

캐릭터 에이전트를 이용한 교육용 컨텐츠 제작에 대한 연구

박상의*, 이해정, 정석태
원광대학교 컴퓨터공학과

e-mail : network@wonkwang.ac.kr

A Study for Education Contents Production using Character Agent

Sang-yi Park*, Hea-Jung Lee, Suk-Tae Joung,
Dept of Computer Engineering, Won-kwang University

요 약

본 논문은 LipSynchro 소프트웨어 개발 키트(SDK)를 이용하여 기존 2차원의 정지된 이미지를 모션생성엔진, 음성분석, 인식엔진과 함께 연동함으로써 캐릭터의 움직임을 자동으로 생성하여 사실적이고 살아있는 캐릭터 에이전트를 만들어, 이를 멀티미디어 교육용 컨텐츠제작 툴과 결합하여 보다 나은 교육용 컨텐츠를 생성할 수 있도록 하였다.

1. 서론

교사와 학생의 요구에 따른 교육 컨텐츠를 개발하고, 이를 공유하여 교육정보화의 목표를 달성하는 것이 현 디지털 교육사업의 당면과제이다. 정보통신 기술은 교사와 학생들이 지식정보사회에서 살아갈 수 있는 힘을 길러주는 교육을 하기 위한 핵심적인 위치를 차지하고 있으며 기존 텍스트만을 이용한 정보 전달에서 영상과 음향이 가미된 멀티미디어 기술은 정보공유 및 전달을 위한 대안으로 제시되고 있다. 캐릭터 음성연동 기술은 멀티미디어 교육 및 정보전달을 위한 디지털 컨텐츠 제작에 안정성 및 편리성을 가져올 수 있으며, 또한 강사의 음성만으로 비디오 영상 클립을 생성함으로써 캠으로 영상을 촬영해야만 했던 기존 방식과 비교하여 보다 나은 교육용 컨텐츠를 생성할 수 있도록 하였다. 빠르고 간편한 교육 영상컨텐츠 제작은 학생들에게 보다 많은 흥미를 유발시키고 유익한 정보를 빠르고 정확하게 전달할 수 있다. 현재는 인간과 컴퓨터간의 상호작용은 단순한 기계적 조작을 벗어나 좀 더 친숙한 형태로 나가고 있다. 그 중에서 궁극적인 상호작용 방법으로 간주되고 있는 것이 음성과 시각을 통한 인터페이스이다. 이는 음성과 시각을 통한 명령과 결과의 전달을 말하며, 이로 인해 음성압축, 음소인식, 캐릭터 에이전트와 연동되는 음성 처리기술 등이 필요하게 되었다. 이와 같은 기술들에 사람과 친숙한

형태를 가진 모델을 구현함으로써 인간과 컴퓨터간의 상호작용을 더욱 친밀하게 하는 인터페이스를 만들 수 있다.^[1]

본 논문은 캐릭터 에이전트의 장점을 살려 멀티미디어 교육과 접목시키기 위해 LipSynchro 소프트웨어 개발 키트(SDK)^[2]를 이용하여 모션생성엔진, 음성 분석, 인식엔진과 함께 연동한다. 모션생성엔진과 연동하여 제작된 3D캐릭터 에이전트는 마치 살아있는 것 같이 보여지며, 자동으로 3D캐릭터 에이전트의 머리와 눈동자 눈썹의 자동움직임과 자동 생성의 기본적인 움직임을 생성하였다.

본 논문의 2장에서는 관련연구와 LipSynchro SDK적용기술에 대해서 설명하였다. 3장에서는 시스템설계 및 구현결과를 보이고, 마지막으로 4장에서는 결론을 맺는다.

2. 관련연구와 LipSynchro SDK 적용기술

2.1 관련연구

현재 교육용 멀티미디어 컨텐츠 제작을 위한 효율적이고 생산적인 저작도구의 개발 및 연구가 활발히 진행되고 있다.^[3] 이로 인해 SMIL^[4] 활용을 위해 쉽고 효과적으로 SMIL문서를 작성할 수 있는 기능을 지원하는 인터페이스 개발과 다양한 기능을 지원하여 대중성을 높이는 많은 SMIL 저작도구^[5]가 출현되고 있으며, 이를 이용한 멀티미디어 프리젠테이션

