 콘텐츠를 랜더링하고 디스플레이 하게 된다. PC 기반의 3D 콘텐츠를 작은 휴대폰으로 옮기는 것은 단말기의 배터리 용량, 디스플레이의 크기 한계, 모바일 3D 콘텐츠의 용량, 다른 UI환경

본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 지역문화산업연구센터(CRC) 지원사업의 연구결과로 수행되었음.

등과 같은 휴대 단말기 차원에서 여러 가지 한계를 가지고 있으며, 특히 3D 게임의 콘텐츠 용량의 한계를 극복하는 것이 중요한 일이다. 최근 PC나 콘솔기반의 대작 게임들이 3D 모바일 게임으로 제작되어 3D 캐릭터의 사실적이며 자연스러운 애니메이션이 강조되고 있다.

본 연구에서는 휴대용 단말기에 수신된 문자 메시지를 모바일 환경에 적합한 음성으로 변환하고, 변환된 음성과 모바일 환경 맞게 모델링된 3D 아바타의 입술모양 변화를 동기화함으로써 문자 음성 변환(TTS) 서비스의 가시적 효과를 높일 수 있는 방법을 제시하고자 한다.

Ⅱ. 관련연구

1. 3D 아바타를 이용한 입술 움직임

컴퓨터의 사용자 편의성을 위하여 음성적인 인식이나 영상처리 각종 센서를 이용한 입력기술이 발달해 왔고 음성 합성이나 3차원 그래픽스를 이용한 다양한 사용자 인터페이스 기술이 발전되어 왔다. 그 중에서도 가상 얼굴을 모델링하여 그 모델이 사용자에게 말하게 함으로써 컴퓨터의 휴먼에이젼트를 인식하도록 하는 기술이 연구 되어 왔다. Parker[6]의 파라메트릭 모델을 시작으로 영어에 대해서는 표정연출, 입술움직임을 통한 발음, 대화할 때 각 얼굴 요소들의 움직임이 연구되어 얼굴 모델에 구현되었고, 최근에도 활발히 연구되고 있다[7]. 이 분야는 사람과 컴퓨터의 메신저 역할로서 사용자에게 친근하게 정보를 전달하며, 사람의 대화를 가시화함으로써 음성학적 연구에 도움이 되기도 한다.

1974년 Parke는 발음하고자 하는 문장을 녹음하여 녹음된 트랙을 읽어 나가면서 발화의 위치를 조사하였다[4]. 그리고 비디오 카메라로 문장을 발음을 촬영한 후 rotoscoping을 통하여 각 프레임별로 변하는 입술의 모양을 추적하는 기법을 사용하였다. 하지만 이 기법은 실시간 처리가 어렵다는 단점과 비디오촬영 후, 음성을 녹음해야하는 경우에는 사용하지 못한다는 단점을 가지고 있다. 1986년 Pearce와 Hill은 사용자가 텍스트를 입력하였을 때 음성을 합성하고, 이와 동기화된 얼굴 애니메이션을 수행하는 시스템을 개발하였다[3]. 이 시스템은 음소들로 구성된 텍스트를 실제 발음된 음소들의 발음 기호 들로 전환하고 이 기호에 맞추어 음성합성과 애니메이션을 수행한다. 하지만 이 시스템은 합성된 음성의 질과 얼굴 이미지가 좋지 않다는 평가를 받았다. 1993년 Water와 Levergood는 이 시스템의 이미지 부분에 2D이면서 텍스쳐 맵핑된 얼굴을 사용하고, 시스템이 실시간으로 수행되도록 하였다[5]. 2차원 이미지이기 때문에 실시간 수행이 가능하였지만, 고개움직임이나 깊이 정보에 대한 현실감 등이 미진하였다.

2. 텍스트에 따른 음성 표현

일반적으로 문장 합성(Text to Speech) 시스템은 입력된 텍스트의 소리를 합성하거나 그래픽 에이젼트가 말하도록 하거나 할때 공통적으로 수행하는 작업이 있다. 이는 곧 텍스트를 분석하여 문자의 수열을 발음하기 직전의 음소수열로 전환하는 일이다. 이미 국어 음성합성분야에서는 문자전환 시스템을 개발되었다. 이러한 시스템들은 텍스트에 작용하는 대부분의 음운규칙을 구현하여 자연스러운 음성신호를 출력하고 있다. 자연스런 문장의 가시적 표현을 위해서 문장의 문자열을 다룰수 있는 방법은 이러한 음성합성기의 음운처리 모듈을 이용하는 것이다. [6]

3. 한글 음절 추출

한글은 초, 중, 종성들의 조합으로 만들어지는데, 한글 조합은 총 14,364자에 이르며, 현재 실용되는 한글 한자 표준코드(KSC5601)에서의 한글은 2350자이다[7].