제작을 위해 국내외에서 개발되어진 여러 저작도구들 있다. 그 중 대표적인 국내 에디터로는 (주)다산기술의 TagFree^{[6][7]}와 스마일미디어의 EZer^[8]가 있고, 외국의 에디터는 RealProducer G2^[9], T.A.G Editor, GRiNS Pro Editor for SMIL 2.0^{[10][11]}이 있다. 이 외에도 많은 SMIL 에디터들이 개발되면서 좀더 편리한 인터페이스를 제공하고 있으나 아직 국내에서의 개발은 미흡한 실정이다. 기존의 저작도구들의 기능을 통합하고 단점을 보완하여 보다 나은 SMIL 에디터의 개발이 요구되고 있다.

따라서 본 논문에서는 편리한 인터페이스를 제공하는 SMIL 에디터에 캐릭터 에이전트를 이용한 플레이어를 결합시킴으로서 더 많은 정보를 정확히 인식할 수 있는 교육용 멀티미디어 콘텐츠 저작 시스템을 설계하였다.

2.2 LipSynchro SDK 적용기술

Lip Synchro SDK는 입력된 음성을 분석/인식하여 자동으로 3D 캐릭터가 입력된 음성에 따라 립싱크하게 된다.

2.2.1 음성압축 (Speech Compression)

음성 압축은 효과적인 음성 신호의 처리로 14Kbps에서의 양방향 동시전송 방식을 지원한다. ITU-T Vocoder (G.723.1, G.723AB와 5.4Kbps, 4.8Kbps, 2.4Kbps, 1.2Kbps 에서 사용되는 Vocoder)를 채택하여 최적의 음성 알고리즘을 적용한다.

2.2.2 음소인식 (Phoneme Recognition)

음소인식은 인공신경 네트워크 기술 (ANN - Artificial Neural Networks Technology)에서 출발하였다. 음소인식은 신경네트워크와 음소 데이터베이스를 이용하여 음성의 특징을 분석하고 음소를 분류하여 개발하였다. 음소인식 엔진은 음성에서 음소

를 자동으로 분석하고 인식하며, 음소 그룹에서 음소를 세분화하고 입술 움직임에 적용시켜 자연스러운 애니메이션을 실시간으로 생성한다.

2.2.3 캐릭터 연동 음성 처리

사실과 같은 입술의 움직임은 얼굴 모델링 (Face Modeling)을 기반으로 생성되며, 이는 음성의 움직임이 3D모델의 프레임 시퀀스를 이용하여 보여진다. 3D 캐릭터의 입 모양을 실시간으로 연동시키는 과정은 음성 신호를 분석해서 얻은 정보를 이용하여 발음된 음소에 맞는 움직임을 보여 주게 된다. 캐릭터의 얼굴을 포함한 애니메이션은 음성만으로 자동 생성하게 된다.

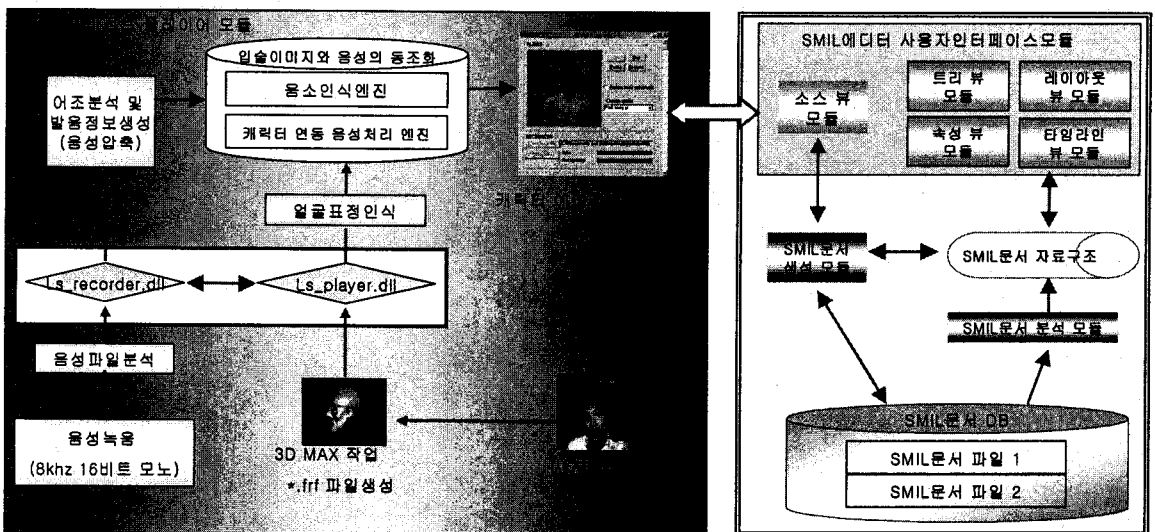
3. 시스템 설계 및 구현

3.1 시스템 설계

기존에는 사용자가 SMIL을 기반으로 교육용 멀티미디어 콘텐츠를 제작하기 위해서는 메모장이나 워드프로세서를 이용해야하므로 SMIL의 태그를 숙지해야하는 어려움과 이를 구현하는데 많은 시간과 노력을 요구하는 등의 문제점을 가지고 있었다. 또한 청각장애자들은 수업강의내용을 들을 수 없어 정보를 얻는데 많은 어려움을 가지고 있었다.

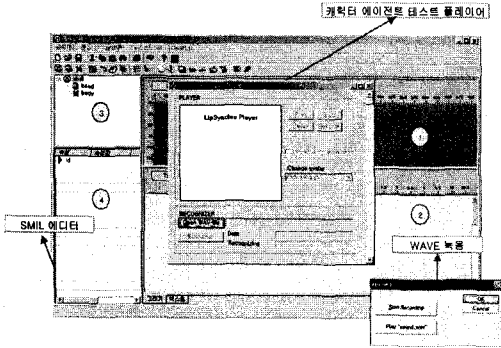
본 논문에서는 이러한 문제점을 해결하고 사용자가 좀더 쉽게 SMIL을 이용하여 교육용 멀티미디어 콘텐츠를 제작할 수 있도록 SMIL과 캐릭터 에이전트를 기반으로 하는 교육용 콘텐츠 저작 시스템을 설계하였으며, 이 시스템에 의해 만들어진 콘텐츠는 오디오정보와 비주얼정보를 상호보완적으로 사용함으로써 교육의 효과를 향상 시킬 수 있다. [그림 1]은 캐릭터 에이전트를 이용한 교육용 콘텐츠 저작 시스템의 전체적인 시스템 구성도를 나타낸다.

본 연구실에서 개발한 멀티미디어 교육용 콘텐츠



[그림 1] 전체적인 시스템 구성도

저작 툴인 SMIL에디터와 캐릭터 에이전트 플레이어를 결합한 전체적인 에디터의 초기화면은 [그림 2]과 같다. 이 그림에서 보는바와 같이 SMIL에디터의 사용자 인터페이스와 캐릭터 에이전트 테스트 플레이어, Wave 레코더가 동시에 작동된다.

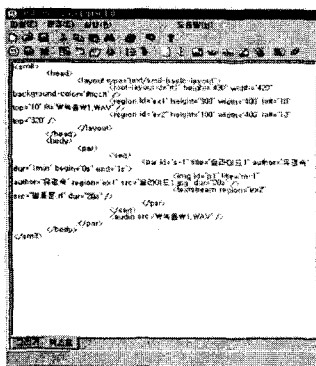


[그림 2] 전체적인 에디터 초기화면

3.1.1 사용자 인터페이스

기존에는 사용자가 SMIL 기반의 콘텐츠를 제작하는데 있어 단순한 텍스트 편집기를 이용하여 직접 코딩해야 하는 불편함과 텍스트 편집 위주의 연속적인 단조로운 작업으로 인해 많은 시간과 노력을 요구하는 문제점을 가지고 있었다. 이러한 문제점을 해결하기 위해 사용자 인터페이스 부분에서는 WYSIWYG 방식으로 멀티미디어 데이터들 사이에 공간적 동기화 정보를 설정하는 방식과 시간적 동기화 정보를 설정하는 시간편집기능으로 분리하였다.

사용자 인터페이스([그림2] 참고)는 그래픽 사용자 인터페이스를 사용하여 SMIL 문법의 기본적 지식만을 가진 사용자가 쉽게 SMIL 문서를 저작할 수 있도록 설계하였으며, 공간적 동기화 정보를 설정하는 ①레이아웃 뷰와 시간적 동기화 정보를 설정하는 ②타임라인 뷰, 그리고 ③트리 뷰, ④속성 뷰, [그림3]의 소스 뷰로 구분되어 있다.