[표 1] 한글 음절 분류

기본자음 기본모음 쌍자음 부자음 부모음	ㄱ,ㄴ,ㄷ,ㄹ,ㅁ,ㅂ,ㅅ,ㅇ,ㅈ,ㅊ,ㅋ,ㅌ,ㅍ,ㅎ ㅏ,ㅑ,ㅓ,ㅕ,ㅗ,ㅛ,ㅜ,ㅠ,ㅡ,ㅣ ㄲ,ㄸ,ㅃ,ㅆ,ㅉ ㄳ,ㄵ,ㄶ,ㄺ,ㄻ,ㄼ,ㄽ,ㄾ,ㄿ,ㅀ,ㅄ ㅐ,ㅒ,ㅔ,ㅖ,ㅘ,ㅙ,ㅚ,ㅝ,ㅞ,ㅟ,ㅢ
초성 중성 종성	기본자음(14)+쌍자음(5) 기본자음(10)+부모음(11) 기본자음(14)+쌍자음(2)+복자음(11)

받침이 있는 음절과 없는 음절로 분류하고, 받침이 없는 음절에 따른 입 모양 분류와, 분류된 입 모양에 받침이 결합하는 경우를 고려하여 한글의 음절을 분류 하게 된다.[8] 받침이 없는 경우, 모음에 의해 입 모양이 좌우된다. 그러나 입술소리인 'ㅁ', 'ㅂ', 'ㅃ', 'ㅍ' 일 경우에는 입술이 닫혀진 상태에서부터 입 모양이 시작되고 차후 모음에 영향을 받아 변하는 함구형으로 분류 될 수 있다. 받침이 있는 경우는 크게 'ㄱ', 'ㄴ', 'ㄷ', 'ㄹ', 'ㅁ', 'ㅂ', 'ㅇ' 의 7개 자음으로만 표현된다. 받침이 있는 경우의 입 모양은 모음에 의해 입 모양의 변화가 이루어지고 최종 생성되는 입 모양은 받침이 있는 음절에 따라 약간씩 차이를 보이지만 미묘하다 할 수 있다. 받침이 있는 경우, 입술소리인 'ㅁ', 'ㅂ', 'ㅃ', 'ㅍ'이 왔을 땐, 역시 입술이 닫혀진 상태로 최종 생성되는 입 모양이 된다[7].

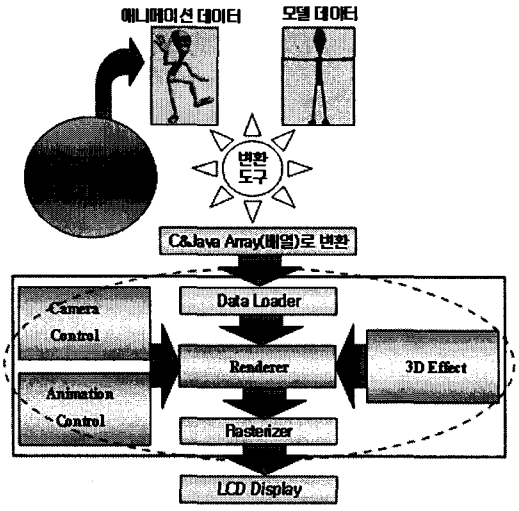
4. 모바일 3D 엔진 구조

모바일 3D 엔진은 저사양의 모바일 기기에서 실시간으로 3D 모델을 렌더링 및 애니메이션할 수 있도록 해주는 것이라

고 할 수 있다. 모바일 3D 엔진 내부에서는 기본적으로 2D와 3D 디자이너가 제작한 리소스들을 랜더링 파이프 라인 구조 형태로 데이터들을 처리한 후, 최종적으로 사용자 화면에 재생시킨다. 현재 시장에 나와 있는 모바일 3D 엔진들의 구조는 대부분 [그림 1]과 비슷하다.

최근 모바일 3D 업체에서 OpenGL-ES와 더불어 주목받고 있는 것이 JSR(Java Specification Request)-184이며, Jave 환경에서 최적화된 모바일 표준 3D 그래픽스 API이다. J2ME 환경에서 3D 그래픽을 구현하기 위해서는 저수준 의 OpenGL을 이용할 경우 코드가 길어져 MIDlet(MIDP Application)의 덩치가 커지므로 속도가 느려질 수밖에 없고, 자바 3D API를 이용할 경우에는 스펙의 양이 너무 방대하기 때문에 역시 MIDP(Mobile Information Device Profiles)에 이용하기엔 적합하지 않다. 이런 문제들로 하여금 만들어진 것이 JSR-184이다.

JSR-184는 OpenGL-ES보다는 고수준의 API로써 콘텐츠 제작 개념에 더 가깝고, SceneGraph 개념의 도입으로 3D World, 애니메이션, 그리고 이벤트 처리 자체를 엔진에서 처리할 수 있도록 정의 되어 있다.



▶▶ 그림 1. 모바일 3D 엔진 구조

Ⅲ. 3D 아바타 입술 모양 제어

1. 한글 발음에 따른 입 모양 분석

한글 음절은 총 9개의 입 모양 패턴으로 분류 하였다. 자음의 입술소리의 입 모양 변화 1개 패턴, 모음에 의한 입 모양 변화 9개 패턴, 받침이 있는 자음 입술소리의 입 모양 변화 1개 패턴으로 모든 한글 음절의 입 모양 변화 처리가 가능하다. 그 중 자음에 의한 입술소리 입 모양 변화 2개 패턴은 서로 입 모양이 일치 한다. 초성 자음이 입술소리일 경우와 받침 있는 자음 중 입술소리일 경우 두 가지 모두 입 모양이 함구된다. 그럼으로, 2개의 패턴을 1개의 패턴으로 사용하여 총 9개의 입 모양 추출 패턴이 된다. 사용자가 입력한 문자를 받은 후, 클라이언트 영역에서 초성, 중성, 종성으로 패턴을 분석 한 후, 자음 입술소리, 단모음, 중모음, 받침 있는 입술소리에 따라 그 인덱스 값을 넘겨받아 아바타의 입 모양과 1:1 매칭을 하여 확인 후 사용자에게 아바타의 입 모양의 변화를 보여 준다.

자음 입술소리의 입 모양 패턴은 기본적인 자음이 아닌 입술소리인 'ㅁ', 'ㅂ', 'ㅃ', 'ㅍ' 로만 입 모양 패턴을 분류 하였다. 나머지 다른 자음은 입 모양 변화가 모음에 따라 입 모양이 변화 되지만 자음의 입술소리는 그렇지 않고 최초 함구되는 것을 알 수 있다. 함구 모습에서 출발 하여 모음에 따라 입 모양이 변화되는 것이다. 받침이 있는 입술소리의 입 모양 패턴은 'ㅁ', 'ㅂ', 'ㅃ', 'ㅍ' 이 받침으로 올 경우에는 모두 입 모양이 함구된다. 최초 입 모양은 자음의 입술소리와 모음에 따라 변하게 되지만 최종 입 모양 변화는 받침이 있는 입술소리가 올 경우 'ㅁ', 'ㅂ', 'ㅃ', 'ㅍ' 에 영향을 받는다. 즉, 최종 입 모양이 함구 되는 것이다[7].

2. 한글 모음에 의한 입 모양 패턴

한글 모음에 의한 입 모양 패턴은 단모음 8개의 패턴을 분석한다.단모음의 경우 'ㅏ(ㅑ), ㅓ(ㅕ), ㅗ(ㅛ), ㅜ(ㅠ), ㅡ, ㅣ, ㅔ, ㅐ' 로 구분하여 초성의 입술소리 이외의 자음이 올 경우 입 모양의 변화를 단 모음에 따라 변하는 패턴을 분석한다. [표 1]은 자음과 단모음에 따른 입 모양 변화 시 적용되는 아바타의 입 모양 나타낸다[7].

[표 1] 자음과 단모음에 따른 입술 모양

자음	단모음			
함구	ㅏ	ㅓ	ㅗ	ㅜ
	ㅡ	ㅣ	ㅔ	ㅐ

3. 문자 메시지에 따른 모바일 3D 아바타의 입술 모양과 음성 동기화

휴대용 단말기에 수신된 문자 메시지에 따라 3D 아바타의 입술 모양을 변형하고, 변형된 3D 아바타와 한글 발음에 따른 음성을 동기화하기 위한 구조는 [그림 2]와 같다. 3D 모델링