[그림 3] SMIL에디터 소스뷰

3.1.2 캐릭터 에이전트 테스트 플레이어

캐릭터 에이전트 플레이어 모듈에서는 먼저 8Khz, 16bit의 wav음성녹음부분에서 mono의 녹음 방식을 사용하며 이 음성을 분석하여 Ls_record.dll 을 통해 음성압축 및 음소인식을 하게 된다. 또한 3D이미지로 추출된 캐릭터 정보를 Ls_Player.dll을 통해 얼굴 표정을 인식하게 된다. 이렇게 어조분석 및 발음정보가 생성되고 얼굴 표정인식 정보를 합하여 캐릭터 연동 음성 처리(3D Phoneme Visualization)를 통해 발음된 음소에 맞는 움직임 보여준다. 마지막으로 교육용멀티미디어 자료와 캐릭터 에이전트 플레이어가 결합된 기능을 수행하게 된다.

3.2 시스템 구현

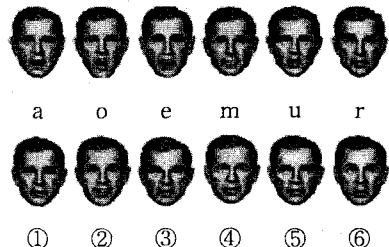
본 논문에서는 SMIL 기반 교육 저작 시스템의 사용자 인터페이스와 LipSynchro 소프트웨어 개발 키트(SDK)을 활용한 캐릭터 에이전트 플레이어에 중점을 두고 설계하였다.

사용되는 시스템 환경은 펜티엄3 64MB이상과 OS는 윈도우98이상의 버전을 사용하며, 먼저 Directx SDK 8.X 버전을 설치하고 LipSynchro 라이브러리파일(ls_player.dll, ls_recorder.dll) 두개를 레지스트리에 등록한다.

3.2.1. 음소(Phoneme)와 캐릭터

본 논문에서 사용되는 얼굴표정은 총 16가지로, 감정을 표현하는 얼굴표정으로는 기본표정, 두눈 감은표정, 왼쪽눈 감은표정, 오른쪽눈 감은표정, 웃는 표정, 놀란 표정, 화내는 표정, 왼쪽눈썹 올린표정, 오른쪽눈썹 올린표정의 총 9가지가 사용되며, 음소에 따른 얼굴표정은 a(아), o(오), e(에), f(프, 으), m(엠펜, 음), u(우, 유) other(기타) 발음으로 총 7가지가 사용된다.

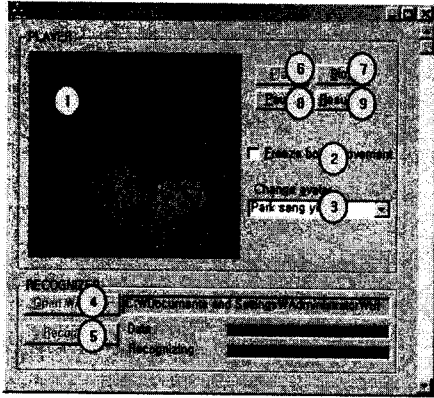
이와 같은 얼굴표정은 몰프타켓방식^[12]으로 디자인되었으며 이렇게 만들어진 각각의 얼굴표정에 Mesh name를 지정해줌으로서 프로그램에서 동작하게 된다. [그림4]의 첫 번째 줄은 필요한 음소를 나타내고 있으며, 두 번째 줄은 감정에 따른 캐릭터의 얼굴표정을 나타내는 것으로 ①은 기본표정, ②는 웃음, ③은 화냄, ④는 오른쪽 눈 감음, ⑤는 눈 감음, ⑥은 놀람을 표현하고 있다.