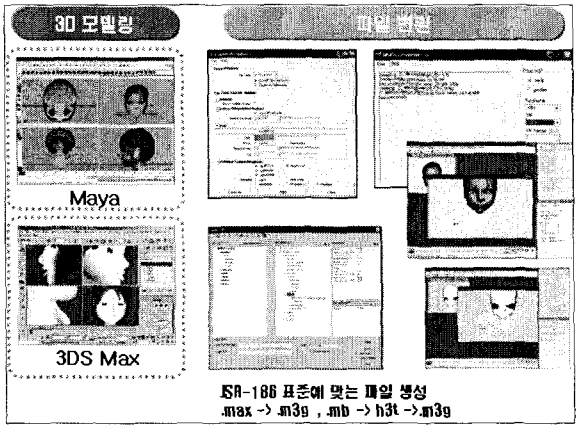
도구인 3dsMax나 Maya를 통해 모바일 환경에 적합한 캐릭터를 모델링 하고, 단말기 환경에서 구동될 수 있도록 JSR-184 규격에 맞는 ".M3G" 파일로 변환하도록 한다. 변환된 파일은 GNEX 환경의 Swerve 3D 엔진에 적재하여 아바타의 제어 애니메이션을 수행 할 수 있도록 한다.

모바일 장치에 수신된 한글 문자 메시지는 완성형 한글 코드를 사용하므로 완성형 한글 코드에 맞는 음성을 녹음하여 GNEX 환경에서 구동될 수 있는 형태로 변환하여 동일 발음을 제거한 후 인덱스를 구성하여 3D 아바타의 입술 모양 변화와 동기화 될수 있도록 한다. 또한 수신된 문자 메시지는 한글 코드에 해당하는 문자만 고려하도록 제한하였다.

▶▶ 그림 2. 모바일 3D 아바타의 입술 모양과 음성 동기화

Ⅳ. 실험결과

본 연구는 Swerve 3D 엔진을 사용하는 WIPI GNEX Emulator 환경에서 수행하였다. 3D 아바타는 Maya를 이용해 모델링하였으며, 한글 완성형 코드를 바탕으로 문자 발음에 해당하는 음성을 녹음하여 인덱스를 생성하고, 타이머를 사용하여 한글 코드와 매칭되는 3D 아바타의 입모양과 해당 음성의 동기화를 수행 하였다.



▶▶ 그림 3. 3D 아바타 생성 및 파일변환 과정

[그림 3]은 3D 아바타의 모델링 및 JSR-184 규격에 맞는 파일 변환 과정을 보이고 있다.

▶▶ 그림 4. 문자 메시지에 따른 3D 아바타와 음성 동기화

[그림 4] 는 문자 메시지에 따른 3D 아바타와 음성 동기화의 결과를 보이고 있다.

Ⅴ. 결론 및 향후 연구 방향

본 연구에서는 휴대용 단말기에 수신된 문자 메시지한국어 발음의 가시적인 표현을 위해 모바일 3D 아바타를 생성하여 GNEX Emulator에 적재하였으며, 모바일 환경에서 사용되는 한글 완성형 코드에 해당하는 음성을 녹음하여 인덱스를 구성한 후 3D 아바타와 동기화 함으로써 문자 음성 변환의 사실적인 가시적 효과를 높였다. 향후 단순 발음 뿐만아니라 문장 맺음 강조 등을 고려한 눈동자 움직임, 눈 깜박임, 고개 동작들을 표현하는 것과 모바일 3D 아바타의 표정 변화 까지 표현하기 위한 연구가 필요하다.

▌참 고 문 헌▐

[1] http://www.kisti.re.kr, "3D 모바일 게임 기술과 특허 동향", 한국과학기술 정보연구원

[2] P. Ekman and W. Friesen. "Manual for the Facial Action Coding System," Consulting Psychologists Press, Inc., Palo Alto, CA, 1978.

[3] David R. Hill, Andrew Pearce and Brian Wyvill, "Animating speech: an automated approach using speech synthesised by rules," The Visual Computer, Vol.3, No.5, pp.277-289, March 1988.

[4] Parke, F., "A parametrized model for facial animation," IEEE Computer Graphics and Applications, pp.61-70, 1982

[5] Keith Waters and Thomas M. Levergood, "DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces," Digital Equipment Corporation, Cambridge Research Lab, CRL 93/4 , September 1993.

[6] 최승걸, 안재우, 김응순, 얼굴 애니메이션을 위한 국어의 가시적 분석, 한국정보처리학회 추계학술발표논문집, 제2권, 제2호, pp.1292-1295, 1997.

[7] 김명수, 이현철, 김은석, 허기택, 얼굴 입모양 변화를 이용한 3D Avatar Messenger, 한국정보처리학회 춘계학술발표논문집, 제12권, 제1호, pp.225-228, 2005.