[그림 4] 음소(Phoneme)와 감정에 따른 캐릭터표정

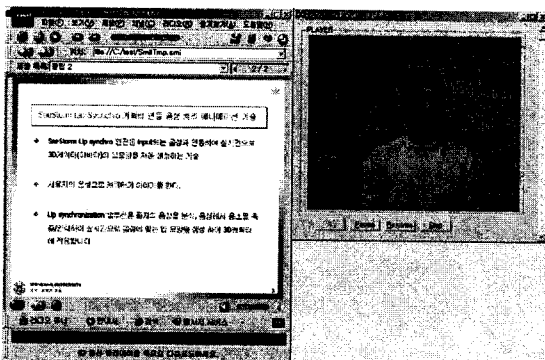
3.2.2 캐릭터 에이전트 플레이어 기능 및 동작

플레이어는 두 부분으로 구성되는데 하나는 원하는 웹 파일을 검색하고 불러들여 동작하는 플레이어부분과 SMIL에디터가 동작할 때 실행되는 플레이어부분으로 구성되며 다음 [그림5]과 [그림6]에서 보여지는 것과 같다.



[그림 5] 캐릭터 테스트 플레이어

먼저 웹파일을 가져와 테스트하는 기능에 대해 살펴보면, [그림6]과 같이 디지털카메라로 찍은 이미지를 3D MAX를 이용하여 캐릭터의 기본모습을 나타나도록 하였다.(①) 이 캐릭터는 몸동작을 움직이게 해주는 기능(②)을 넣어줄 수 있으며 원하는 캐릭터를 선택(③)하여 바꾸어 줄 수 있다. 캐릭터 에이전트를 동작하기 위해서는 음원을 가져와야 하는데 파일을 불러오는 버튼(④)이 필요하며 가져온 웹파일과 캐릭터 에이전트와 병합시켜주는 기능 버튼(⑤)이 있다. 음원을 불러들이면 실행버튼(⑥)이 활성화되어 실행될때 캐릭터 에이전트가 소리에 따라 동작하게 되며, 중단할 경우 Stop버튼을(⑦)실행하고 동작 중 일시정지와 해제는 (⑧, ⑨)버튼을 사용하면 된다.



[그림 6] 실행화면

SMIL에디터에서 강의를 제작한 것과 같이 동작되는 캐릭터 에이전트는 [그림 6]에서 보는바와 같이 강의에 필요한 기능만 추가하였으며 이 SMIL에디터 제작시 캐릭터 에이전트의 연동을 설정하여 캐릭터 에이전트와 동시에 작동되도록 한다.

4. 결론

본 논문에서는 인간에게 가까운 측면에서의 컴퓨터 사용을 돕는 3D 캐릭터를 이용하여 캐릭터 에이전트 플레이어를 구현 하였다. 이 캐릭터 에이전트 플레이어는 음성만으로 음소인식을 통해 정확한 발음에 따른 입 모양을 보이며 캐릭터 연동 음성 처리를 통해 발음에 따른 머리, 눈동자, 눈썹, 눈감박임 등의 다양한 움직임을 자동으로 생성하여 살아있는 캐릭터 에이전트를 멀티미디어 교육용 콘텐츠 저작툴과 결합하여 보다 나은 교육용 콘텐츠를 생성할 수 있도록 하였다.

참고문헌

- [1] <http://www.greenart114.co.kr/menu2/avatar.htm>
- [2] SeeStorm LipSynchro SDK
<http://www.seestorm.com/lipsynchrosdk.jsp>
- [3] Roger C., et. al, "Multimedia Application for Education and Training : Revolution or Red Herring?". ACM Computation Surveys, vol.27, No.4, pp. 633-635 December 1995
- [4] W3C, Synchronized Multimedia Integration Language(SMIL)1.0 Specification, <http://www.w3.org/TR/1998/12EC-smil-19980615,1998>
- [5] <http://my.dreamwiz.com/gojirael/html/smil/smil.html>
- [6] Tagfree, Tagfree 2000 SMIL Editor v1.0, <http://www.tagfree.com/english/product/producto2.asp?menu=2>
- [7](주)다산기술, <http://www.tagfree.com>
- [8] 스마일미디어, EZer SMIL 1.0, <http://www.smilmedia.com> No.4, pp. 633-635 December 1995.
- [9] RealNetworks, RealPlayer, <http://www.real.com/?PV> =6
- [10] Bulterman, D., et. al., "GRiNS : A Graphical Interface for creating and playing SMIL documents," Computer Networks and, ISDN Systems, Vol.30, No.1-7, April,1998.
- [11] Oratrix, GRiNS Pro Editor for SMIL 2.0, http://www.oratrix.com/products/G2E/index_html
- [12] <http://www.comet-cartoons.com/toons/3ddocs/headpolymodel/>